Journal of International Economics 47 (1999) 245–266

Journal of
INTERNATIONAL
ECONOMICS

Relative labor productivity and the real exchange rate in the long run: evidence for a panel of OECD countries

Matthew B. Canzoneri*, Robert E. Cumby, Behzad Diba

Department of Economics, Georgetown University, Washington, DC 20057, USA

Received 16 December 1996; received in revised form 6 August 1997; accepted 17 September 1997

Abstract

The Balassa-Samuelson model, which explains real exchange rate movements in terms of sectoral productivities, rests on two components. First, it implies that the relative price of non-traded goods in each country should reflect the relative productivity of labor in the traded and non-traded goods sectors. Second, it assumes purchasing power parity holds for traded goods. We test both of these using a panel of OECD countries. Our results suggest that relative prices generally reflect relative labor productivities in the long run. The evidence on purchasing power parity in traded goods is less favorable, at least when we look at US dollar exchange rates.

Keywords: Real exchange rates; Productivity

JEL classification: F11

1. Introduction

Changes in real exchange rates — defined as the relative price of national outputs — have been so persistent that they question the very notion of purchasing power parity (PPP). Even the evidence suggesting that deviations from PPP are temporary points to a half life of 4–5 years.[1] Explanations of persistent real

*Corresponding author. Tel.: +1-202-6875911; fax: +1-202-6876102; e-mail: canzonem@gunet.georgetown.edu

[1] Frankel and Rose (1996) and Lothian (1997) report half lives of this magnitude using aggregate price levels as do Wei and Parsley (1995) using sectoral prices.

PII: S0022-1996(98)00021-X

exchange rate changes have often followed the lead of Balassa (1964) and Samuelson (1964), who divide national output into traded and non-traded goods and explain real exchange rates in terms of sectoral productivity. The Balassa-Samuelson hypothesis divides real exchange rate movements into two components. Competitive behavior implies that the relative price of non-traded goods depends on the ratio of the marginal costs in the two sectors, and we will show that for a wide class of technologies the ratio of marginal costs is proportional to the ratio of average labor products in the two sectors. So, the first component of the hypothesis is the assumption that the relative price of non-tradeables is proportional to the ratio of average labor products. The second component is the assumption of PPP for traded goods.

The two components combine to produce a simple model of real exchange rate movements. If, for example, the ratio of traded goods productivity to non-traded goods productivity is growing faster at home than abroad, then the relative price of non-traded goods must be growing faster at home than abroad, and the price of home national output must be rising relative to the price of foreign national output (since by assumption the prices of traded goods equalize). In other words, if traded good productivity relative to non-traded good productivity is growing faster at home than abroad, then the home country should experience a real appreciation.

How well does the Balassa-Samuelson model explain real exchange rate movements? In Fig. 1 we plot four bilateral real exchange rates along with the ratio of labor productivity in the traded and non-traded goods sectors for each

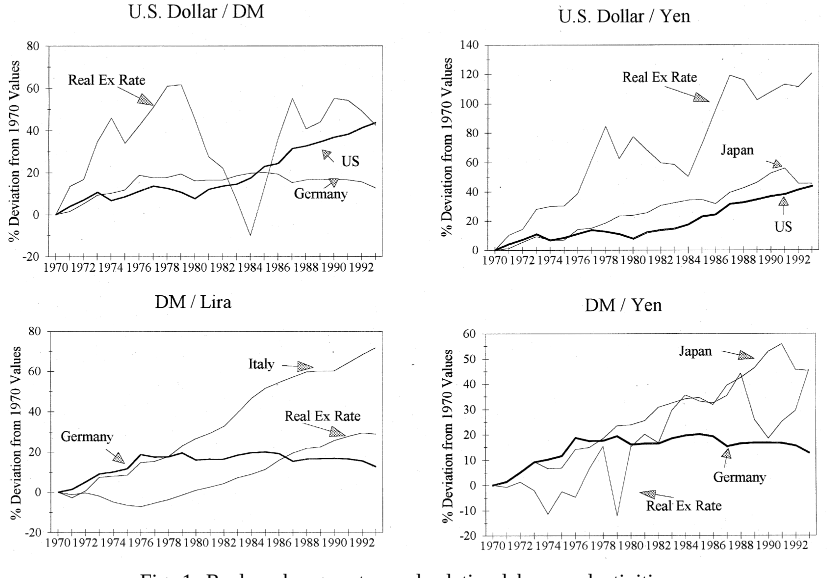

Fig. 1. Real exchange rates and relative labor productivities.

country. The Balassa-Samuelson hypothesis clearly fails to explain the short run movements in real exchange rates and Fig. 1 suggests that there may also be problems in explaining the long run movements, especially where the US dollar is concerned. Relative traded goods productivity has risen much faster in the United States than in Germany, especially since the late-1970s. This should have led to a real appreciation of the dollar (a fall in the real exchange rate as plotted in Fig. 1); instead, the dollar has depreciated. And the negligible difference between US and Japanese productivity trends has little hope of explaining the real depreciation of the dollar against the yen. The Balassa-Samuelson hypothesis does seem to fare better with DM real exchange rates. Since the mid-1970s, relative traded goods productivity has been growing more rapidly in both Italy and Japan than it has in Germany, and both the lira and the yen have appreciated in real terms against the DM.

Fig. 1 points to some problems with the Balassa-Samuelson hypothesis, but it does not reveal the source of the problems. Nor does it explain why there might be a problem with the dollar real exchange rates. Either component of the hypothesis can be challenged, and the special problem with the dollar — if one exists — could reside in either place.

In this paper we examine the two components of the Balassa-Samuelson hypothesis using a panel of 13 OECD countries. We do not require that either part of the hypothesis holds in the short run. In fact, both theory and a substantial body of evidence suggest that short-run deviations from both can be substantial. Instead, we focus our tests on long-run or trend behavior.

Two unfortunate facts have combined to plague the time series literature on long run real exchange rates: we typically only have 20 to 30 years of data on any pair of countries, and unit root tests have notoriously low power in small samples to distinguish between series that are non-stationary and series that are stationary but highly persistent.[2] Recently developed techniques allow us to deal with non-stationary data in heterogeneous panels, and in fact combining the data from 13 countries yields substantial benefits. We are able to confirm the existence of a long run (or cointegrating) relationship between the relative price of non-tradeables and the ratio of average labor products, and we are able to estimate its parameters with a surprising degree of precision.

The precision is so great that we are seemingly able to reject the first component of the Balassa-Samuelson model. This leads us to perform some Monte Carlo experiments to show that the remaining small sample bias is capable of explaining our apparent rejection of the model. We will argue that this firmly establishes the first component of the Balassa-Samuelson hypothesis.

We will argue instead that the problems with the Balassa-Samuelson hypothesis lie in the failure of PPP to explain traded goods prices, especially for the US

[2] Hakkio (1984) is the first to exploit the benefits of pooling when examining PPP.

dollar. Once again, using panel data we are able to establish that the appropriate long run (or cointegrating) relationship exists, but our estimates of its parameters are much less precise, and none of the parameter estimates are close to the values implied by PPP. And, as Fig. 1 would seem to suggest, we find that the failure of PPP to hold for traded goods may be largely a US dollar phenomenon. These large and persistent deviations from PPP in traded goods dominate US dollar real exchange rate movements and it is therefore difficult to explain those movements with differences in sectoral productivities.[3] When we use the DM as reference currency, the results are much more favorable.

The plan of the paper is as follows. In Section 2 we describe our notation and set out the two components of the Balassa-Samuelson hypothesis. We review the econometric methods we use in Section 3, we present our empirical results in Section 4. In Section 5 we offer some concluding remarks.

2. The analytical framework

We begin with the link between the relative price of non-traded goods and the relative productivities in the traded and non-traded goods sectors. The analytical framework we use is quite general. In each country, capital and labor are employed in the production of traded goods, X, and non-traded goods, H. Competition implies that labor is paid the value of its marginal product, and labor mobility implies that the nominal wage rate, W, is equal in the two sectors.

$$\frac{\partial X/\partial L^X}{\partial H/\partial L^H} = \frac{W/P^X}{W/P^H} = \frac{P^H}{P^X} = q \tag{1}$$

Eq. (1) states the familiar condition that the relative price of non-traded goods, which we denote as q, is equal to the slope of the production possibility curve.

To measure the marginal products we assume that the marginal product of labor is proportional to the average product of labor in each sector.

$$\frac{\partial X/\partial L^X}{\partial H/\partial L^H} = \frac{\varphi(X/L^X)}{\psi(H/L^H)} \tag{2}$$

With Cobb Douglas technologies, φ and ψ are the labor shares in value added in the traded and non-traded goods sectors. But Eq. (2) will hold under assumptions that are much less restrictive than Cobb Douglas. Average and marginal products will be proportional if the production functions can be expressed as,

---

[3]Engel (1995) examines real exchange rate variability over various horizons and reaches the same conclusion. Our results suggest that his conclusions might be sensitive to the choice of reference currency.

$$X = F(K^X)(L^X)^{\varphi}, \; H = G(K^H)(L^H)^{\psi} \tag{3}$$

where $F(.)$ and $G(.)$ are arbitrary functions that may, for example, depend on inputs other than the firms' choices of labor and capital as in the endogenous growth literature.[4] Moreover, if labor and capital are both mobile across sectors, then (2) holds in equilibrium even for some technologies that cannot be represented in the form of (3). For example, CES production functions with constant returns to scale satisfy (3) in equilibrium if both factors are mobile.

When average and marginal products are proportional, the relative price of non-traded goods, q, is proportional to the ratio of the average products of labor, x/h, in the two sectors:

$$q = \frac{P^H}{P^X} = \frac{\varphi}{\psi}\frac{X/L^X}{H/L^H} = \frac{\varphi}{\psi}\frac{x}{h} \tag{4}$$

Adding subscripts to denote country i at date t and taking logarithms, we get

$$\ln(q_{i,t}) = \ln\left(\frac{\varphi_i}{\psi_i}\right) + \ln\left(\frac{x_{i,t}}{h_{i,t}}\right) \tag{5}$$

Although the model described above does not distinguish short-run and long-run fluctuations, in our empirical tests, we interpret Eq. (5) as a restriction on the long-run trends in the relative price of non-traded goods and relative labor productivity in the two sectors. Thus, in Section 4 we test whether $\ln(q)$ and $\ln(x/n)$ are cointegrated and whether the cointegrating slope is one.

The recent literature has often focused on total factor productivity (TFP), reflecting an interest in assessing the relative importance of supply shocks (as proxied by TFP) and demand shocks (as proxied by government expenditures, etc.) in explaining real exchange rate movements.[5] Instead, we follow Hsieh (1982), Marston (1987), and others and use average products of labor. We therefore implicitly allow both supply and demand shocks to affect real exchange rates. We make this choice for four reasons. First, interpreting movements in Solow residuals as exogenous supply shocks is problematic.[6] Second, we do not need data on sectoral capital stocks, which are likely to be less reliable than data on sectoral employment and value added. Third, we have shown that our development of the first component of the Balassa-Samuelson hypothesis holds for a broader class of technologies than the Cobb-Douglas production function, which is used to

---

[4]The production functions specified in (3) are the general solutions to the differential equations: $\partial X/\partial L^X = \varphi(X/L^X)$ and $\partial H/\partial L^H = \psi(H/L^H)$. We require only that they satisfy the standard properties of production functions.

[5]De Gregorio et al. (1993), (1994), and Chinn and Johnston (1996) are examples.

[6]Evans (1992), for example, shows that measured Solow residuals are Granger caused by money, interest rates, and government spending and finds that one fourth to one half of their variation is attributable to variations in aggregate demand.

compute Solow residuals. And finally, our specification allows a sharper test of the first component of the Balassa-Samuelson hypothesis; in particular, we do not need to rely on outside estimates of labor's share in production.[7]

We now turn to the second part of the Balassa-Samuelson model, the assumption that traded goods prices are characterized by purchasing power parity. Let $E_{i,t}$ be the nominal exchange rate of currency i relative to currency 1 (the reference currency) at time t. If $E_{i,t}$ is expressed in units of the reference currency per unit of currency i, purchasing power parity implies that $E_{i,t}P^T_{i,t} = P^T_{1,t}$. If we define the PPP exchange rate for country i at date t as

$$r_{i,t} = \frac{P^T_{1,t}}{P^T_{i,t}} \tag{6}$$

purchasing power parity implies that the nominal exchange rate is equal to the PPP exchange rate. As is well known, purchasing power parity fails dramatically in the short run. The interesting question is whether it holds in the long run for traded goods. We therefore test whether $\ln(E_{i,t})$ and $\ln(r_{i,t})$ are cointegrated and whether the cointegrating slope is unity.[8]

3. A brief review of the econometric methods

Because the tests that we consider focus on the long-run or trend behavior of the relative prices, relative productivities and nominal and PPP exchange rates, we begin by examining the long-run or trend behavior of each of the series and then test the restrictions that the hypotheses imply on that long-run behavior. In all of the tests that we implement, we allow the short-run dynamics to be relatively unconstrained and focus only on the long-run behavior of the data.

In order to determine whether each of the series is better characterized by stationary deviations from a deterministic trend or by stochastic trends, or possibly by both a deterministic and stochastic trend we carry out augmented Dickey-Fuller (1979) unit root tests.[9] Im et al. (1995) propose testing for a of a unit root in a

[7]Asea and Mendoza (1994) also examine the link between relative prices and relative productivities. They find that the coefficient on total factor productivity has the correct sign and that its magnitude is reasonable given other estimates of factor shares. Our set-up, using average products, yields the sharper hypothesis that the slope is unity.

[8]We use two reference currencies, the US dollar and the DM. Because we include a constant term in our estimates, we consider relative, not absolute, purchasing power parity.

[9]We use the general-to-specific procedures suggested by Hall (1994) and Ng and Perron (1995) to determine the number of lags and do tests both with and without deterministic trends.

heterogeneous panel that is based on the average over the N cross sections of the ADF t-ratio, $\bar{t}_N$. They show that the statistic,

$$\sqrt{N}\left(\frac{\bar{t}_N - a_T}{\sqrt{b_T}}\right) \tag{7}$$

is distributed asymptotically as a standard normal. The mean and variance of $\bar{t}_N$, a_T and b_T depend only on T and the average lag length and are tabulated by Im et al. (1995).[10]

If the data contain stochastic trends, the Balassa-Samuelson model implies that pairs of series must share the same stochastic trend. That is, they must be cointegrated. For each country, we test for cointegration using both standard, residual-based tests and the panel tests proposed by Pedroni (1995). We estimate the regression, $y_{i,t} = \alpha_i + \beta_i\, z_{i,t} + \epsilon_{i,t}$ and test the null hypothesis that the estimated residuals have a unit root as suggested by Engle and Granger (1987) and Phillips and Ouliaris (1990). Pedroni (1995) proposes a series of tests of cointegration in heterogeneous panels that can be viewed as extensions of these single-equation tests.[11] Two of his tests extend the semiparametric Phillips-Ouliaris (1990) tests. The third extends the ADF test.[12]

The Balassa-Samuelson model implies not only that pairs of variables should be cointegrated, but also that the slopes, β_i, of the cointegrating relationships should have the common value of 1.0. We explore this hypothesis in two ways. First, we impose $\beta_i = 1.0$ and test whether that restriction is consistent with the data. Next, we compute the Phillips-Hansen fully modified OLS estimates of β and the corresponding standard errors and test the null hypothesis that $\beta_i = 1.0$ both country-by-country and with the panel tests proposed by Pedroni (1996).

We test whether the restriction $\beta_i = 1.0$ is consistent with the data in two ways. First, we test for unit roots in the difference $y_{i,t} - z_{i,t}$ using ADF tests for each cross section along with the Im-Pesaran-Shin tests for the panel. In addition, as proposed by O'Connell (1997), we pool the data and estimate a common value of the coefficient on $y_{i,t-1} - z_{i,t-1}$ in the ADF regressions using a GLS estimator to take account of cross-equation correlation in the errors. We then do panel unit root test proposed by Levin and Lin (1992). Second, we carry out tests proposed by

---

[10]As they suggest, we use common time dummies to account for correlation across cross section units. Our cross sections differ in the number of observations that we have available so we use all of the available data to compute the t-ratios for each cross section.

[11]The parameters α_i and β_i in the cointegrating regression for each cross section are allowed to differ as are the dynamics of $y_{i,t}$ and $z_{i,t}$.

[12]Our test statistics are a slight modification of Pedroni's that allows for a different number of time series observations in each cross section. Again, we account for cross section correlation by using common time dummies to remove shocks common to all cross sections.

Horvath and Watson (1995) that are valid when the cointegrating vector contains known parameters. Their test is based on the error correction representation,

$$\begin{aligned}\Delta y_{i,t} &= \kappa_{1,i} + \lambda_{1,i}(y_{i,t-1} - z_{i,t-1}) + \sum_{j=1}^{s_i} \phi_{1,i,j}\Delta y_{i,t-j} + \sum_{j=1}^{s_i} \psi_{1,i,j}\Delta z_{i,t-j} + \nu_{1,i,t}\\ \Delta z_{i,t} &= \kappa_{2,i} + \lambda_{2,i}(y_{i,t-1} - z_{i,t-1}) + \sum_{j=1}^{s_i} \phi_{2,i,j}\Delta y_{i,t-j} + \sum_{j=1}^{s_i} \psi_{2,i,j}\Delta z_{i,t-j} + \nu_{2,i,t}\end{aligned} \tag{8}$$

Under the null hypothesis that the variables are not cointegrated with $\beta_i = 1.0$, $\lambda_{1,i} = \lambda_{2,i} = 0$. If the constant terms are zero, the Wald and likelihood ratio statistics have standard distributions. But, as is the case here, when the constants are non-zero, the test statistics have a non-standard distribution that is tabulated by Horvath and Watson. Unfortunately, the properties of a panel Horvath-Watson test have not yet been derived so we are restricted to country-by-country tests.

The fully modified OLS tests are based on the representation,

$$\begin{aligned}y_{i,t} &= \alpha_{1,i} + \beta_i z_{i,t} + \epsilon_{i,t}\\ \Delta z_{i,t} &= \alpha_{2,i} + \varsigma_{i,t}\end{aligned} \tag{9}$$

where we make no assumptions about the exogeneity of the regressors. The FMOLS estimates correct for both endogeneity and serial correlation in the errors. Pedroni (1996) proposes two tests of the hypothesis that the (common) cointegrating slope for heterogenous panels, β, is equal to some value, β_0. Both tests can be thought of as extensions of fully modified OLS to heterogeneous panels in which the intercepts $\alpha_{1,i}$ and $\alpha_{2,i}$ can differ across i as can the dynamics of $\eta_{i,t} = (\epsilon_{i,t}, \zeta_{i,t})$. The first is the group-mean t-ratio (as in Im et al., 1995) and the second is based on a t-ratio from the pooled data. Both have standard normal distributions.

4. Empirical results

Before examining the hypotheses of interest, we examine the trend behavior of each series, and find little evidence against the null hypothesis of a unit root in relative productivities, relative prices of non-traded goods, nominal exchange rates, or PPP exchange rates.[13] Because the evidence points overwhelmingly to unit roots we proceed to look at cointegrating relationships.

[13] The results are available in Canzoneri et al. (1996). Post-unification German data present a problem — there is a sharp divergence of relative productivities and relative prices. This could be due to a data problem associated with unification or to transition-related developments. The question needs further study once the data issues are more settled.

4.1. Relative prices and relative productivities

We begin with the hypothesis that, in the long run, the relative price of non-traded goods reflects relative productivities in the traded and non-traded goods sectors. Table 1 contains the results of tests of the null hypothesis that relative prices and relative productivities are not cointegrated. The tests using individual country data yield mixed evidence. The ADF-type tests reject the null hypothesis of no cointegration for eight of the 13 countries at the 10% level, and seven of these eight test statistics are also significant at the 5% level. The other two tests provide less evidence against the null. The panel tests provide strong evidence that relative prices and relative productivities are, in fact, cointegrated. We reject the null of no cointegration at the 5% level with the ADF-type tests and at the 10% level with the semiparametric *t*-tests.

We conclude from Table 1 that the relative prices of non-traded goods and the relative productivities in the traded and non-traded goods sectors are cointegrated as the Balassa-Samuelson model predicts. In Table 2 and Table 3 we turn to the

Table 1
Tests for cointegration of $\ln(q_{i,t})$ and $\ln(x_{i,t}/h_{i,t})$

$$\ln(q_{i,t}) = \alpha_i + \beta_i \ln(x_{i,t}/h_{i,t}) + \epsilon_{i,t} \Delta\hat{\epsilon}_{i,t} = (\rho_i - 1)\hat{\epsilon}_{i,t-1} + \sum_{j=1}^{k_i} \gamma_{ij} \Delta\hat{\epsilon}_{i,t-j} + \nu_{i,t}$$

Country	$T(\hat{\rho}-1)$	$\hat{t}_\rho$(PP)	$\hat{t}_\rho$(ADF)
United States	* −18.361	* −3.378	** −3.859
Canada	−11.419	−2.451	** −4.664
Japan	−8.643	−3.052	** −3.521
Germany	−15.629	* −3.173	** −3.837
France	−8.943	−2.339	−2.972
Italy	−5.009	−1.536	−2.842
Great Britain	−4.969	−1.276	−1.356
Belgium	−17.931	** −4.161	** −4.254
Denmark	−6.184	−1.728	−1.796
Sweden	−7.142	−1.882	** −4.278
Finland	−15.888	−3.034	** −3.852
Austria	−11.102	−3.021	−3.066
Spain	−9.809	−2.494	* −3.403
Panel tests of cointegration with common time dummies			
All countries	−19.508	* −8.267	** −8.803
G-7 countries	−14.360	−5.621	** −6.950

Significance at the 95% and 90% levels are noted by ** and *, respectively. The $T(\hat{\rho}-1)$ and $\hat{t}_\rho$(PP) statistics are Phillips-Perron (Phillips and Perron, 1988) tests applied to the residuals from the first regression and are computed with $k_i = 0$. The $\hat{t}_\rho$(ADF) is the augmented Dickey-Fuller statistic. Critical values are taken from the tables compiled by Phillips and Ouliaris (1990). The panel cointegration tests are those proposed by Pedroni (1995), which is the source of the critical values.

Table 2
Tests Using $\delta_{i,t} = \ln(q_{i,t}) - \ln(x_{i,t}/h_{i,t})$. A: tests for unit roots (H_0: $\rho_i = 1$)

$$\Delta\delta_{i,t} = \theta_i + (\rho_i - 1)\delta_{i,t-1} + \sum_{j=1}^{k_i} \gamma_{ij}\, \Delta\delta_{i,t-j} + \eta_{i,t}$$

B: Horvath-Watson tests (H_0: $\lambda_{1,i} = \lambda_{2,i} = 0$)

$$\Delta\ln(q_{i,t}) = \kappa_{1,i} + \lambda_{1,i}\delta_{i,t-1} + \sum_{j=1}^{s_i} \phi_{1,i,j}\, \Delta\ln(q_{i,t-j}) + \sum_{j=1}^{s_i} \psi_{1,i,j}\, \Delta\ln(x_{i,t-j}/h_{i,t-j})$$
$$+ \nu_{1,i,t}\Delta\ln(x_{i,t}/h_{i,t}) = \kappa_{2,i} + \lambda_{2,i}\delta_{i,t-1} + \sum_{j=1}^{s_i} \phi_{2,i,j}\, \Delta\ln(q_{i,t-j})$$
$$+ \sum_{j=1}^{s_i} \psi_{2,i,j}\, \Delta\ln(x_{i,t-j}/h_{i,t-j}) + \nu_{2,i,t}$$

Country	$\hat{t}_{\rho-1}$ (ADF)	LR(λ)
United States	−2.245	8.069
Canada	*2.669	7.557
Japan	−1.742	5.758
Germany	*−2.875	**14.433
France	−1.716	4.810
Italy	**−3.361	**15.630
Great Britain	−1.149	5.319
Belgium	−0.199	6.022
Denmark	−1.570	4.488
Sweden	−2.452	4.977
Finland	−1.003	**11.296
Austria	*−2.709	*9.198
Spain	**−3.819	3.632
Panel unit root test with common time dummies		
All countries	**−3.762	
G-7 countries	**−2.422	

Significance at the 95% and 90% levels is noted by ** and *, respectively. The $\hat{t}_{\rho-1}$(ADF) is the augmented Dickey-Fuller statistic and LR(λ) is the Horvath-Watson likelihood ratio test. The joint tests are the panel unit root tests proposed by Im et al. (1995) and are distributed as standard normal.

stronger predictions of the Balassa-Samuelson model that the slope in the cointegrating relationship is 1.0. In Table 2 we report the results of tests of the null hypothesis that there is a unit root in the difference between the (log) relative price and the (log) relative productivities and tests of whether the coefficients on the lagged differences are jointly zero in the error correction representation. Once again the evidence is mixed when we look at the individual country tests. The ADF tests reject the null hypothesis of a unit root in the difference at the 5% level for only two of the 13 countries and at the 10% level for three additional countries.

Table 3
Estimates of the cointegrating slope coefficient of $\ln(q_{i,t})$ and $\ln(x_{i,t}/h_{i,t})$

$$\ln(q_{i,t}) = \alpha_i + \beta_i \ln(x_{i,t}/h_{i,t}) + \epsilon_{i,t}$$

Country	$\hat{\beta}_{i,\mathrm{OLS}}$	$\hat{\beta}_{i,\mathrm{FMOLS}}$	$t(\hat{\beta}_{i,\mathrm{FMOLS}}=1)$
United States	0.877	0.869	** −4.209
Canada	0.633	0.675	** −2.602
Japan	1.237	1.181	**3.145
Germany	1.074	1.063	1.570
France	0.802	0.779	** −3.297
Italy	0.900	0.922	−1.212
Great Britain	0.412	0.443	** −7.002
Belgium	0.773	0.763	** −12.219
Denmark	0.552	0.510	** −7.888
Sweden	0.624	0.571	** −3.713
Finland	0.804	0.770	** −3.194
Austria	0.959	0.929	−1.398
Spain	0.895	0.859	** −2.141
Panel tests of $\hat{\beta}=1$ with common time dummies			
	$t_{\hat{\beta}}$	$\sqrt{N}\,\bar{t}$	
All countries	** −7.309	** −6.206	
G-7 countries	** −2.055	* −1.693	

The second column contains the ordinary least squares estimates of the slope coefficient. Column 3 contains the Phillips and Hansen (1990) fully modified OLS estimates of the slope and column 4 contains the *t*-ratio formed by subtracting one from the fully modified OLS estimate of the slope and dividing by the corresponding standard error. The numbers of lags used in computing the fully modified OLS estimate are chosen using the data-dependent procedure proposed by Newey and West (1994). The panel tests are those proposed by Pedroni (1996).

Similarly, the Horvath-Watson tests reject the null hypothesis that the series are not cointegrated with $\beta_i = 1$ at the 5% level for only three countries and at the 10% level for an additional country. Once again, the gain in power from the panel tests is clear. When we include common time dummies, we reject the null hypothesis of a unit root at the 5% level for all countries and for the G-7 countries.[14]

Monte Carlo evidence presented by Im et al. (1995) suggests that their test is generally quite reliable even for relatively small values of N and T (provided the lag length is chosen correctly) but that size distortions can arise for some types of serial correlation in the data. In order to determine how the test performs in samples like our unbalanced panel, we do Monte Carlo experiments in which the data are generated with a unit root and serial correlation that matches our data. In experiments with 2000 replications in which the lag length is data determined, we

[14] We also reject the null hypothesis of a unit root in the difference both for all countries and for the G-7 countries at the 95% level when we use GLS estimates to account for cross sectional dependence.

find only slight size distortions. The tests reject 6.25% of the time at the 5% level and 10.35% of the time at the 10% level. Small sample problems with the tests do not appear to be responsible for the results in Table 2.

The evidence in Table 2 is consistent with the long-run proportionality of relative prices and relative productivities. An alternative means of testing whether the slope of the cointegrating relationship is 1.0 is to estimate the slope and do a *t*-test. We report the results of these tests in Table 3. With three exceptions, Denmark, Britain, and Sweden, the slope coefficients are generally close to 1.0. The average fully modified OLS slope estimate is roughly 0.8. Excluding the three countries with slopes around 0.5, the average is nearly 0.9. The slopes are fairly precisely estimated, however, and we can reject the null hypothesis that the slope is 1.0 at the 5% level for 10 of the 13 countries. The panel tests confirm these results. When we include common time dummies the *t*-ratio for the null that $\beta = 1.0$ is around -7.0 when we use all of the countries in the sample, although it is considerably lower when we use only the G-7 countries.

The results of the formal tests are reflected in Fig. 2 where we plot the (log) relative price of non-traded goods and the (log) relative labor productivities in traded and non-traded goods (normalized so that $1970 = 0$) for four countries: Germany, Japan, Italy, and the United States. The differences between the two series appear to be transitory but the data for the United States and Italy exhibit a tendency, common to many of the countries we examine, for the relative

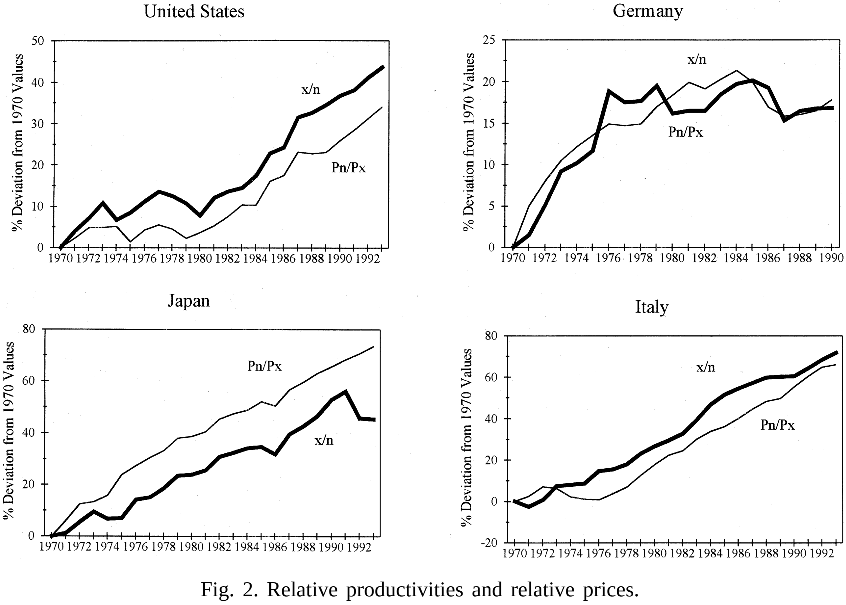

Fig. 2. Relative productivities and relative prices.

productivities to grow by more than the relative price. This tendency is reflected in the slope estimates, which are generally less than one.

4.2. Monte Carlo experiments

Table 2 and Table 3 contain contradictory results. The results in Table 2 are consistent with a slope of 1.0 while the results from Table 3 are not. One explanation for the difference is that the unit root tests in Table 2 have low power against alternatives with $\beta_i \neq 1$ (as suggested by Evans and Lewis, 1994). Another is the small sample properties of the FMOLS estimator. Phillips and Hansen (1990) find that FMOLS estimates can exhibit considerable small sample bias when ϵ and ζ (from (9)) are persistent and when they are highly correlated. We explore whether small sample problems might be responsible for the differences between the results in Table 2 and Table 3, with a series of Monte Carlo experiments. First we verify that the FMOLS estimates perform well in time series of 25 observations when ϵ and ζ are iid. Then we choose three countries: one (United States) with a slope estimate around 0.9, one (Belgium) with a slope estimate around 0.8, and one (Denmark) with a slope estimate around 0.5. For each, we estimate a VAR for $\eta = (\epsilon, \zeta)'$ and use the parameters and error covariance matrix to generate 15 000 samples of 25 observations in which the true value of β is 1.0. We also examine the small sample properties of the pooled tests statistics from the corresponding 1000 panels of 15 countries and 25 observations.

In all three cases we find that the FMOLS estimate of the slope is biased and that using the asymptotic distribution would lead to rejecting the null hypothesis of $\beta = 1.0$ too frequently. The mean FMOLS estimates of β are: 0.90 when we use the US parameters, 0.77 when we use the Belgian parameters, and 0.86 when we use the Danish parameters. Thus for two of the three we find that our point estimate from Table 3 is roughly equal to the mean value from the Monte Carlos in which the true value of β is unity. The point estimate computed from the Danish data (0.54) is below the average for the third and is significantly different from 1.0 at the 5% level even using the empirical Monte Carlo distribution. The tendency for the FMOLS estimates to reject the null of $\beta = 1.0$ too frequently is also reflected in the panel test statistics.

The Monte Carlo evidence suggest that one would tend to find slope estimates below 1.0 (relative productivity changes outstripping relative price changes) even if the true slope was, in fact, 1.0. Thus the test statistics reported in Table 3 should probably not be taken as evidence against the null hypothesis that β is one.[15]

Fig. 2, the results in Table 1, Table 2 and Table 3, and the Monte Carlo experiments taken together lend support to the first part of the Balassa-Samuelson

[15]Pedroni (1996) presents extensive evidence on the panel tests' small sample properties.

hypothesis. The relative prices of non-traded goods and the relative productivities in traded and non-traded goods appear to be cointegrated and the slopes of the cointegrating relationships are close to 1.0 as the hypothesis predicts. Thus relative prices and relative productivities appear to be proportional in the long run.

4.3. Purchasing power parity in traded goods

Next we turn to testing long run purchasing power parity in traded goods. In order to determine if any of our tests are sensitive to the choice of reference currency, we carry out all tests using both the US dollar and the DM.[16]

Table 4 presents the results of tests of the hypothesis that the nominal and PPP exchange rates are not cointegrated. The tests carried out on the data from individual countries yield mixed evidence. The ADF tests reject the null hypothesis of no cointegration at the 5% level for five of the 12 currencies when the dollar is the reference currency. When the DM is used as the reference currency we reject the null at the 5% level for the three Scandinavian countries and at the 10% level for one additional country. The benefit of pooling the data in the panel tests is, once again, clearly apparent. When we include common time dummies, all three tests reject the null of no cointegration at the 5% level for both reference currencies.[17]

The results in Table 4 suggest that nominal exchange rates and PPP exchange rates are cointegrated. If purchasing power parity holds for traded goods, they should be cointegrated with a cointegrating slope of 1.0. Next we test this stronger restriction of purchasing power parity and report the results in Table 5 and Table 6. If purchasing power parity holds in the long run for traded goods, the difference between the nominal exchange rate and the PPP exchange rate should be stationary. In Table 5 we report the results of tests of the null hypothesis that there is a unit root in the difference as well as the Horvath-Watson tests of the null hypothesis that nominal and PPP exchange rates are not cointegrated with a slope of 1.0. The tests using the data from each country individually again provide mixed evidence. Regardless of whether the dollar or the DM is the reference currency, the ADF tests reject at the 10% level for five of the 12 currencies. The evidence from the Horvath-Watson tests is also mixed, with more evidence against the null when the dollar is used as the reference currency than when the DM is

[16] Some existing evidence suggests that it might. Frenkel (1981) finds that PPP held more closely in the 1970s when the DM is used as the reference currency than when the dollar is used. More recently, Jorion and Sweeney (1996), Papell (1997), Papell and Theodoridis (1997), and Edison et al. (1997) all find a similar influence of the choice of reference currency in their samples.

[17] The similarity of the panel test statistics across the two reference currencies is no accident. Changing the reference currency amounts to adding the same log exchange rate to each cross-sectional unit, it will have no impact on tests computed using deviations from cross-section means, at least for large N. With $N=12$, a change of reference currency results in only minimal changes to the test statistics. See O'Connell (1997) and Engel et al. (1997).

Table 4
Tests for cointegration of $\ln(E_{i,t})$ and $\ln(r_{i,t})$

$$\ln(E_{i,t}) = \alpha_i + \beta_i \ln(r_{i,t}) + \epsilon_{i,t}$$

$$\Delta\hat{\epsilon}_{i,t} = (\rho_i - 1)\hat{\epsilon}_{i,t-1} + \sum_{j=1}^{k_i} \gamma_{ij}\Delta\hat{\epsilon}_{i,t-j} + \nu_{i,t}$$

	Reference currency: US dollar			Reference currency: DM		
Country	$T(\hat{\rho}-1)$	$t_{\hat{\rho}}$ (PP)	$t_{\hat{\rho}}$(ADF)	$T(\hat{\rho}-1)$	$t_{\hat{\rho}}$(PP)	$t_{\hat{\rho}}$(ADF)
Canada	−10.196	−2.294	** −3.459	−8.261	−2.052	−2.404
Japan	−9.062	−2.404	−2.466	−9.193	−2.221	−2.663
Germany	−8.261	−2.052	−2.404	−13.245	−2.661	−2.178
France	−10.345	−2.321	** −3.481	−8.285	−2.325	−2.096
Italy	−8.695	−2.112	−3.074	−9.870	−2.729	−2.995
Great Britain	−15.880	−3.049	** −4.713	−8.863	−2.311	−2.673
Belgium	−7.141	−1.822	−2.635	−0.476	−0.439	−0.793
Denmark	−7.166	−1.909	−2.740	* −20.531	** −4.817	** −4.170
Sweden	−7.081	−1.863	−2.414	−17.552	** −5.591	** −5.881
Finland	−12.350	−2.400	** −3.602	* −18.882	** −3.486	** −4.114
Austria	−7.618	−2.025	−2.071	−10.713	−2.679	−2.728
Spain	−8.692	−2.094	** −3.668	−15.143	−3.055	* −3.119
Panel tests of cointegration with common time dummies						
All countries	** −45.769	** −9.839	** −10.587	** −45.448	** −9.827	** −11.782
G-7 countries	** −40.717	** −9.475	** −11.169			
European countries				** −38.040	** −8.824	** −10.349

Statistical significance at the 95% level or greater is signified by **. Significance at the 90% level or greater is signified by *. The $T(\hat{\rho}-1)$ and $t_{\hat{\rho}}$(PP) statistics are Phillips and Perron tests applied to the residuals from the first regression and are computed from regressions setting $k_i = 0$ in the second regression. The $t_{\hat{\rho}}$(ADF) is the augmented Dickey-Fuller statistic. Significance levels for these tests are taken from the tables compiled by Phillips and Ouliaris (1990). The joint tests are the panel cointegration tests proposed by Pedroni (1995). Significance levels for these tests are computed from tables compiled by Pedroni.

used. The panel test points to a rejection of the null hypothesis when common time dummies are used.[18]

In Table 6 we present estimates of the slopes of the cointegrating relationships along with tests of the hypothesis that the slope is 1.0. When we use the dollar as the reference currency, we reject the null hypothesis that the slope is 1.0 for eight

[18]The test statistics differ slightly for the two reference currencies. These differences arise from slight difference in sample size (as we note in Appendix A, we do not use post-unification German data) and because $N = 12$. In addition, Papell and Theodoridis (1997) argue that panel tests will be invariant to the choice of reference currency only when the same number of lags is used for all currencies. Once again, the unit root tests using GLS to account for cross-sectional dependence reject the null hypothesis at the 95% level.

Table 5
Tests using $\mu_{i,t} = \ln(E_{i,t}) - \ln(r_{i,t})$A: tests for unit roots (H_0: $\rho_i = 1$)

$$\Delta\mu_{i,t} = \theta_i + (\rho_i - 1)\mu_{i,t-1} + \sum_{j=1}^{k_i} \gamma_{ij}\Delta\mu_{i,t-j} + \eta_{i,t}$$

B: Horvath-Watson tests (H_0: $\lambda_{1,i} = \lambda_{2,i} = 0$)

$$\Delta\ln(E_{i,t}) = \kappa_{1,i} + \lambda_{1,i}\mu_{i,t-1} + \sum_{j=1}^{s_i} \phi_{1,i,j}\Delta\ln(E_{i,t-j}) + \sum_{j=1}^{s_i} \psi_{1,i,j}\Delta\ln(r_{i,t-j}) + \nu_{1,i,t}$$

$$\Delta\ln(r_{i,t}) = \kappa_{2,i} + \lambda_{2,i}\mu_{i,t-1} + \sum_{j=1}^{s_i} \phi_{2,i,j}\Delta\ln(E_{i,t-j}) + \sum_{j=1}^{s_i} \psi_{2,i,j}\Delta\ln(r_{i,t-j}) + \nu_{2,i,t}$$

Country	Reference currency: US dollar		Reference currency: DM	
	$t_{\hat{\rho}-1}$(ADF)	LR(λ)	$t_{\hat{\rho}-1}$(ADF)	LR(λ)
United States			−1.562	7.694
Canada	** −3.605	3.591	−2.182	3.309
Japan	−1.762	5.355	−2.401	**22.272
Germany	−1.562	*9.265		
France	** −3.012	*8.776	−1.825	2.705
Italy	* −2.848	4.364	* −2.913	2.074
Great Britain	−0.675	5.152	−2.162	2.271
Belgium	−2.484	1.037	−0.453	3.627
Denmark	−2.244	**12.613	** −4.061	5.736
Sweden	* −2.960	2.913	−0.444	2.551
Finland	−1.959	6.208	* −2.466	6.163
Austria	−1.244	**11.391	* −2.639	1.994
Spain	* −2.875	**15.002	** −3.087	1.371
Panel unit root test with common time dummies				
All countries	** −2.382		* −1.775	
G-7 countries	** −5.319			
European countries			−1.565	

Statistical significance at the 95% level or greater is signified by **. Significance at the 90% level or greater is signified by *. The $T(\hat{\rho}-1)$ and $t_{\hat{\rho}}$(PP) statistics are tests proposed by Phillips and Perron and are computed setting $k_i = 0$. The $t_{\hat{\rho}}$(ADF) is the augmented Dickey-Fuller statistic, the t ratio. The panel unit root tests are proposed by Im et al. (1995) and are distributed as standard normal.

of the 12 countries. Not surprisingly, the panel test also points to a rejection of the null hypothesis that the slopes are jointly equal to one. We also find a large number of rejections of the null when we look at the cointegrating slopes for relative prices and relative productivities. But there the rejections arise because the slope estimates, which are fairly close to 1.0, are precisely estimated. In contrast, here the point estimates are generally far from 1.0 and we reject the null despite fairly

Table 6
Estimates of the cointegrating slope coefficient of $\ln(E_{i,t})$ and $\ln(r_{i,t})$

$$\ln(E_{i,t}) = \alpha_i + \beta_i \ln(r_{i,t}) + \epsilon_{i,t}$$

Country	Reference currency: US dollar			Reference currency: DM		
	$\hat{\beta}_{i,\mathrm{OLS}}$	$\hat{\beta}_{i,\mathrm{FMOLS}}$	$t(\hat{\beta}_{i,\mathrm{FMOLS}}=1)$	$\hat{\beta}_{i,\mathrm{OLS}}$	$\hat{\beta}_{i,FMOLS}$	$t(\hat{\beta}_{i,FMOLS}=1)$
United States				1.888	1.592	1.357
Canada	0.651	0.610	**−2.615	1.455	1.322	*1.874
Japan	1.657	1.521	1.912	1.416	1.169	0.707
Germany	1.888	1.592	1.357			
France	0.495	0.390	**−3.072	1.337	1.242	**2.703
Italy	0.868	0.973	−0.221	1.040	1.037	0.923
Great Britain	0.506	0.497	**−7.104	0.954	0.886	*−1.754
Belgium	0.422	0.116	**−2.857	−1.045	−0.563	−0.672
Denmark	−0.060	−0.032	**−3.341	1.058	1.043	*1.680
Sweden	0.727	0.734	−0.927	1.159	1.133	**4.229
Finland	0.322	0.308	**−6.328	0.861	0.850	**−3.822
Austria	2.118	1.912	**2.234	1.058	1.017	0.036
Spain	0.652	0.635	**−2.784	1.023	1.005	0.184
Panel tests of $\hat{\beta}=1$ with common time dummies						
	$t_{\hat{\beta}}$	$\sqrt{N}\bar{t}$	$t_{\hat{\beta}}$	$\sqrt{N}\bar{t}$		
All countries	**−2.948	**−4.587	−0.849	−0.815		
G-7 countries	**−5.671	**−10.296				
European countries			**−6.696	**−6.600		

The second and fifth columns contain the ordinary least squares estimates of the slope coefficient. Columns 3 and 6 contain the Phillips and Hansen (1990) fully modified OLS estimates of the slope, and columns 4 and 7 contain the *t*-ratio formed by subtracting one from the fully modified OLS estimate of the slope and dividing by the corresponding standard error. The numbers of lags used in computing the fully modified OLS estimate and its standard error are chosen using the data-dependent procedure proposed by Newey and West (1994). The panel tests are those proposed by Pedroni (1996).

large standard errors. The wide range of point estimates that we report in Table 6 is roughly consistent with previous work. Froot and Rogoff (1995) report that it is common in the literature to find slopes that vary widely across countries and are frequently far from the value of 1.0 implied by purchasing power parity.

Could these results also be due to the small-sample properties of the FMOLS estimates? Monte Carlo evidence suggests that it is not. We choose six countries, three with estimated slopes exceeding one (Japan, Germany, and Austria) and three with estimated slopes below one (France, Britain, and Finland). For each, we estimate a VAR for $\eta=(\epsilon,\zeta)'$ and use the parameters and error covariance matrix to generate 15 000 samples of 25 observations in which the true value of β is 1.0. Although we find evidence that the FMOLS estimates are biased, the bias is not sufficient to explain either the dispersion or the magnitudes of the estimated slopes in Table 6. In all six experiments the mean slope is below one, so that small sample problems cannot explain the estimated slopes that exceed one. The

FMOLS standard errors understate the dispersion of the slopes found in the experiments. Using the empirical distribution of slopes, we reject the null of a unit slope at the 95% level for two countries (Japan and France), at the 90% level for two countries (Germany and Austria) and at levels below 90% for the other two.

In contrast, when we use the DM as the reference currency, the point estimates are much closer to 1.0 than are those obtained using the dollar.[19] Even though we can reject the null hypothesis for some currencies, the point estimates are close to one. Interestingly, the point estimates do not indicate that purchasing power parity in traded goods is any more likely to hold for European currencies than for non-European currencies relative to the DM.

The results in Tables 4–6 lend somewhat favorable support for the hypothesis of purchasing power parity in traded goods, especially when the DM is the reference currency. Nominal exchange rates and PPP exchange rates appear to be co-integrated. Our estimates suggest that the slopes of the cointegrating relationships are far from the hypothesized value of unity when the dollar is used as the reference currency. But nominal and PPP exchange rates are nearly proportional when the DM is used as the reference currency.

In Fig. 3 we plot the (logs of the) nominal and PPP exchange rates for the DM

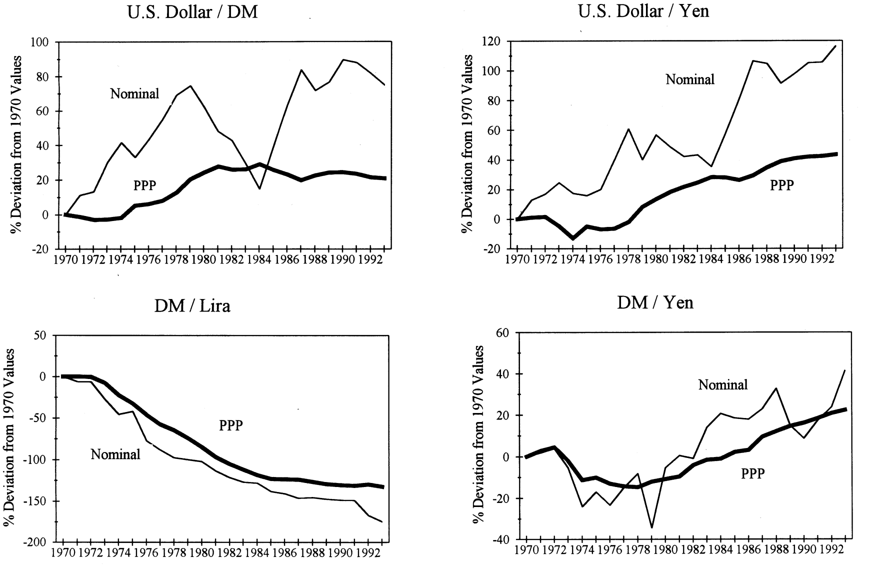

Fig. 3. Nominal and purchasing power parity exchange rates.

[19]The one notable exception is the Belgian franc regression where we obtain a negative point estimate and a large standard error. The slopes for the US dollar–DM exchange rate are, of course, identical.

and yen relative to the US dollar and the lira and yen relative to the DM. Again, we normalized so that 1970 = 0. Large and long-lasting deviations from PPP in traded goods are evident for both exchange rates relative to the US dollar. As is consistent with the formal tests, these deviations are smaller and less persistent when we examine the two exchange rates relative to the DM.

5. Concluding remarks

How well does the Balassa-Samuelson model explain the behavior of real exchange rates? The evidence from a panel of 13 OECD countries supports the hypothesis that the relative price of non-traded goods reflects the relative labor productivities in the traded and non-traded goods sectors. The results suggest that the relative prices of non-traded goods and the relative productivities in the traded and non-traded goods sectors are cointegrated and that the slope of the cointegrating relationship is generally close to 1.0. For some countries, growth in relative productivities appears to outstrip growth in relative prices (or equivalently the slope is less than 1.0), but Monte Carlo evidence suggests that this difference is within the range of sampling variation in a sample like ours. Thus relative prices and relative productivities appear to be proportional in the long run.

The Balassa-Samuelson model also assumes that traded goods prices are characterized by purchasing power parity. As is consistent with the evidence presented in Engel (1995), we find large and long-lived deviations from PPP in traded goods when we look at US dollar exchange rates. Although nominal exchange rates and PPP exchange rates appear to be cointegrated we find that the slopes of the cointegrating regressions vary widely and differ substantially from one. But when we examine DM exchange rates the evidence is considerably more favorable to purchasing power parity in traded goods. Nominal and PPP exchange rates appear to be cointegrated and the slopes of the cointegrating regressions are generally close to one.

Acknowledgements

Helpful discussions with José Vinals and generous assistance and numerous suggestions from Peter Pedroni along with the comments of seminar participants at Johns Hopkins University, Indiana University, North Carolina State University, the University of Maryland, and the University of Pennsylvania, the referees and the editor are acknowledged with thanks.

Appendix A

The data

Nominal exchange rates are from *International Financial Statistics* (line ae). For Belgium, Canada, Denmark, Great Britain, Finland, France, Germany, Italy, Japan, Sweden, and the US, sectoral price and productivity data come from the OECD *International Sectoral Database*, 1995. For Austria and Spain, they come from national statistics. (Francisco de Castro of the Bank of Spain collected and documented the data.) These sources provide annual data on nominal and real value added and number of employees. Sectoral prices are implicit deflators.

Traded goods consist of the 'manufacturing' and 'agriculture, hunting forestry and fishing' sectors. Non-traded goods consist of the 'wholesale and retail trade, restaurants and hotels', 'transport, storage and communication', 'finance, insurance, real estate and business services', 'community social and personal services', 'non-market services' sectors. 'Non-market services' include the 'producers of government services' and 'other producers' subsectors.

Data consistency is always an issue. We are aware of these anomalies in the OECD data: (1) the German market services employment figures do not include the 'real estate and business services' sector; value added figures do. (2) The Italian, British and Belgian value added and employment figures do not include 'real estate and business services' sector. (3) British value added and employment figures consist of the 'producers of government services' sector.

The productivity data allow us to begin in 1960 for three countries (the United States, Germany, and Finland), in 1961 for Canada, in 1964 for Spain, in 1967 for Denmark, and in 1970 for the remainder. The data end in 1993 for all countries except Canada, Austria, Spain (all 1992), Great Britain (1991), and Belgium (1990). Apart from Germany, where the post-unification data are problematic, we use all available data for estimation.

References

Asea, P.K., Mendoza, E.G., 1994. The Balassa-Samuelson model: a general equilibrium appraisal. Review of International Economics 2, 244–267.

Balassa, B., 1964. The purchasing power parity doctrine: a reappraisal. Journal of Political Economy 72, 584–596.

Chinn, M.D., Johnston, L., 1996. Real exchange rate levels, productivity and demand shocks: evidence from a panel of 14 countries. Working paper, Department of Economics, University of California, Santa Cruz.

Canzoneri, M.B., Cumby, R.E., Diba, B.T., 1996. Relative labor productivity and the real exchange rate in the long run: evidence for a panel of OECD countries. NBER Working Paper no. 5676.

De Gregorio, J., Giovannini, A., Krueger, T., 1994. The behavior of non-tradable goods prices in Europe: evidence and interpretation. Review of International Economics 2, 284–305.

De Gregorio, J., Giovannini, A., Wolf, H., 1993. International evidence on tradables and nontradables inflation. European Economic Review 38, 1225–1244.

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

Edison, H.J., Gagnon, J.E., Melick, W.R., 1997. Understanding the empirical literature on purchasing power parity: the post-Bretton Woods era. Journal of International Money and Finance 16, 1–17.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 55, 251–276.

Engel, C., 1995. Accounting for US real exchange rate changes. NBER working paper no. 5394.

Engel, C., Hendrickson, M.K., Rogers, J.H., 1997. Intra-national, intra-continental, and intra-planetary PPP. NBER working paper no. 6069.

Evans, C., 1992. Productivity shocks and real business cycles. Journal of Monetary Economics 29, 191–208.

Evans, M.D.D., Lewis, K.K., 1994. Do stationary risk premia explain it all? Evidence from the term structure. Journal of Monetary Economics 33, 285–318.

Frankel, J.A., Rose, A.K., 1996. A panel project on purchasing power parity: mean reversion within and between countries. Journal of International Economics 40, 209–224.

Frenkel, J.A., 1981. Flexible exchange rates, prices, and the role of 'news': lessons from the 1970s. Journal of Political Economy 89, 665–705.

Froot, K., Rogoff, K., 1995. Perspectives on PPP and long-run real exchange rates. In: Grossman, G., Rogoff, K. (Eds.), Handbook of International Economics, vol. 3. North-Holland, Amsterdam, pp. 1647–1688.

Hall, A., 1994. Testing for a unit root in time series with pretest data-based model selection. Journal of Business and Economic Statistics 12, 461–470.

Hakkio, C.S., 1984. A re-examination of purchasing power parity: a multi-country and multi-period study. Journal of International Economics 17, 265–277.

Horvath, M.T.K., Watson, M.W., 1995. Testing for cointegration when some of the cointegrating vectors are prespecified. Econometric Theory 11, 984–1014.

Hsieh, D.A., 1982. The determination of the real exchange rate: the productivity approach. Journal of International Economics 12, 355–362.

Im, K., Pesaran, M.H., Shin, Y., 1995. Testing for unit roots in heterogeneous panels. Working paper, Department of Applied Economics, University of Cambridge.

Jorion, P., Sweeney, R.D., 1996. Mean reversion in real exchange rates: evidence and implications for forecasting. Journal of International Money and Finance 15, 535–550.

Levin, A., Lin, C.F., 1992. Unit root tests in panel data: asymptotic and finite sample properties. Working paper, Department of Economics, University of California, San Diego.

Lothian, J., 1997. Multi-country evidence on the behavior of purchasing power parity under the current float. Journal of International Money and Finance 16, 19–35.

Marston, R., 1987. Real exchange rates and productivity growth in the United States and Japan. In: Arndt, S., Richardson, J.D. (Eds.), Real and Financial Linkages Among Open Economies. MIT Press, Cambridge.

Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.

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

O'Connell, P., 1997. The overvaluation of purchasing power parity. Journal of International Economics, in press.

Papell, D.H., 1997. Searching for stationarity: purchasing power parity under the current float. Journal of International Economics, in press.

Papell, D.H., Theodoridis, H., 1997. The choice of numeraire currency in panel tests of purchasing power parity. Working paper, Department of Economics, University of Houston.

Pedroni, P., 1995. Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. Working paper, Department of Economics, Indiana University.

Pedroni, P., 1996. Fully modified OLS for heterogeneous panels. Working paper, Department of Economics, Indiana University.

Phillips, P.C.B., Hansen, B.E., 1990. Statistical inference in instrumental variables regressions with I(1) processes. Review of Economic Studies 57, 99–125.

Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. Econometrica 58, 165–193.

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

Samuelson, P., 1964. Theoretical notes on trade problems. Review of Economics and Statistics 46, 145–154.

Wei, S.J., Parsley, D.C., 1995. Purchasing power disparity during the floating rate period: exchange rate volatility, trade barriers, and other culprits. Working paper, Kennedy School of Government.