Country comparisons and the real exchange rate

Oct 22, 2026

By Risto Herrala *

Abstract The paper contributes to the integration of the work on country comparisons and on the real exchange rate. It formalizes an equivalence between measurement discrepancies in the field of country comparisons and real exchange rate variants. It reviews the previous work and provides a set of stylized facts about the real exchange rate variant that is most relevant for GDP comparisons.

* Risto Herrala, Adviser, Bank of Finland Institute for Emerging Economies (BOFIT), Risto.Herrala@bof.fi.

DRAFT

I. Introduction

This paper is an attempt toward integration of the work on country comparisons and the real exchange rate. It is well known (ICP 2025; Deaton and Heston, 2010) that comparing the size of national economies is non-trivial and that the difference between the estimates given by the Purchasing Power Parity (PPP) -adjusted and the vehicle currency denominated GDP is often large. For example, in 2024 China exceeds the USA by 27 log-points in the PPP -adjusted GDP comparison and falls short of it by -44 log-points in the current USD denominated GDP comparison.[1] The discrepancy between the two relative GDP estimates, +71 log-points from the Chinese perspective, is so large that there is ambiguity about which of the two countries has the world's largest national economy.

In the field of country comparisons, the measurement discrepancy between PPP adjusted and vehicle currency denominated data is also called the inverse relative price level (Feenstra et al, 2015). Yet another name has been given to it in open economy macroeconomics, where it is called the real exchange rate. The +71 log-point GDP measurement discrepancy can therefore be equally referred to as the natural logarithm of the GDP price level of USA relative to China, and the real exchange rate between China and the USA calculated using GDP prices.

Even though the equivalence between inverse relative price level and the real exchange rate has been recognized (Feenstra et al, 2015), it is fair to say that the work on the unifying variable is not well integrated. For example, the term "real exchange rate" is not mentioned in the flagship reports on country comparisons (ICP, 2025; ICP, 2020), even in the parts where measurement discrepancies and relative price levels are in focus. The major review of the real exchange rate debate in open economy macroeconomics by Itskhoki (2020) does not discuss country

[1] This comparison uses data from IMF WEO April 2025.

comparisons. Such fragmentation risks compartmentalizing findings and making their interpretation dependent on the conventional assumptions of each field—signs of which we report evidence below.

The first unification challenge to be tackled is conceptual. In the field of country comparisons, the researcher is faced with various relative price levels and a range of relative prices of the many national account items. In contrast, the discussion on the real exchange rate focuses on only few variants, which are deemed important from a macroeconomic perspective. General statements like "the inverse relative price level is the real exchange rate" therefore seem inadequate.

To set the discussion on a conceptually sound footing, we formalize the link between real exchange rates and country comparisons. The formalism supports a simple terminology, where the real exchange variant calculated using the prices of any item equals the ratio of the real and vehicle currency denominated relative value (ie. the inverse relative price) of that item. Under this terminology, there are not only few but instead a large variety of real exchange rates, each calculated using a different price assumption, and therefore corresponding with a different comparison between two countries.

The work on country comparisons and on the real exchange rate overlaps mainly when it comes to variants of the unifying variable that build on macroeconomically important price level aggregates. A survey of this work reveals an almost universal inability to understand the unifying variable.

Since the early 1950's, when first version of the system of national accounts emerged, the main country comparison at the macroeconomic level has been the GDP comparison (IPC, 2025; Gilbert and Kravis, 1954). The discrepancy between real and vehicle currency denominated relative GDP has been recognized at least since the first PPP -adjusted comparisons by Gilber and Kravis (1954). Part of the discrepancy, namely the real GDP comparison, is considered well understood, whereas the vehicle currency nominated comparison is not well understood.

In exchange rate economics, the pioneering contributions by Cassel (1922, 1918, 1916) brought into focus the "abnormal" deviation of nominal exchange rates from PPP, which has become known as the real exchange rate. The mainstream view in the field (Engel, 2014) is that no theory can at present explain it.

In open economy macroeconomics, the theories of the real exchange rate by Balassa (1964) and Samuelson (1964) have been rejected in empirical studies. The current mainstream view (Itskhoki, 2020) is that the real exchange rate arises in general equilibrium, and that no partial theory can explain it.

The empirical focus differs across the three fields in terms of the variant of the unifying variable and its data frequency. The current focus is on the consumer prices -based variant in the real exchange rate debate, and on the GDP price -based variant in the field of country comparisons. In open economy macroeconomics the preferred data frequency is quarterly and in country comparisons it is annual.

In the empirical analysis, we take the perspective of country comparisons and use the survey of the real exchange rate debate by Itskhoki (2020) for comparison. Building on the latest Penn World Table (PWT 11), we study the unifying variable in annual frequency over the largest number of countries (185) and over the longest available time period (73 years).

The empirical analysis reveals significant differences relative to the findings reported by Itskhoki (2020). In our data, the weak PPP hypothesis, the status of which is controversial in the open economy marco debate, is soundly rejected. Two mainstream views, referred to as the Backus-Smith puzzle and the excess sensitivity of the real exchange rate to nominal exchange rates, mostly do not hold in this data.

As a novel empirical approach, we pioneer a decomposition of the unifying variable into the PPP -adjusted relative GDP and the vehicle currency denominated relative GDP component. We find that of the two components, the latter contributes by far more to variation of the unifying variable than the former. This finding calls for future efforts to explain nominal GDP variation.

The next section is devoted to a formal analysis of the link between the real exchange rate and country comparisons. A review of the literature, and an empirical study follows. We conclude with some discussion about the next steps.

II. Formalizing a link between country comparisons and the real exchange rate

A. The real exchange rate and GDP comparisons

The standard definition of the real exchange rate Q of country a toward b is:

(1) $Q_{a,b} \equiv E_{a,b} * \frac{P_b}{P_a}$

, where $E_{a,b}$ denotes the nominal exchange rate (the price of one unit of b's currency in terms of a's currency); and P_a and P_b the price levels in local currency terms.

We make the standard assumption (Obstfeld and Rogoff, 1996) that any nominal exchange rate can be viewed as chain via the "vehicle currency" of country v:

(2) $E_{a,b} = E_{a,v} * E_{v,b}$

Denote by Y real GDP and by $N_v Y_a \equiv E_{v,a} * P_a^Y * Y_a$ the nominal GDP of country a in the currency of country v, with the understanding that P_a^Y refers to GDP prices, and that $E_{a,a} = 1$. The links in the chain (2) can then be expressed as ratios of the local currency and vehicle currency denominated GDP:

(3) $E_{a,v} = \frac{N_a Y_a}{N_v Y_a};\ E_{v,b} = \frac{N_v Y_b}{N_b Y_b}$

Applying (3) to (2), inserting to (1), and rearranging yields an intermediate result:

(4) $Q_{a,b} = \frac{N_a Y_a}{N_v Y_a} * \frac{N_v Y_b}{N_b Y_b} * \frac{P_b}{P_a}$

$= \frac{N_a Y_a}{P_a} * \frac{P_b}{N_b Y_b} * \frac{N_v Y_b}{N_v Y_a}$

The main result applies to the real exchange rate variant that uses GDP prices ($P = P^Y$). Implementing this price level assumption in (4) yields:

$$(5)\ Q_{a,b}^Y = \frac{N_a Y_a}{P_a^Y} * \frac{P_b^Y}{N_b Y_b} * \frac{N_v Y_b}{N_v Y_a} = \frac{Y_a/Y_b}{N_v Y_a/N_v Y_b}$$

Eq. (5) establishes the equality of the real exchange rate variant that uses GDP prices, and the ratio of the real and vehicle currency denominated relative GDP. By (5) $Q_{a,b}^Y$ is larger/smaller than one if the real relative GDP is larger/smaller than the vehicle currency denominated relative GDP.

Notably, the equivalence (5) is unaffected, if the numerator and denominator of relative real GDP are both multiplied by P_v^Y. This operation replaces real GDP with the "PPP -adjusted GDP" $PPP_v^Y Y_a \equiv \frac{N_a Y_a}{PPP_{a,v}^Y} = Y_a P_v^Y$, where $PPP_{a,v}^Y$ is the "Purchasing Power Parity" towards the vehicle country calculated using GDP prices. We may therefore use the terms real and PPP -adjusted relative GDP interchangeably when discussing equation (5).

We summarize these findings for further reference:

Result 1 Assume that the chain rule (Eq. 2) holds. Then the real exchange rate variant calculated using GDP prices is equivalent to the ratio of the real and vehicle currency denominated relative GDP (Eq. 5). It is also equivalent to the ratio of the PPP -adjusted and vehicle currency denominated relative GDP, when the PPP adjustment is made using the GDP prices of the vehicle country.

While Result 1 is presented by Feenstra et al (2015), it has received little attention elsewhere. It is more common to formulate the link between country comparisons and the real exchange rate indirectly using the concept of the inverse relative price level. This link emerges simply from the definition of PPP, which implies that

$Q^Y_{a,b} = \frac{E_{a,b}}{PPP^Y_{a,v}}$, the right-hand side of which is the inverse relative GDP price level (Feenstra et al, 2015). Thet the equivalence between the real exchange rate and relative prices implies (5) can be shown using a similar argument that we use above.

We want to emphasize Result 1 instead, because it links the real exchange rate with the relative GDP comparison, the quintessential problem in the field of country comparisons. This is the problem which the PPPs are built to solve. Result 1 makes it explicit that the work on the real exchange rate variant that uses GDP prices in exchange rate economics and open economy macroeconomics, and the work on the GDP measurement problem in country comparisons are about a single unifying variable from two different perspectives.

B. Generalization

To integrate the work on the real exchange rate and on country comparisons, we equate variants of the real exchange with specific country comparisons. To this end, we show that Eq. (5) generalizes to the value comparison of any items Z that has a price P^Z in the vehicle country and in the two countries that are under comparison. Namely, the links of the chain (3) can be re-expressed in terms of that item ($E_{a,v} = \frac{N_a Z_a}{N_v Z_a}$, $E_{v,b} = \frac{N_v Z_b}{N_b Z_b}$). By applying these to (2) and inserting to (1), it is then found that the real exchange variant that uses P^Z is equivalent to the measurement discrepancy of relative Z:

$$(6)\ Q^Z_{a,b} = \frac{Z_a / Z_b}{N_v Z_a / N_v Z_b}$$

Result 2 Assume that the chain rule (Eq. 2) holds. Then the real exchange variant that uses the prices of item Z is equivalent to the ratio of real and vehicle currency denominated relative Z. It is also equivalent to the ratio of the PPP -adjusted and

vehicle currency denominated relative Z, when the PPP adjustment is made using the prices of Z in the vehicle country

For example, based on Result 2 the real exchange rate variant that uses consumer prices is equivalent to the measurement discrepancy of relative consumption expenditure. The real exchange rate variant that uses unit labor costs is equivalent to the measurement discrepancy of relative total labor costs. More generally, result 2 implies that apart from the widely used real exchange rate variants, there is a broad range of other variants, each of which corresponds with the measurement discrepancy of a different country comparison. Result 2 is anticipated by Itskhoki and Mukhin (2021), who link real exchange rate and relative price variants.

III. Revisiting the literature

A. The measurement problem of country comparisons

From the times of the earliest civilizations, governments have kept account of the stocks and flows of assets in government institutions (Premchand, 1995). Until the early modern period in Europe, accounts cannot be found at the national level. Arguably, the absence of national accounts during the early times does not reflect lack of skill - the first national accounts compiled in the early modern period were simple approximations that would seem to have been within reach during earlier times - but, rather, lack of interest.

In his authoritative review of country comparisons, Studentski (1958) views the Mercantilist studies in 16th and 17th century Europe as a foundational expansion in focus from government accounts toward national accounts. The Mercantilists considered that the national stock of gold was the primary measure of the strength of a nation and therefore made efforts to estimate and compare it across countries (Magnusson, 2015).

Studentski credits the emergence of the national income approach in the seventeenth century primarily to William Petty from England and Pierre le Pesant de Boisguilbert from France. In “Political Arithmetick” Petty (1677) numerically compares the wealth of various countries, associating it with military strength. The units of comparison are variable (people to people, acres to acres), with monetary comparisons based on conversion rates that are not completely transparent today. In “Verbum Sapienti” (1691), Petty motivates wealth accounting in national currency units by taxation, which is in those units. The approach can therefore guide the distribution of taxation toward those who have the most ability to pay and set the boundaries of how much can be taxed. He introduces national income accounting to study how the income from national wealth and labor are distributed in the economy to its different uses.

During the eighteenth and nineteenth centuries, the focus of national accounting shifted from wealth to income reflecting a shift in the analytical agenda from national strength to welfare analysis. As explained by Marshall (1890) ”The money income, or inflow, of wealth gives a measure of a nation's prosperity, which, untrustworthy as it is, is yet in some respects better than that afforded by the money value of its stock of wealth”. In the early 20th century, prominent contributions to national income measurement included Clark (1932) and Kuznets (1934). Kuznets et al (1941) promotes net national income, which is gross national income minus depreciation of capital, as the relevant welfare metric at the national level.

Another major shift in the perspective of national accounting occurred after the second World War (WWII), when the analytical agenda shifted from welfare back toward strength. At the time, the term Gross Domestic Product (GDP) entered professional vocabulary[2], and a standardized approach to measure GDP by a system

[2] The term appears, for example, in the Monthly Digests of Statistics of H.M. Stationery office in the UK in the late 1940’s.

of national accounts (SNA) was developed by the United Nations (UN, 1953). Presently, the SNA is under development in collaboration with various international organizations in the Intersecretariat Working Group on National Accounts (ISWGNA). Since GDP is an aggregate measure of production, which abstracts from depreciation and imports, it is primarily a measure of national productive strength rather than of economic welfare.

During the post WWII period, when the US was a globally dominant economic power, country comparisons often built on current United States Dollar (USD) denominated GDP, calculated by applying the nominal USD exchange rates to local currency denominated GDP (Gilbert and Kravis, 1954). However, this approach was recognized early on as problematic due to significant evidence that the purchasing power of the USD varied across countries.

Clark (1937)[3] and Gilbert and Kravis (1954) are among the early users of PPP -adjusted GDP in country comparisons. In such studies, nominal exchange rates are replaced by purchasing power parities, which are relative price levels towards some vehicle currency.[4] When the USA is used as the vehicle country, division by PPP transforms the local currency GDP to the US price level.

In 1968, the International Comparisons Program (ICP) was set up as a joint venture of the United Nations Statistical Division and the International Comparisons Unit of the University of Pennsylvania. At present, the program is managed by the World Bank under the UN Statistical Commission.[5] At irregular intervals of several years, The ICP generates comparable price level aggregates in participating countries and uses these to compute the PPP and PPP -adjusted GDP

[3] We have the second revised edition which came out in 1951

[4] For a formal definition of PPP, see the discussion around Eq. (7).

[5] World Bank, International Comparisons Project, https://www.worldbank.org/en/programs/icp/history.

data. The number of countries participating in the ICP rounds has increased from 10 in 1970 to 176 countries during the latest round in 2021.

In a pioneering study, Summers, Kravis and Heston (1981) extrapolated the PPP data to annual frequency and to non-participating countries to generate the Penn World Tables (PWT). These are regularly updated, initially by the University of Pennsylvania, and currently by the University of California and the University of Groningen. Various UN affiliated institutions such as the IMF and the World Bank also publish continuous time series of PPP -adjusted GDP at annual frequency.

Many empirical studies during the past and present century have demonstrated that the PPP adjusted and USD denominated GDP comparisons differ systematically: the latter tends to emphasize the GDP of high-income countries, and the former that of low-income countries. The prevailing interpretation of this finding, as summarized by Summers and Heston (1991), is that it is unmistakable evidence against the strong PPP doctrine. The literature therefore shows clear awareness of the relevance of the PPP debate in exchange rate economics (which we discuss below) to country comparisons.

The equivalence between inverse relative price level and the real exchange rate is well known. For example, the Result 1 is discussed by Feenstra et al. (2015). Recent work connecting real exchange rates and relative price levels include Couharde et al (2021), and Bergin et al (2024). Both make use of the equivalence between real exchange rates and relative price levels in their work. Anticipating Result 2, Itskhoki and Mukhin (2021) map real exchange rate variants with corrsponding relative prices.

In general, the work on country comparisons seems somewhat distant from the debate on the real exchange rate. For example, in the most recent flagship reports on country comparisons (ICP, 2025, 2020) the real exchange rate is not mentioned even in instances when the relation of PPP -adjusted and the nominal exchange rate -based GDP comparisons are discussed.

B. Deviations from the PPP hypothesis in exchange rate economics

In his pioneering article, Cassel (1916) proposes that the nominal exchange rate reflects the relative price level between the two countries:

(7) $E_{a,b} = \frac{P_a}{P_b}$

In 1918, he names the relative price level on the right of (7) "the Purchasing Power Parity" and notes that certain nominal exchange rates had deviated from it. He considered the deviation abnormal as, in his view, the parity should be maintained if there is "anything like" trade and free movement of goods between the two countries. In his 1922 book, Cassel promotes the view that differing price levels would cause trade imbalances that are unsustainable in the long run. Proposition (7) is known as the strong PPP hypothesis, and the proposition of long-term convergence toward (7) as the weak PPP hypothesis.

Cassel's analysis of the channels through which the nominal exchange rate reaches the PPP is somewhat opaque. Keynes (1930, 1923) noted that, in principle, spatial (cross-border) competition at the goods markets could lead to the prices of goods, expressed in a common currency, being equalized, thereby implementing (7). The equality of the price of a good across borders is referred to as the Law of One Price (Rogoff, 1996). Assuming, firstly that LOP applies to all goods, and secondly that the baskets of goods are similar in the two countries, we get:

(8) $P_b E_{a,b} = P_a$

which rearranges to (7). The "Law of One Prile Level" (8) is temptingly simple from a modelling perspective. However, as recognized early on (Samuelson, 1964; Keynes, 1930, 1923) it relies on the strong assumptions outlined above.

Although the PPP hypothesis is still used in modelling, its popularity is somewhat diminished by the weak theoretical foundation, and large persistent empirical deviations that are not difficult to spot (Engel, 2014; Rogoff, 1996; Dornbush,

1988; Samuelson, 1964). Presently, the mainstream theory of exchange rates is the interest parity theory:

(9) $E_{a,b} = \frac{i_a}{i_b} F\left[E_{a,b}\right]$

where i are the interest rates, and F is the forward or expected value. Equation (9) has been known at least since the early past century as a financial arbitrage condition, where F is the forward exchange rate (Keynes, 1923). Under the rational expectations approach, F is the expected value of the nominal exchange rate (Dornbush, 1976). Based on that approach, nominal exchange rates depend on the structure of the economy via the interest rates and the expectation channels. The fields of exchange rate economics and open economy macroeconomics are therefore tightly integrated under this approach.

The interest rate parity theory does not necessarily replace PPP theory but rather complements it by providing a second potential mechanism, financial arbitrage, through which exchange rates are influenced. Whether any form of the PPP hypothesis holds, and how exchange rates deviate from it, depends on the macroeconomic structure. For example, if there is perfect cross-border arbitrage at the goods markets so that LOP brings about (7), then (9) provides a transmission channel of the effects of this arbitrage via the financial market. In case LOP does not hold, then exchange rates are determined by other factors via (9).

Notwithstanding the discoveries of (9) and (7), it is widely considered that the explanatory power of nominal exchange rate theory is unsatisfactory. Echoing the earlier critical assessment by Dornbush (1988), Engel (2014) concludes after his extensive survey that no theory can explain exchange rate movements.

In the absence of theoretical clarity, there has been significant empirical work documenting deviations of nominal exchange rates from PPP. Surveying the empirical literature, Rogoff (1996) finds that the strong PPP hypothesis is overwhelmingly rejected in micro and macroeconomic data. In his view the early

empirical studies failed to find convergence toward PPP, but the more recent research was consistent with convergence, with deviations having a half- life of about 3 years. Rogoff's optimism about the weak PPP hypothesis remains controversial, however. In his survey, Itskhoki (2020) concludes that there is no conclusive evidence of PPP convergence.

C. The real exchange rate in open economy macroeconomics

In mainstream Keynesian open-economy theory (Meade, 1951; Mundell, 1960; Fleming, 1961) PPP is not imposed, and convergence toward PPP is not assumed. Deviations from PPP are brought into focus by Balassa (1964) and Samuelson (1964), who explicitly link them with trade frictions. They propose that LOP applies to goods that are tradeable across countries, whereas the prices of non-tradables vary. Since tradeable-goods prices are tied internationally, higher tradable productivity translates into higher wages. Labor mobility transmits higher wages to the non-tradable sector, which implies higher prices of non-tradable goods. The theory therefore predicts that the real exchange rate is higher in high productivity countries than in low productivity ones.

The Balassa-Samuelson theory is the first step in understanding the real exchange rate from an open economy macro perspective. The theory makes certain qualitative predictions, which have shown promise in empirical studies. Given that openness to trade can have deep institutional roots, the theory correctly predicts that deviations from the strong PPP hypothesis may be long standing. The theory also predicts that the price level tends to be higher in high productivity – high income countries relative to low productivity – low-income countries. This regularity has been confirmed in many studies.

Subsequent empirical work has, however, rejected various key predictions of the Balassa-Samuelson theory. Expanding the theory in a dynamic multi-country

setting, Backus and Smith (1993) arrive at the prediction that the real exchange rate is positively correlated with relative consumption. For example, if the real exchange rate between two countries is very variable, then the ratio of consumption between them should also be. However, they find little evidence of such correlation. The finding that the real exchange rate is only weakly correlated with relative real variables is known as the Backus-Smith Puzzle.

Further evidence against the Balassa-Samuelson theory is provided by Engel (1999) based on microeconomic data. He tests the theory by studying whether real exchange rate changes between the USA and other countries reflect relative price changes of non-tradables as suggested by the theory. He finds that relative prices of nontraded goods appear to account for almost none of the movement. This result undermines the central assumption in Balassa-Samuelson theory that LOP applies for tradeable goods.

Parallel to the work on the Balassa-Samuelson theory, the evidence against the PPP hypothesis prompted research to find other mechanisms of exchange rate formation in open economy macro models. In his influential paper, Dornbush (1976) introduces the interest parity condition as an important structural equation. The term "real exchange rate" seems to have appeared in open economy macroeconomics around this time. For example, Dornbusch does not use the term in his 1976 paper, but he does so along with the formal definition in his 1980 book.

The application of exchange rate assumptions in open economy models has been variable. For example, the PPP hypothesis holds and the real exchange rate is therefore unity by construction in real business cycle models such as Backus et al (1994); in the sticky-price model by Obstfeld and Rogoff (1995); and the New-Keynesian open economy model by Clarida et al (2002).

Other modelers attempt to achieve more realistic exchange rate behavior by imposing significant trade and financial frictions, often at the cost of model tractability. For example, Obstfeld and Rogoff (2002) restrict international real and

financial trade; Devereux and Engel (2002) restrict pricing and marketing of goods, and international arbitrage; and Itskhoki and Mukhin (2025) present a model where the segmentation of the global financial market is a key contributing factor.

All in all, the open economy macro literature has not been able to converge on a quantitative model of the real exchange rate. The PPP hypothesis has been rejected, and the main theories about PPP deviations by Balassa (1964) and Samuelson (1964) have been shown to lack explanatory power in empirical studies. Itskhoki (2020) summarizes the current understanding that the real exchange rate is inherently a general-equilibrium variable, which depends on the full model structure and policy regime, and therefore partial theories like the PPP theory are insufficient to explain it.

IV. Empirical analysis

A. The data

The empirical analysis builds primarily on the latest Penn World Tables version 11.0 (PWT 11). The sample period is 1950-2023 (73 years). The number of countries is 185, and the number of country pairs is therefore 34040 (=185*184). PWT 11 contains the variable "pl gdpo", which gives the real exchange rates, calculated in terms of output-side GDP prices, of the USA toward other countries (ie. $\text{pl gdpo}_a = \text{Q}^Y_{USA,a}$). Using the chain rule, we compute for all country pairs the complete set of the unifying variable in natural logarithms[6]:

$$(10)\ q^Y_{a,b} = -ln(pl\ gdpo_a) + ln(pl\ gdpo_b)$$

[6] To confirm that the baseline calculation method (10) is irrelevant we also computed the real exchange rate based on Eq. (5) using the PPP adjusted and nominal relative GDP. The correlation of that estimate with (10) is perfect.

Calculation (10) yields about 1.8 million observations of the unifying variable (Table 1). Each country has a maximum of 184 observations per year over the 73 - year period, which gives a maximum of 13432 observations per country and a maximum of about 2.5 million observations overall. The number of missing values (the difference between the maximum 2.5 mil and the actual 1.8 mil) is therefore about 0.7 million. These are mostly found during the earlier period and involving the poorer countries.

Definition (1) implies that $q_{a,b} = -q_{b,a}$, which explains why the sample mean of the unifying variable is zero and the distribution is perfectly symmetric (Table 1). Throughout the analysis, we include in the sample both $q^Y_{a,b}$ and the reciprocal $q^Y_{b,a}$. While possibly inefficient, this sampling approach does not otherwise disturb the analysis.

The standard deviation, minima and maxima of the unifying variable show very large variation. Transforming the logarithmic units into real units, the unifying variable is about 0.0001 at minimum and almost 9000 at maximum. By Result 1, the real GDP comparison exceeds/falls short of the vehicle currency denominated comparison by these factors at the extreme tails. A relative GDP measurement discrepancy of this scale is macroeconomically very large: it is roughly equal to the GDP gap in nominal USD terms in 2023 between the US and Suriname.

It is problematic to say the least that the two main ways to size up national economies can differ by this much for reasons that are not well understood. Large measurement discrepancies cannot only be attributed to outliers. For example, in almost one third of the cases, the unifying variable exceeds 2 or falls below 0.5 which implies that the discrepancy in the real and PPP adjusted country comparison exceeds 100%.

TABLE 1—SUMMARY STATISTICS

Variable	Obs	Mean	Std. Dev.	Min	Max

q^Y	1,818,597	0	.7334337	-9.099904	9.099904

Notes: q^Y is the natural logarithm of the real exchange rate computed using (10). Obs: number of observations; Std.Dev.=standard deviation; Sample period 1950-2023; Data Source: PWT 11.

B. The PPP hypothesis

The **strong PPP hypothesis** is rejected as a general theory by the large spread of real exchange rate values in the data. Random noise is ruled out as the sole explanation for the rejection by the finding that the country means show substantial and statistically significant deviations from zero: Based on the t-test, the hypothesis that q is on average zero, is rejected in almost all countries at the 1% significance level (the country-level findings are not tabulated). The point estimates of country-level averages are in the interval $(-2, +2)$ and their distribution seems somewhat asymmetric (Fiqure 1).

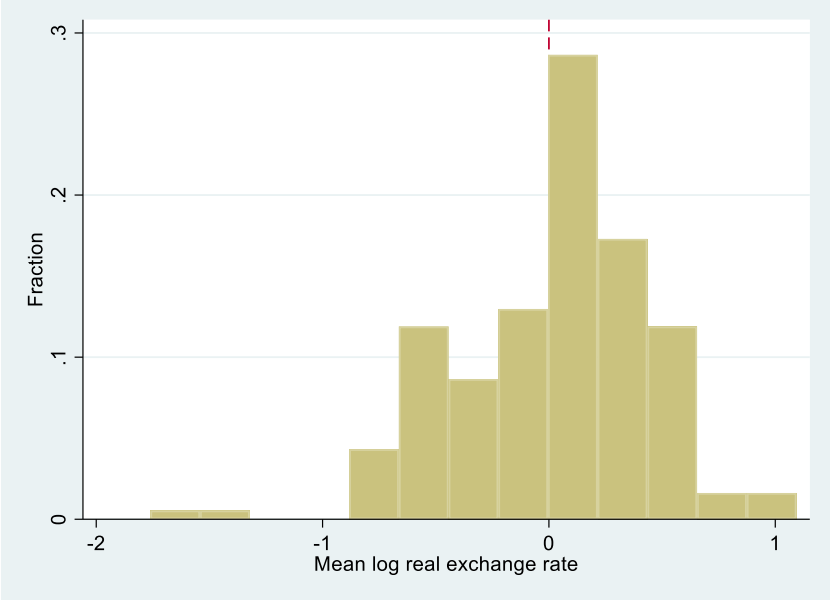


FIGURE 1. A HISTOGRAM OF COUNTRY LEVEL AVERAGES

Notes: The chart shows a histogram of the average q at country level.

The **weak PPP hypothesis** is also rejected as a general theory in this data. To arrive at this conclusion, we first visually inspect selected cases involving countries of different size and level of development (Figure 2) and then perform statistical tests in the full data sample.

Figure 2 shows four cases:

- $q_{USA,GBR}$: The graph shows an increasing trend for decades from negative toward positive values, until the unifying variable seemingly stabilizes around zero, only to turn back downwards the negative territory during the final decade. If a stable level exists, it seems completely unclear where that level is. Convergence to a hypothetical stable level is very slow.
- $q_{IND,CHN}$: The graph remains broadly stable for three decades, followed by an off-the-cliff drop which broadly coincides with the start of China's economic reform and opening, followed by a slow climb back to the old parity. In this case, the stable level, if it exists, does not seem to be at zero as would be required by the weak PPP hypothesis but perhaps somewhere between 0.5 and 1.
- $q_{BRA,SWE}$: The graph shows an explosive jump from around zero toward unity in the early sample, followed by a slow decline toward 0.5. Again, if a stable level exists, it is not at zero and convergence is very slow.
- $q_{DEU,FRA}$: The visual looks somewhat promising from the pow of the weak PPP hypothesis, albeit with very slow convergence.

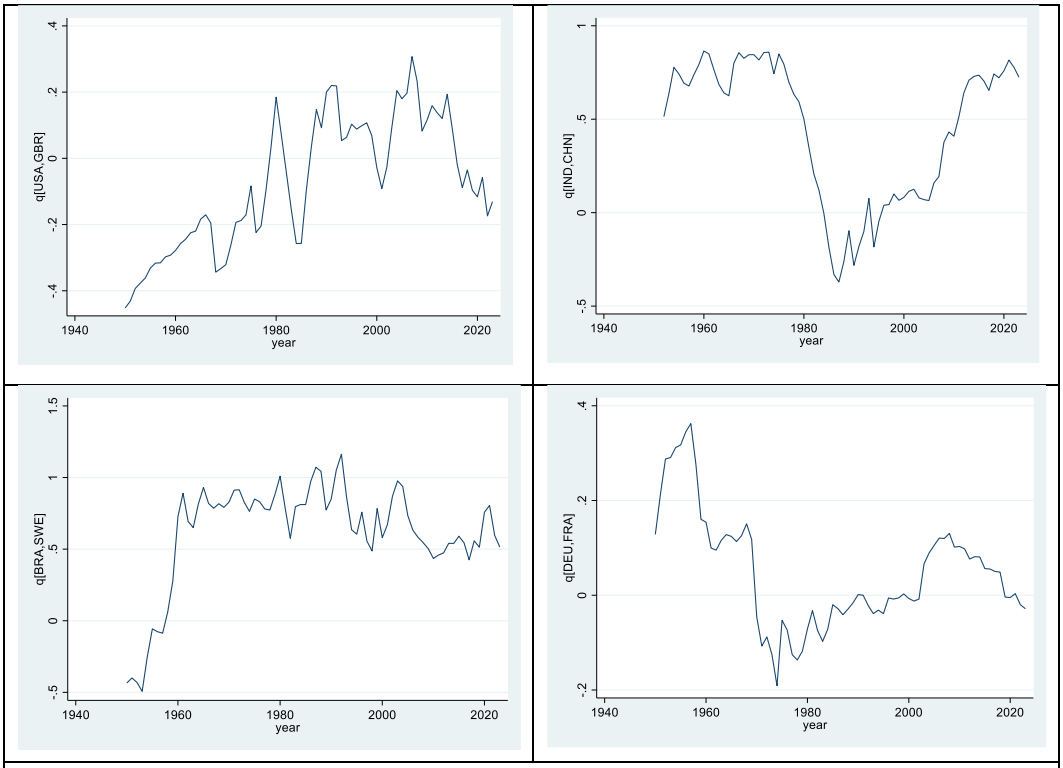


FIGURE 2. TIME SERIES OF UNIFYING VARIABLE IN SELECTED COUNTRY PAIRS (IN NATURAL LOGS)

Notes: USA= United States of America; GBR=Great Britain; IND=India; CHN=China (mainland); BRA=Brazil; SWE=Sweden; DEU=Germany; FRA=France.

In summary, rather than the convergence to zero predicted by the weak PPP hypothesis, the visuals indicate idiosyncratic dynamics. To strengthen this conclusion, we test the weak PPP hypothesis in the 34040 country pairs based on the ADF model with one lag:

(11) $\Delta q^{Y}_{a,b,t} = \alpha + \beta q^{Y}_{a,b,t-1} + \gamma \Delta q^{Y}_{a,b,t-1} + \varepsilon_t$

, where (α, β, γ) are estimated parameters and ε is a residual. Convergence of q to some stable level implies that $\beta = 0$, ie that q does not have a unit root. Thirty percent (10174) of the country pairs pass the unit root test at the 10% significance level. Based on the t -test of the constant term ($\alpha = 0$), less than a third (2796) of the country pairs that pass the test for the absence of a unit root show convergence

toward zero. Jointly, the unit root test and the zero mean test are consistent with the weak PPP hypothesis in 8 % of the country pairs.

C. The Balassa-Samuelson effect

The Balassa-Samuelson theory predicts that the unifying variable is higher in high-productivity countries than in low productivity countries. Given that higher productivity at country level implies higher real GDP per capita, the theory predicts a positive relationship between the unifying variable and real GDP per capita. This prediction, which is sometimes also called the Penn-effect, accords well with observations in the country comparisons literature (ICP, 2025; Feenstra et al, 2015; Kravis et al 1981).

To test this prediction in PWT 11, we study the relationship between $q_{USA,b}$ and $gdppc_b$, the real GDP per capita[7] of its country pairs. In each annual cross section, we estimate by OLS the regression model:

(12) $q^{Y}_{USA,b} = \alpha + \beta * ln(gdppc_b) + \varepsilon$

, where (α, β) are parameters and ε is a residual. The predicted positive relationship between the unifying variable and GDP per capita is assessed by the sign and statistical significance of the parameter β.

During most but not all years, the estimations yield the predicted positive, statistically significant β estimate (Figure 3). The exception is the 1950's, during which a positive relationship is absent. The predicted positive point estimate appears for the first time in 1959, after which it persists in all annual cross sections, and becomes consistently statistically significant in the 1970's. The parameter shows an increasing trend until the turn of the millennium, and a decreasing trend thereafter.

[7] This is calculated by dividing the real GDP variable in PWT (rgdpo) by population (pop) and then taking the natural logarithm.

We therefore conclude that the Balassa Samuelson theory fails as a general theory in this test. While the predicted empirical regularity is present during most years, the test fails during the early sample period. This overall result agrees well with the earlier findings by Bergin et al (2006).

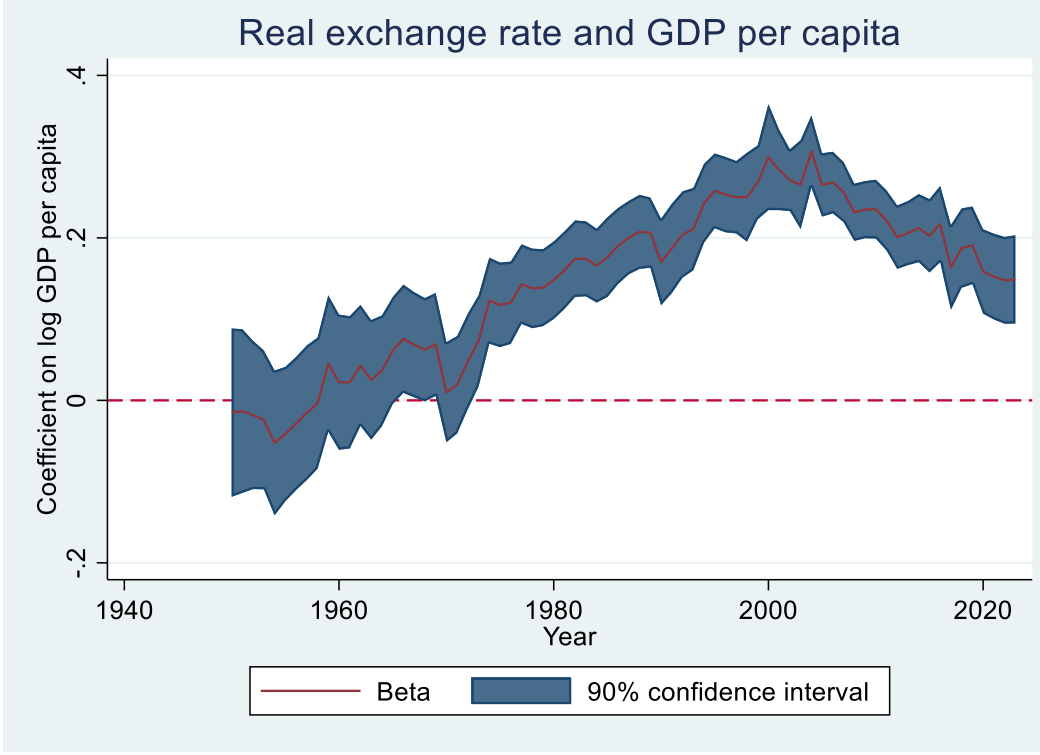


FIGURE 3. THE POINT ESTIMATE OF β AND THE 90% CONFIDENCE INTERVAL.

Notes: The chart shows a line connecting the point estimates and the confidence interval of (11), estimated in the annual cross sections. The confidence interval is approximate, based on $\beta \pm 1.645 * std.dev.(\beta)$*, where stad.dev. is the standard deviation*

D. The Backus-Smith puzzle

Backus and Smith (1993) predict using a dynamic model of an exchange economy with traded and non-traded goods, that there should be a correlation between changes in the unifying variable and changes in relative real consumption.

Yet, empirical studies have found little evidence of such correlation or, indeed, any correlation between changes in the unifying variable and changes in real macroeconomic variables (Itskhoki, 2020).

To test the Backus-Smith -puzzle, we investigate the empirical relationship between changes in the unifying variable and changes in two relative real variables: real consumption and real GDP per capita. For all country pairs with sufficient data, we estimate the two regression models:

(13) $\Delta q^{Y}_{a,b,t} = \alpha + \beta\Delta\left[\ln(consumption_{a.t}) - \ln\left(consumption_{b,t}\right)\right] + \varepsilon_t$

(14) $\Delta q^{Y}_{a,b,t} = \alpha + \beta\Delta\left[\ln(gdppc_{a.t}) - \ln\left(gdppc_{b,t}\right)\right] + \varepsilon_t$

, where Δ indicates the annual change, (α, β) are estimated parameters and ε is a residual. The t- test of parameter β for the country pair a, b indicates in model (13) whether changes in relative real consumption expenditure and changes in the unifying variable are correlated, and in model (14) whether relative real GDP per capita and changes in the unifying variable are correlated.

Of all country pairs:

- Eq (13): 24301 had enough observations for the estimations. Of those, a quarter (6066) show a statistically significant (at 10% level) β. Of those, about a quarter (1454) show a significantly positive estimate.
- Eq (14): Two thirds (23146) show a statistically significant (at 10% level) β, Of those, 96% (22230) show are positive point estimate.

We therefore conclude that the Backus-Smith puzzle, that changes in the unifying variable and changes in relative real macroeconomic variables are only weakly correlated, is not generally supported in this data. Changes in the unifying variable are in some cases significantly correlated with relative real consumption expenditure changes and in most cases strongly positively correlated with relative real GDP per capita changes.

E. Decomposition analysis

Using variance decomposition analysis, Mussa (1986) found that, for selected countries that apply a flexible exchange rate regime, the unifying variable is highly correlated with the nominal exchange rate. This view remains widely held in the open economy macro literature (Itskhoki, 2020). The

Following Mussa, and building on (1), we compute a variance decomposition:

$$(15)\ var\left(\Delta q_{a,b}^{Y}\right) = var\left(\Delta E_{a,b}\right) + var\left(\Delta \frac{PGDP_a}{PGDP_b}\right) + 2 * cov\left(\Delta E_{a,b}, \Delta \frac{PGDP_a}{PGDP_b}\right)$$

The main difference of our analysis with Mussa's (1986) analysis is that we use relative GDP prices instead of consumer prices to compute the unifying variable and the relative price term. We also study a much larger set of countries than he did over a much longer period.

Apart from analyzing the traditional decomposition, we build on Eq. (5) to pioneer another decomposition, which links the variance of the unifying variable with the relative GDP components:

$$(16)\ var\left(\Delta q_{a,b}^{Y}\right) = var\left(\Delta \frac{PPPY_a}{PPPY_b}\right) + var\left(\Delta \frac{N_c Y_a}{N_c Y_b}\right) + 2 * cov\left(\Delta \frac{PPPY_a}{PPPY_b}, \Delta \frac{N_c Y_a}{N_c Y_b}\right)$$

As regards the traditional decomposition (15), it is observed from the second column, rows 4-6 of Table 2 that, in the full sample of observations, the contributions to the variance of the unifying variable of nominal exchange rate variance and relative price level variance are 0.15 and 0.13 respectively. The variances of the two components are therefore not that far from each other, which means that nominal exchange rate variation and relative GDP price level variation contribute in broad terms similarly to variation in the unifying variable in this sample. The largest contributor is the covariance term at -0.24.

These findings imply that nominal exchange rates and relative prices tend to vary much more than the unifying variable because the negative covariance of the two variables mutes the combined impact of their variation. Under the PPP hypothesis, the covariance term would perfectly counter the contributions of the two variances, which does not hold in this data.

The third and fourth columns of Table 2 show the sub-period 1999-2023, during which it is possible to differentiate between floating exchange rate countries ("floater") and other countries. A comparison of the third and fourth columns indicates that the variance of the unifying variable was similar among pairs of floaters than among other country pairs. The contribution of nominal exchange rate variance is larger than that of relative price level variance in both groups, and it is not dominant in either group. In both groups, the negative covariance term is in absolute terms the largest contributor to the variance of the unifying variable.

All in all, the view that the unifying variable is closely correlated with nominal exchange rates, does not hold generally in this data. While nominal exchange rate variance contributes more than relative price level variance to changes in the unifying variable, the two contributions are not that different. In all cases, the large negative co-variation between nominal exchange rates and relative prices dampens the variance of the unifying variable.

TABLE 2—VARIANCE DECOMPOSITIONS

	all country pairs	floater pairs	other pairs
$var(\Delta q^{Y}_{a,b})$	0.04	0.02	0.02
Decomposition (15)			
$var(\Delta E_{a,b})$	0.15	0.06	0.06
$var\left(\Delta \frac{PGDP_a}{PGDP_b}\right)$	0.13	0.04	0.05
$2 * cov\left(\Delta E_{a,b}, \Delta \frac{PGDP_a}{PGDP_b}\right)$	-0.24	-0.08	-0.10
Decomposition (16)			

$avar\left(\Delta\frac{PPPY_a}{PPPY_b}\right)$	0.01	0.00	0.01
$var\left(\Delta\frac{N_cY_a}{N_cY_b}\right)$	0.04	0.02	0.02
$2*cov\left(\Delta\frac{PPPY_a}{PPPY_b},\Delta\frac{N_cY_a}{N_cY_b}\right)$	0.01	0.01	0.01
Sample period	1950-2023	1999-2023	1999-2023
Obs	1784372	27823	846811

Notes: Obs indicates the number of observations. "floater pairs" are those country-year pairs in which both countries are classified as "Independent floating" (before 2008) or as "Free floating" (after 2007) in the IMF AREAER database. "other pairs" indicates all other country pairs Data Sources: PWT 11, IMF AREAER.

Rows 7-10 of Table 2 quantify the novel decomposition (16), which links the variance of the unifying variable with the variances of PPP adjusted and vehicle currency denominated relative GDP. We observe that the variance of relative vehicle currency denominated GDP is markedly larger than the variance of the PPP -adjusted relative GDP and the correlation term. The correlation term is positive in this case: the co-movement of the PPP -adjusted and nominal relative GDP amplify changes in the unifying variable.

F. Summary of empirical findings

All in all, the empirical analysis leads to a list of rejections of previous theories and commonly held views about the unifying variable (Table 3). The rejection of the strong PPP hypothesis is very much in line with previous findings. However, there has not been a consensus about the status of the weak PPP hypothesis.

TABLE 3—SUMMARY STATISTICS

Hypothesis	Test result	Description
Strong PPP hypothesis	Rejected	Large systematic deviations observed.
Weak PPP hypothesis	Rejected	For most countries, conditions needed for convergence not found.

Balassa Samuelson theory	Rejected	For most but not all years, the prediction positive correlation found.
The Backus-Smith puzzle	Rejected	Significant correlation of real exchange rate changes with relative real consumption changes in some cases, and with relative real GDP per capita changes in most cases.
High correlation between unifying variable and nominal exchange rate	Rejected	Nominal exchange rate and relative price level variance almost equally contribute to real exchange rate variance.

Notes: q is the natural logarithm of the unifying variable computed using (10). Obs: number of observations; Std.Dev.=standard deviation; Sample period 1950-2023; Data Source: PWT 11.

The rejection of the Balassa-Samuelson theory during the early sample is consistent with the earlier findings by Bergin et al (2006). The failures of the Backus-Smith puzzle and the hypothesis that the unifying variable is highly correlated with the nominal exchange rate contrast with previous empirical findings. We find that while these assumptions may hold in some cases, in this data they are in most cases inaccurate.

V. Discussion and concluding remarks

The formal analysis supports integration by establishing a conceptual framework where real exchange rate variants correspond with comparisons of items. The mapping unlocks a broad range of integrated problems in the fields of country comparisons, exchange rate economics, and open economy macroeconomics. Itskhoki and Mukhin's (2021) efforts demonstrate how exploiting this range may help build more realistic macroeconomic models.

The literature review reveals just how little is really known about the unifying variable at the macroeconomic level, despite decades and even centuries of inquiry. This message is reinforced by the shocking empirical findings. The spread of the real exchange rate is huge in macroeconomic terms for reasons that are not known. The idiosyncratic dynamics (Fiqs. 2 and 3) of the unifying variable challenges the

assumption of convergence and stability, which underscores much of macroeconomic analysis. The series are characterized by long cycles, which are in many cases unstable in levels and in first differences.

The main gain from integration is therefore not the spread of information about the unifying variable. Rather, the hope is that the broadening of the perspective brought about by integration could inspire a more fruitful century of research.

D. References

Backus D, Kehoe P, Kydland F (1994) Dynamics of the trade balance and terms of trade: The J -curve, American Economic Review, Vol 84, No 1, 84-103.

Backus D, Smith G (1993) Consumption and real exchange rates in dynamic economies with non-traded goods, Journal of International Economics 35, 297-316.

Bergin P, Glick R, Taylor A (2006) Productivity, tradability, and the long-run price puzzle, Journal of Monetary Economics, Vol 53, Issue 8, 2041-2066.

Balassa B (1964) The Purchasing Power Parity doctrine: A reappraisal, Journal of political Economy, Vol 72, No 6, 584-596.

Bergin P, Choi W, Pyun J (2023) Catching up by "Deglobalizing": capital account policy and economic growth, NBER 30944.

Cassel G (1916) The present situation of the foreign exchanges, Economic Journal, Vol 26, No 101, 62-65.

Cassel G (1918) Abnormal deviations in international exchanges, Economic Journal, Vol 28, No 112, 413-415.

Cassel G (1922) Das Geldwesen nach 1914, G.A. Glocckner, Verlagsbuchhandlung in Leipzig.

Clarida R, Gali J, Getler M (2002) A simple framework for international monetary policy analysis, Journal of Monetary Economics 49, 879-904.

Clark C (1932) The national income and the net output of industry, Journal of the Royal statistical Society, Vol 96, No 4, 651-659.

Clark C (1937) National income and outlay, McMillan & Co, London.

Couharde C, Grekou C, Mignon V (2021) MULTIPRIL, a new database on multilateral price levels and currency misalignments, International Economics, Vol 165, 94-117.

Deaton A, Heston A (2010) Understanding PPPs and PPP based national accounts, American Economic Journal: Macroeconomics, 2:4, 1-35.

Devereux M, Engel C (2002) Exchange rate pass-through, exchange rate volatility, and exchange rate disconnect, Jounal of Monetary Economics 49, 913-940.

Dornbusch R (1976) Expectations and exchange rate dynamics, The Journal of Political Economy, Vol 84, Issue 6, 1161-1176.

Dornbush R (1988) Real exchange rates and macroeconomics: a selective survey, NBER Working Paper 2775.

Engel C (1999) Accounting for U.S. real exchange rate changes, Journal of Political Economy, Vol 107, No 3, 507-538.

Engel C (2014) Exchange rates and interest parity, Handbook of international economics, Volume 4, Chapter 8, 453-521.

Feenstra R, Inklaar R, Timmer M (2015) The next generation of the Penn World Table, American Economic Review, 105(10), 3150-3182.

Fleming M (1961) Domestic financial policies under fixed and under floating rates, The international Monetary Fund staff papers, 369-380.

Gilbert M, Kravis I (1954) An international comparison of the national products and the purchasing power of currencies: a study of the United States, The United Kingdom, France, Germany, and Italy, Organisation for European Economic Co-operation.

ICP (2025) Purchasing Power Parities and the size of world economies: results from the international comparison program 2021.

ICP (2020) Purchasing Power Parities and the size of world economies: results from the international comparison program 2017.

Itskhoki O (2020) The story of the real exchange rate, NBER working paper 28225.

Itskhoki O, Mukhin D (2021) Exchange rate disconnect in general equilibrium, Journal of Political Economy, vol 129, no 8, 2183-2232.

Itskhoki O, Mukhin D (2025) Mussa puzzle redux, Econometrica, Vol 93, No 1, 1-39.

Keynes J (1923) A tract on monetary reform, McMillan and Co., Limited, London.

Keynes J (1930) A Treatise on money Volume I, McMillan and Co. Limited, London

Kravis I, Heston A, Summers R (1981) New insights into the structure of the world economy, The review of income and wealth, Vol 27, Issue 4, 339-355.

Kravis I, Lipsey R (1986) The assessment of national price levels, NBER Working Paper No 1912.

Kuznets S (1934) National Income 1928-1932, National Bureau of Economic Research Bulletin,49, June 7, 1934.

Kuznets S, Epstein L, Jenks E (1941) National Income and Its Composition, 1919-1938, Vol 1. NBER vol 0-87014-039-6.

Magnusson L (2015) The political economy of Mercantilism, in Routledge explorations of economic history, ed Magnusson L, Routledge, London and New York.

Marshall A (1890) Principles of Economics, McMillan & Co Ltd.

Meade J (1951) The theory of International Economic Policy Volume one Balance of Payments, Oxford University Press.

Mundell R (1960) The monetary dynamics of international adjustment under fixed and flexible exchange rates, The Quarterly Journal of Economics, Vol 74, No 2, 227-257.

Mussa M (1986) Nominal exchange rate regimes and the behavior of real exchange rates: evidence and implications, Carnegie-Rochester conference series on public policy 25: 117-214.

Obstfeld M, Rogoff K (1995) Exchange rate dynamics redux, Journal of Political Economy, Vol 103, No 3, 624-660.

Obstfeld M, Rogoff K (1996) Foundations of international macroeconomics, The MIT Press.

Obstfeld M, Rogoff K (2002) Global implications of self-oriented national monetary rules, Quarterly Journal of Economics, Vol 117, Issue 2, 503-535.

Officer, L (1975) The purchasing power theory of exchange rates: a review article, International Monetary Fund Staff Papers.

Petty, William (1691) Verbum Sapienti, Reprinted from the Online Library of Liberty, A project of Liberty Fund Inc.

Petty, William (1677) Political Arithmetick, Reprinted from the Online Library of Liberty, A project of Liberty Fund Inc.

Premchand A (1995) Effective government accounting, IMF publication.

Rodrik D (2008) The Real Exchange rate and economic growth, Brookings Papers on Economic Activity.

Rogoff K (1996) The purchasing power parity puzzle, Journal of Economic Literature, Vol XXXIV, 647-668.

Samuelson P (1964) Theoretical notes on trade problems, The Review of Economics and Statistics, Vol 46, No 2, 145-154.

Summers R, Heston A (1991) The Penn World Table (Mark 5): An expanded set of international comparisons, 1950-1988, Quarterly Journal of Economics, Vol. 106, No 2, 327-368.

Summers R , Kravis I, Heston A (1981) New insights into the structure of the world economy, The review of income and wealth, Vol. 27, Issue 4, 339-355.

Studenski P (1958) The Income of Nations. Theory, Measurement. and Analysis: Past and Present, New York University Press.

The World Bank (2021) International Comparison program 2021, global highlights, report available at https://thedocs.worldbank.org/en/doc/0274411350395ce53ccd3e91a431e924-0050022024/original/FINAL-ICP-2021-Global-and-regional-highlights.pdf

The World Bank (2013) Measuring the real size of the world economy, International Bank for Reconstruction and Development / The World Bank, Washington DC, the USA.

United Nations (1953) A system of national accounts and supporting tables, United Nations Studies in methods No 2.