

Measurement Matters: Productivity-Adjusted Weighted Average Relative Price Indices*

Douglas L. Campbell[†]
New Economic School (NES)

September, 2015

Abstract

Commonly used trade-weighted real exchange rate indices are computed as indices-of-indices, and thus do not adequately account for growth in trade with developing countries. Weighted Average Relative Price (WARP) indices solve this problem but do not control for productivity differences, as developing countries are observed to have lower price levels via the Penn Effect. I remedy these problems in two ways. First I propose a Penn Effect productivity adjustment to Weighted Average Relative Price indices (P-WARP). Secondly, I introduce a Weighted Average Relative Unit Labor Cost index (WARULC) for manufacturing and show that this measure does a much better job predicting trade imbalances and declines in manufacturing employment than the IMF's Relative ULC measure created as an index-of-indices. The new series reveal that for many countries currently mired in liquidity traps, relative prices reached historic highs heading into the financial crisis of 2008. I document that in 2002 – during the surprisingly sudden collapse in US manufacturing – US relative prices had not been that high relative to trading partners since the worst year of the Great Depression.

JEL Classification: F10, F31, N70, C43

Keywords: Real Exchange Rate Indices, Relative Unit Labor Cost Indices, Weighted Average Relative Prices, Balassa-Samuelson, Trading Partner Substitution Bias

*I am indebted to comments received from four anonymous referees, from seminar participants at UC Davis, Colby College, the New Economic School, Santa Clara, and at the All-UC Economic History conference at Berkeley. I would also like to thank Ju Hyun Pyun, Paul Bergin, Robert Feenstra, Chris Meissner, Kim Ruhl, and John Devereux for their suggestions. I thank Nadezhda Kotova and Yulia Zhestkova for their stellar research assistance. Special thanks to Barry Eichengreen and the Berkeley Economic History Lab for providing access to data resources. I would also like to thank the hardworking public servants at the BLS, the BEA, and the OECD for responding to data inquiries. Note that this paper was previously circulated under the title “Through the Looking Glass: A WARPed View of Real Exchange Rate History.”

[†]Visiting scholar, Berkeley Economic History Lab. Address at New Economic School: 100 Novaya Street, Skolkovo, Moscow, Russia, 143025, Tel.: 7-925-629-6600, e-mail: dolcampb@gmail.com, Homepage: dougcampbell.weebly.com.

1 Introduction

One of the most important prices in any open economy is the real exchange rate. Trade-weighted real exchange rate indices thus provide a useful guide to both policymakers and academic economists as rough measures of the competitiveness of a currency in international trade.¹ In this paper I examine the methodology used to create these indices with the central goal of producing theoretically-appropriate measures of real exchange rate indices that an applied researcher could use to, for example, gauge the impact of RER movements on the US manufacturing sector. I find that the new series cast many key events in economic history in new light, including the Asian Financial Crisis, the swift decline of American manufacturing, the ongoing structural US trade deficit, the Great Depression, and especially the “Lesser Depression” today.

The most commonly used real exchange rate indices are constructed by the Federal Reserve, the IMF, and the OECD as indices-of-indices, which lack clear theoretical foundations. The levels of these series are not internationally comparable and they suffer from what I term a “trading partner substitution bias” problem, as they do not adequately account for growth in trade with developing countries. This problem is analogous to the “outlet substitution bias” problem in the CPI (Reinsdorf 1998) and to what is called either “offshoring bias” (Houseman *et al.* 2011) or “input-sourcing bias” (Inklaar 2015 and Reinsdorf and Yuskavage 2014).² India and China are assigned the same base value in these price indices as are Switzerland and Germany, even though the latter have much higher prices for all years, which becomes problematic when trade increases with India and China relative to countries with higher price levels. In a seminal contribution, Thomas, Marquez, and Fahle (2008) rewrote the prior 20 years of US real exchange rate history by showing that a simple Weighted Average Relative Price (WARP) index implies that the dollar appreciated substantially more from 1990 to 2006

¹I began this project while doing research on the impact of exchange rate movements on various economic variables. I soon discovered, as Thomas *et al.* (2008) did, that the real exchange rate indices created by the Fed, the IMF, and the OECD, which have appeared widely in academic research, are not suitable for many tasks for which they are often employed. In addition, there are no appropriate indices which are publicly available for easy downloading, even for the modern era, much less historically. Any economist or policymaker who wants to consult a real exchange rate index must choose between plotting a series likely to mislead (often unwittingly), or else engage in the time-consuming task of creating a series from scratch. Thus most central bank presidents and heads of state, even in severely depressed economies such as Ireland, have never seen a real exchange rate index for their own country that accounts for compositional changes in trade for the simple reason that none exist. Thus, part of the value-added of this paper is that I provide these indices for many countries on my website for free, easy downloading.

²Diewert *et al.* (2014) provides a nice overview of the general issue, which they call “sourcing substitution bias” for the context of changing intermediate input sources.

compared to “divisia” based indices-of-indices produced by the Federal Reserve Board and the IMF. Thomas *et al.* (2008) also find that a geometric WARP index does a much better job of explaining trade balances from the period 1970-2006 than divisia-based alternatives.

The problem with traditional real exchange rate indices is visible in Figure 1. When the Fed’s Broad Trade-Weighted RER index appreciated sharply in the 1980s (Panel a), the goods trade balance ex-oil, as a share of GDP, declined markedly. The puzzle is that the dollar (ostensibly) appreciated much less rapidly in the late 1990s and early 2000s, and yet the goods trade deficit was much larger. Using the methodology of Thomas *et al.* (2008), the first contribution of this paper is to create an updated version of WARP for the period 1950-2011 for 154 countries and the last three versions of the Penn World Tables (v7.1, 8.0, and 8.1), using trade weights from the other 153 countries which have complete data.³ Version 8.1 implies that the dollar appreciated much more sharply than the Fed’s index (due both to the rise of China and differing underlying data sources) and nearly 18% more sharply from the period 1990 to 2002 than WARP using PWT v7.1.⁴ This series, which appears to mirror the trade balance, can thus explain the relatively large trade deficit and the surprisingly sudden collapse in tradables-sector employment in the early 2000s, and gives credence to Ben Bernanke’s claim that the savings glut, trade, and China are at the heart of the so-called “secular stagnation” experienced by the US since 2000.⁵ In Figure 1, Panel (b), by contrast, it can be seen that both the IMF’s RULC index and the Terms of Trade (computed from v8.1 of the PWT) appear to bear little resemblance to trade.⁶ The IMF’s index implies that US manufacturing has become steadily more competitive over time, while the Terms of Trade appears to be orthogonal to the trade balance.

However, one problem with using WARP as a measure of competitiveness is that

³I thank Isaac Rabbani and Andrew McCallum of the Federal Reserve for suggesting I extend WARP internationally for a larger sample of countries for central bank research purposes.

⁴One of the benefits of WARP is that it provides a quick summary of changes in the Penn World Tables, which makes it a good way check for possible irregularities. Indeed, soon after WARP was created for a full sample of 157 countries, anomalies were found in the PWT for several countries, prompting a revision of the PWT – one of the most widely used data sets in all of the social sciences.

⁵Bernanke (2005) first raised this possibility while Fed Chair, and also wrote a series of blog posts on the topic in 2015, such as this: <http://www.brookings.edu/blogs/ben-bernanke/posts/2015/04/01-why-interest-rates-low-global-savings-glut>.

⁶I thank several anonymous referees for suggesting that the Terms of Trade are actually the most appropriate measure of competitiveness. However, given the dollar price stickiness of many goods, large changes in the terms of trade appear to be driven by swings in commodity prices such as oil. In addition, as I will discuss, it is actually not clear that we would want to exclude non-traded goods and services in a measure of competitiveness. Even so, part of the value-added of this paper is that I have also computed new Terms of Trade indices for 154 countries using v8.1 of the PWT and have made these publicly available on my webpage.

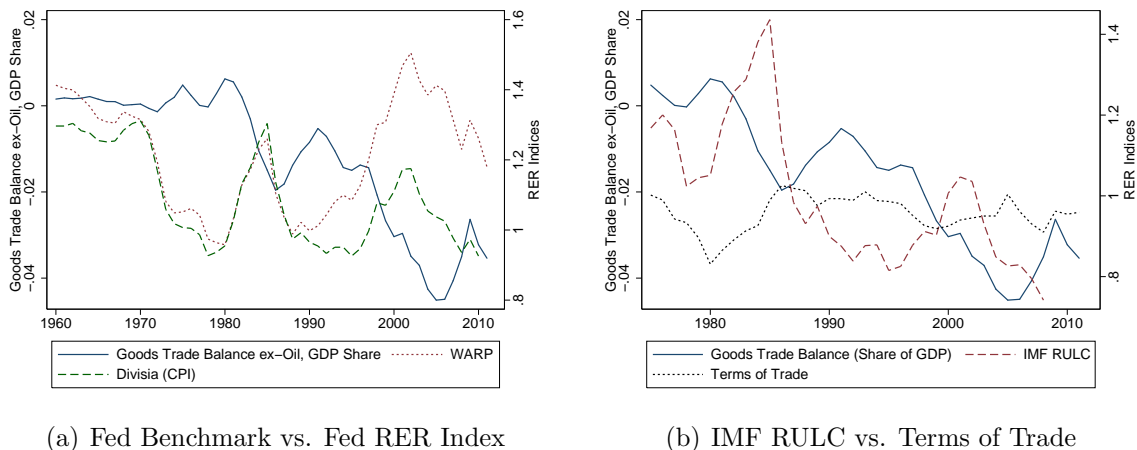


Figure 1: Measures of RER Indices vs. Trade Balance

Sources: The goods trade balance (ex-oil) comes from the BEA, while the WARP index was computed using PWT v8.1 data, and the Divisia index was computed with national CPI data using the Federal Reserve’s methodology in its Broad Trade Weighted RER Index. The Terms of Trade were also computed using PWT v8.1.

poor countries should theoretically have lower price levels according to the Balassa-Samuelson effect. Having a price level twice that of Japan in 1946 has very different implications for bilateral competitiveness than having a price level twice that of Japan in 1986. A straightforward resolution to this problem is to make a Balassa-Samuelson adjustment due to the observed Penn Effect to WARP (P-WARP).⁷ Increased trade with less-developed countries will only result in a stronger dollar index if these countries are undervalued relative to their level of development. The index is conceptually similar to the Balassa-Samuelson residuals used by Rodrik (2008) and many others in the literature on real exchange rates and growth, except that the measure proposed in this paper is a trade-weighted average of the difference between the US residual and the residuals of US trading partners.

The level of the P-WARP index indicates a substantially more competitive dollar relative to WARP for all years from 1950 to 2011, with the dollar actually 4.7% *undervalued* by 2011. This finding was not anticipated and is counterintuitive given the large structural trade deficit. However, after the dollar’s dramatic rise and fall in the 1980s, it also took several years before trade was balanced, giving rise to an academic literature on hysteresis. In the US case, I find that directional changes in P-WARP are broadly

⁷Note that in previous drafts, this index was referred to as the BS-WARP index, for “Balassa-Samuelson” adjusted WARP, but was changed due to the obvious negative connotation after a suggestion by Feenstra.

similar to the directional changes in WARP (the differences are far more pronounced for countries growing or contracting quickly, such as Ireland, Korea, and Poland). The similarity between WARP and P-WARP for the US after 2002 was not easily anticipated – the Balassa-Samuelson adjustment lowers the RER for countries growing quickly, such as China, so it could have been expected that after 2002, the P-WARP index would show a more moderate depreciation as trade with fast-growing China increased. Using PWT v7.1, the P-WARP index does show a more moderate depreciation after 2002, and was still 20% overvalued as of 2010. However, PWT version 8.0 marked up the growth in Chinese prices after 2005 and thus marked down the growth rate of Chinese GDP per capita by 21% over this period, partly moderating the impact. The P-WARP index also seems to do a slightly better job of predicting US trade balances than the WARP index in terms of both in-sample fit and out-of-sample predictive ability, although the differences are not large.⁸

Of course, it has long been recognized that real exchange rate indices need to be adjusted for productivity. This is why economists have generally preferred to use real exchange rate indices computed using unit labor costs in manufacturing rather than those based on other measures, such as consumer prices, as the key measures of competitiveness in international trade. Commonly used real exchange rate indices computed by the IMF and the OECD based on relative unit labor costs are also computed as indices-of-indices and thus suffer from trading partner substitution bias. In addition, they use fixed trade weights that do not include China. To solve this index numbers problem, I propose a simple geometric Weighted Average Relative Unit Labor Cost index (WARULC), where Unit Labor Costs are computed as total labor income in manufacturing converted to the local currency (dollars) at exchange rates divided by total manufacturing output converted to the local currency at manufacturing PPP.⁹ I compute manufacturing PPP using PWT version 8 methodology described in Feenstra *et al.* (2015), applying the Geary-Khamis indexing method to the manufacturing basic headings of all six publicly available International Comparison Program (ICP) benchmark years, and interpolating using manufacturing value-added growth rates reported by country specific sources for the years in between. The index I create shows a much greater dollar appreciation over time than the IMF or OECD indices do, and by 2001 stood 32% higher than the IMF’s index relative to 1975. This index appears to do a superior job predicting trade

⁸However, the US is somewhat of a special case, as generally, I find there is no relation between RER indices and trade balances. This is likely because of obvious endogeneity problems – a decline in productivity (such as from the Asian Financial Crisis) could lead to both a decline in exports and a RER depreciation. In the US case, the RER movements are more likely to have been exogenous.

⁹RULC’s are then the ratio of ULC’s in the US to costs in other countries.

imbalances and periods when relatively more import-competing manufacturing sectors experience relative declines in employment compared with existing series. Reassuringly, this series yields broadly similar conclusions about US competitiveness over time as WARP and P-WARP, and thus solves the puzzle discussed by Chinn (2006) about the different implications of the IMF RULC index vs. other measures of the RER.

As of 2009, while China employed about 9 times as many man-hours in manufacturing than did the US to produce slightly more manufactured goods, Chinese hourly wages in dollars were just \$1.74 compared to \$35.18 for the US.¹⁰ Thus I calculate that Chinese unit labor costs were about 37% of US unit labor costs in 2009. Although full Chinese data on employment and hours worked was unavailable through 2011, Chinese hourly wages went from being 5% of US wages to 7% of US wages in those two years alone, while production rose 24% in China versus just 10% for the US. Thus, while the gap appears to be closing, the picture that emerges of competitiveness from relative unit labor costs in the US vs. China is different from what emerges with the Balassa-Samuelson adjustment.

To the extent possible, I extend all indices over both space and time. For the US, I extend both “divisia” and WARP indices for the US historically for the period 1820-2010. The Thomas *et al.* (2008) WARP series spans 1970-2006, while the Fed’s broad trade-weighted real exchange rate index starts in 1973. The Fed’s series commences at an inopportune time as it misses the large depreciation at the end of the Bretton Woods period. I extend both series back to 1950 using the same sample of countries, trade-weighting scheme and indexing methodology as the Federal Reserve. I also extend these series back to 1922 on a consistent sample of 30 countries, and back to 1820 for a sample of five countries. Compared to divisia, WARP implies a lower US price level in the period before WWII relative to the Bretton Woods period and exhibits a slightly sharper dollar appreciation during the Great Depression, with a difference from 1928-1932 of 3%. Additionally, I propose an improvement to the Federal Reserve Board’s trade weights, but find that this leaves all indices little-changed, resulting in an increased dollar appreciation from 1992-2002 of an additional 1%.

Internationally, I produce WARP and P-WARP indices for 154 countries for the period 1950 to 2011, a far larger sample of countries using a consistent methodology than *any* other RER indices. (Even the World Bank’s WDI, which provides CPI-based

¹⁰These estimates use OECD data on US manufacturing employment and hours, which are based on household survey data for the US which are used for international comparability, and government data for Chinese employment. The better-known manufacturing employment numbers in the US come from the establishment survey, which shows 2 million workers in manufacturing. Chinese manufacturing output from the World Bank was converted into dollars at manufacturing PPP estimates, but would not be substantially different in 2009 converting at exchange rates. The hourly wage data comes from the BLS.

RER indices courtesy of the IMF, includes data for only 92 countries for the period from 1975.) I also provide WARULC indices for a handful of major economies. I find that for Italy, Greece, the UK, and the Russian Federation, the WARP and P-WARP indices reveal a much greater real appreciation since 1990 than do the IMF's divisia-based series. For example, in 2010 the P-WARP series for Italy stood more than 25% higher than the IMF's series relative to 1990, and in 2007, the UK's WARP index also stood 24% higher relative to 1990 than the IMF's RER index. Conversely, Germany's P-WARP index is similar to its IMF CPI-based real effective exchange rate. This revision of relative prices is not merely an academic curiosity given the economic problems now facing Europe. It accentuates the difficulties the European Central Bank faces in devising one monetary policy for countries with very disparate trends in relative prices.

The Balassa-Samuelson adjusted index also reveals a substantial appreciation for the relative price level of Greece, demonstrating that currency appreciation since 1990 cannot merely be explained by Greece's convergence in GDP (before the crisis) as is often thought. In addition, I show that WARP and P-WARP indices for Iceland appreciated much more rapidly than did the IMF's measures leading up to the financial crises in 2008, and that these measures have also depreciated more markedly since. By contrast, I find that WARP and P-WARP imply a more gradual appreciation for relative prices in the Russian Federation than the IMF's REER index in the 2000s.

I provide WARP and P-WARP series for China, Korea, and Japan. Once again, these indices are substantially different from commonly used divisia-based indices. I find that China's real exchange rate was undervalued by 46% in 2005, but by 2011 it was undervalued by just 20% on a Balassa-Samuelson-adjusted basis. In 2005 China's price level was 62% lower than its trading partners, but by 2011 this difference had fallen to just 39%. Japan's two decades spent mired in a liquidity trap have been accompanied by a domestic price level on average 86% higher than that of its trading partners, with an increase about 6% larger from 1990-2000 than the IMF's divisia-based counterpart.

The rest of the paper proceeds as follows: First I extend the Divisia and WARP indices to 1820 for the US, and then I introduce a Balassa-Samuelson adjustment to WARP and a Weighted Average Relative Unit Labor Cost (WARULC) measure. Next I present international extensions of these indices, and test them vs. the data.

2 Benchmarking the Fed, with Historical Extensions

2.1 Post-War Benchmark

In Figure 2(a), I benchmark the Fed’s Broad Trade-Weighted Real Exchange Rate Index using data collected from various sources detailed in Appendix Table 1. Using the same methodology and largely the same sample as the Fed, I extend the Federal Reserve Board’s series back to 1950, capturing the large depreciation after the end of Bretton Woods. While the reconstruction does not mimic the Fed’s index exactly, it never deviates by more than seven-tenths of one percent. The difference could be due to data sources, as I used national sources when WDI or IMF data were missing, such as for Taiwan. There also may be slight differences due to the Fed’s annual series being an average of underlying monthly data, or from special issues arising from the creation of the Euro or end of the Soviet Union which are not discussed in detail in the Fed’s literature on the creation of its index.

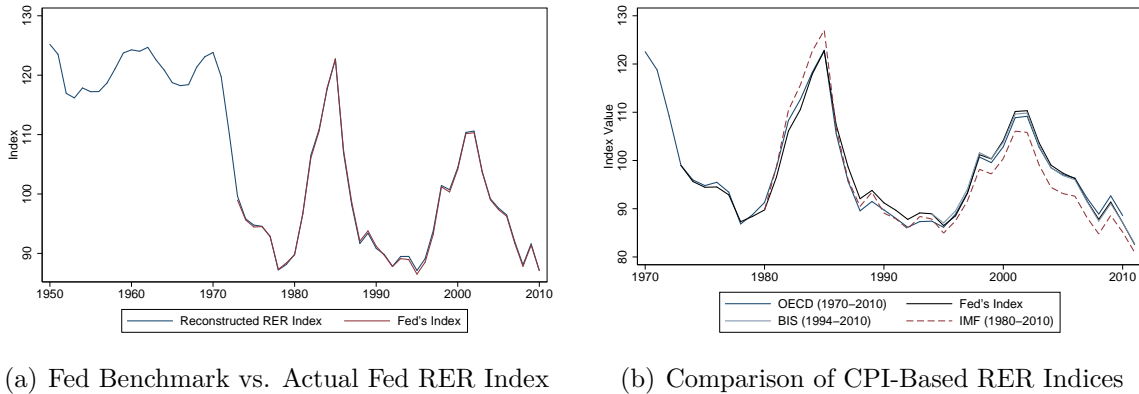


Figure 2: Benchmarking

For the period 1973-2010, I use the Federal Reserve’s trade weights, which are based on trade net of oil. For the period prior, I calculate the trade-weights using the Federal Reserve’s trade-weighting methodology, trade data from the IMF DOTS, and oil trade data from the UN’s Comtrade database. The bilateral oil trade data is not available before 1962, so I assumed that the share of each country’s trade represented by oil for the period 1950-1961 is the same as the average for the period from 1962-1966. Even ostensibly major revisions to the weights during this period seem to have minor impacts on the overall index—if I use total trade with no adjustment for oil before 1962, the series is little changed.

In Figure 1(b), I show that other CPI-based effective exchange rate indices, created

by the IMF, the OECD, and the Bank of International Settlements, all use very similar methodology as the Fed and yield very similar results. Of these, the IMF is the outlier, perhaps because it uses constant trade-weights for the entire period (see Bayoumi *et al* 2006).

2.2 Historical Series

Figure 3 below extends the Fed’s series using historical data for 30 US trading partners, and compares it to the Fed’s post-war broad trade-weighted real exchange rate index (indexed to be equal in 2010). Six partners have data from 1820, eight from 1830, 12 by 1861, and 19 by the 1880s; the sample reaches its max of 30 in 1922 (the historical sample of countries with the starting date for each country is in the appendix where it is compared to the Fed’s sample). This series builds on the scholarship of Taylor (2002) and Fratzscher, Mehl, and Vansteenkiste (2011), who provide real exchange rate indices for the US from 1880, and Catao and Solomou (2000), who provide trade-weighted RER indices for a variety of countries from 1870-1914. Taylor’s series uses a simple average as historical trade data was not then available, while Fratszher *et al.*’s series includes six trading partners.¹

The historical trade weights are computed using trade data from the IMF DOTS, Comtrade, Jacks, Meissner, and Novy (2011), Barbieri (2002), and the Historical Statistics of the United States. Barbieri (2002) and the Jacks, Meissner, and Novy (2011) data begin in 1870 for a broad sample of countries, while the HSUS data begin in 1790 for a smaller sample. The prewar trade weights are simply the share of imports and exports in total US trade, while post-WWII I also include third country weights using the methodology which I discuss in Section 3.2 (I plot the weights over time for major trading partners when I introduce historical WARP in section 3.4). To extend the coverage of countries in the trade-weighted sample, for three countries I imputed trade shares for several decades based on the earliest recorded trade and changes in each country’s share of world GDP as estimated by the most recent revision to the Maddison project.

¹Taylor (2002), notes that a trade-weighted index would be ideal, suggests it as a direction for future research, but writes that it would be a “significant undertaking”.

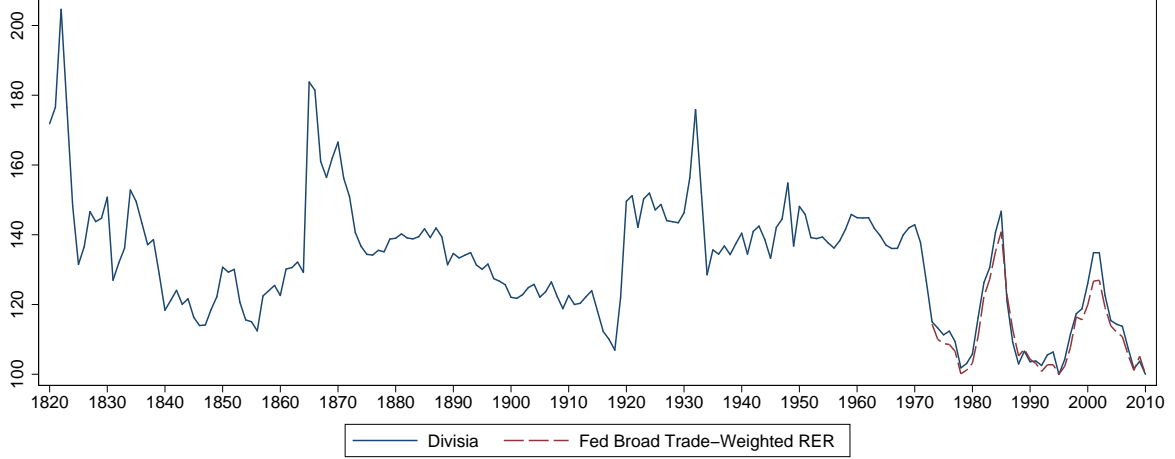


Figure 3: Historical Index Benchmarked to the Post-War Fed Index

3 Indexing Methods

3.1 A Review of Divisia vs. WARP

The Fed’s Broad Real Exchange Rate Index is computed as a weighted average of changes in underlying bilateral real exchange rate indices (this method is called “divisia”, or a Tornquist index), where the base year value of each bilateral index is arbitrary. This is the appropriate construction of a real exchange rate index based on country-specific CPIs, as CPI indices only contain relevant information when movements are plotted over time, while the base level of a CPI index is arbitrary. The lack of comparability of CPIs internationally necessitates the choice of “divisia” as an indexing method in this case, and was likely to have been the sole motivation for what is otherwise an atheoretical index from the perspective of economic theory.¹¹

The Fed’s real exchange rate index is:

$$I_t^d = I_{t-1} \times \Pi_{j=1}^{N(t)} \left(\frac{e_{j,t} p_t / p_{j,t}}{e_{j,t-1} p_{t-1} / p_{j,t-1}} \right)^{w_{j,t}}. \quad (3.1)$$

Where $e_{j,t}$ is the price of a dollar in terms of the currency of country j at time t , p_t is the US consumer price index at time t , $p_{j,t}$ is the consumer price index of country j at time t , $N(t)$ is the number of countries in the basket, and $w_{j,t}$ is the trade weight of country j at time t . The base year is set at an arbitrary level (by necessity), both

¹¹In the Theoretical Appendix, I present a basic Armington model, and show that it implies that relative prices matter, and not merely just changes in relative prices. By contrast, I am not aware of any theoretical arguments for using divisia instead of WARP when information on price levels exist.

for the index and for each bilateral real exchange rate. The trade weight is a weighted average of each country's share of imports, exports, and the degree of competition in third markets (trade weights are discussed later in this section).

Note that while directional changes in real exchange rates will affect the index, changes in trade weights between countries with different levels of real exchange rates will not. Note that while, if we use CPI-based RER indices, the levels of real exchange rates are not known, but that if we use relative price level data from the PWT instead, then this information would be discarded when using a divisia index. The problem with divisia becomes especially apparent when there is a shift in trade from countries with similar price levels to countries with very different price levels.

TABLE 1
Comparing Indexing Methods

Year	Share of Trade with Country A	RER A	RER B	Weighted Average	Geometric Average	Fed's Indexing Method
1	87.0%	0.92	3.68	1.28	1.11	100
2	77.9%	0.97	3.06	1.44	1.26	100.1
3	74.2%	0.89	3.11	1.46	1.23	94.1
4	71.3%	1.11	2.92	1.63	1.46	107.9
5	63.8%	1.25	2.98	1.88	1.71	117.3
6	53.5%	0.97	2.73	1.79	1.57	98.2
7	49.8%	0.85	2.09	1.47	1.34	80.5
8	43.9%	0.94	2.01	1.54	1.44	82.4

Table 1 describes a potentially problematic example of how a divisia-based RER index can be misleading. It compares several possible real exchange rate indices: a simple weighted average, a geometric average (used by Thomas *et. al.* (2008)), and the Fed's method. In this example, the bilateral real exchange rate for country A varies without trend, while the real exchange rate for country B appreciates substantially over the period relative to the home country. Reflecting this, the Fed's indexing method (also a geometric average) reveals a substantial depreciation. However, at the beginning of the period, the home country is trading mostly with country A (87% with country A implying 13% with country B), which has a similar price level, while at the end of the period a majority of trade is with country B, which has a much lower price level. This is reflected in a simple weighted average, or in the geometric average, which both show that by the end of the period the home country's currency is much higher vs. a weighted average of its trading partners than it was at the beginning of the period.

In addition, the simple weighted average of real exchange rates has an intuitive interpretation. For example, its value of 1.28 in the first year means that the price level is 28% higher at home than in a weighted average of its trading partners, and about 20% lower than in the eighth year. By contrast, the value of 100 for year one using the Fed’s method is just an arbitrary number with no economic meaning by itself.¹²

The choice of indexing method has dramatically differing implications for the behavior of the index in the special example above, but do examples such as this happen in practice? In fact, the data in Table 1 are real: “country A” is the Euro Area, “country B” is China, the trade shares are taken from the actual trade-weighting scheme used by the Federal Reserve in the creation of its broad trade-weighted real exchange rate series. The real exchange rates are computed using relative prices (specifically, the ratio of the relative price of output-side real GDP via the Penn World Tables v8.0) for every several years between 1990 and 2010.¹³ In 1990, the US traded a small amount with China relative to the Euro Area, but by 2010 trade the US traded more with China, which has a comparatively inexpensive price level. As a result the Fed’s method and a simple weighted average yield dramatically diverging series.

This example illustrates why divisia-based indices can present a very misleading view of competitiveness, since the relative price levels should theoretically matter more for competitiveness than changes in price levels from arbitrary base year values.¹⁴ When a large country with a low price level, such as China, goes from autarky to free trade in a single year, but has a price level that is initially fixed due to capital controls or government control of the nominal exchange rate in concert with nominal rigidities, both intuition and trade models (such as a simple Armington model) would suggest that this

¹²Note that while the arithmetic average appears to be easier to intuit than the geometric average, which is less affected by extreme values, the geometric average is clearly superior here. Instead of using the price of goods in the US relative to countries A and B, I could have used the prices of goods in those countries relative to the US. Inverting the results using the arithmetic mean (which would be the harmonic mean), would yield very different results. By contrast, with a geometric mean, inversion yields the same results. Otherwise I might favor the arithmetic mean, since from a competitive perspective, having a currency overvalued by 20% with respect to one trading partner is probably more damaging than having your currency overvalued by 1% with respect to 20 countries, although there is not necessarily a lot of evidence for this assertion other than the recent apparent impact of China on the US manufacturing sector. The arithmetic average will yield the same results for these two cases, while the geometric average will yield a lower value for the first scenario.

¹³Note that this is the PWT v8.0 equivalent of the measure of relative prices that Thomas *et al.* (2008) use, as the variable definitions changed slightly in version 8.0 of the PWT as PPP was no longer included by itself.

¹⁴By competitiveness, I mean the ability of an economy (or the manufacturing sector) to compete with trading partners. Thus, what is desired is a measure of competitiveness that could gauge the severity of trade shocks.

country would pose a competitive threat.¹⁵ This threat is reflected in the WARP index, but it would not be reflected in the Divisia index. And while it may be possible to write down a model where the Divisia index is more relevant than a simple Weighted-Average of Relative Prices, most standard trade models would favor WARP. In fact, China's opening of trade has posed a competitive threat (see Autor, Dorn, and Hanson, 2013, and Pierce and Schott, 2015), while Thomas *et al.* (2008) show that WARP seems to do a better job explaining US trade flows.

3.2 Trade Weights: Fixed Import and Export Share Weights vs. Adjustable

Another issue in choosing an exchange rate index is the choice of trade weights. An ideal trade-weighting scheme would simply be a weighted-average of trade exposure in each market:

$$w_{us,j,t}^{k,ideal} = \sum_i^{N(t)} \varepsilon_{us,i,t} \varepsilon_{j,i,t}, \quad (3.2)$$

where $\varepsilon_{us,j,t}$ is the bilateral export share from the US to country j, defined as:

$$\varepsilon_{us,j,t} = \frac{X_{us,j,t}}{\sum_{k=1}^{N^*(t)} X_{us,k,t}}. \quad (3.3)$$

Where $X_{us,j,t}$ are exports from the US to country j at time t. Note that here, $N^*(t)$ includes all countries, including the US and country k, thus it requires data on each country's total shipments, including shipments to itself, which is usually not readily available.¹⁶

¹⁵A referee suggested that the Armington model implies that Divisia is actually the correct index. In the appendix, I demonstrate that this is not the case. Virtually any model in which relative prices matter will imply that WARP is theoretically appropriate. However, it is not necessarily the case that Divisia is without any theoretical foundations. One argument for Divisia is that if data on the levels of relative prices between countries is either missing or of poor quality, then it could be preferred. A second argument is that it yields different information than WARP, so looking at both measures may be useful.

¹⁶Admittedly, one could use data on total GDP or on total manufacturing production, and then subtract exports. Challenges with this approach include (1) some countries have exports larger than GDP, since products can be imported and then exported, so some (arbitrary) adjustment would have to be made to the raw data, and (2) manufacturing data leaves out trade in other sectors, and will also not be complete for the time span of the created series. The manufacturing data with the widest coverage, UNIDOs, is notoriously incomplete, even for many major countries, while OECD data lacks developing countries and the WDI only contains value-added and not shipments. Thus one would have to settle on country-specific sources for complete data. While these adjustments, and others, may be worthwhile, a key finding in this literature is that methodological improvements in the trade weights tend to have little impact on the trade weights themselves, and thus on the indices. Simulations based

Likely due to these data limitations, the Fed’s trade-weighting scheme is a weighted average of three measures – import exposure, destination-market export exposure, and the degree of competition in third markets. It attaches weights symmetrically – 50% to imports and 25% each to destination export markets and competition in third-country markets. Of course, since U.S. trade is often unbalanced, a preferable approach, used by the Bank of International Settlements (Klau and Fung, 2006, and also preferred by Chinn, 2006), is to decide the import weight based on the share of imports in total trade. The Federal Reserve’s own documentation admits that its trade-weighting scheme is arbitrary on the grounds that moderate adjustments in the direction of plausibility add complication but seem to have little impact (Loretan, 2005).

The Fed’s trade weight for country j at time t is:

$$w_{j,t}^{FED} = \frac{1}{2}\mu_{us,j,t} + \frac{1}{2}\left(\frac{1}{2}\varepsilon_{us,j,t} + \frac{1}{2}\tau_{us,j,t}\right). \quad (3.4)$$

Where $\mu_{j,t}$ is country j ’s share of merchandise imports into the US, equal to:

$$\mu_{us,j,t} = \frac{M_{us,j,t}}{N(t) \sum_{k=1} M_{us,k,t}}. \quad (3.5)$$

$M_{us,j,t}$ are imports from country j for $N(t)$ markets at time t . The bilateral export share, $\varepsilon_{us,j,t}$, is defined as in equation (3.3), except now this measure only includes exports and excludes domestic shipments. Since US exports also compete with the exports of country j in third-market economies, $\tau_{us,j,t}$ measures the competitiveness in third markets:

$$\tau_{us,j,t} = \sum_{k \neq j, k \neq us}^{N(t)} \varepsilon_{us,k,t} \mu_{k,j,t} / (1 - \mu_{k,us,t}). \quad (3.6)$$

Where $\mu_{k,j,t}$ is the fraction of country k ’s merchandise imports from country j in year t and the multiplicative factor $(1 - \mu_{k,us,t})$ ensures that the weights sum to 1.

Instead, I propose using the share of imports in total trade, denoted m_t , for the import weight. I then follow the Federal Reserve Board in splitting the export weight

on assumptions of what domestic shipments actually look like suggest that even the Fed’s approach does not contain obvious large biases, which was quite contrary to my initial suspicion. In the appendix, I provide estimates for domestic shipments and production for the US, and create a separate RER series that adjusts for domestic competition.

between the destination and third-country markets.

$$w_{j,t}^{Alt} = m_t \mu_{us,j,t} + (1 - m_t) \left(\frac{1}{2} \varepsilon_{us,j,t} + \frac{1}{2} \tau_{us,j,t} \right). \quad (3.7)$$

When the alternative trade-weights are used which reflect the actual composition of imports and exports in trade, I confirm Loretan’s (2005) finding that the trade-weights between these two approaches are little-changed, and thus for simplicity report all results other than the benchmarks using these adjusted trade weights. However, this does not imply that using fixed trade weights instead would also not matter. In that case, the trade weights and results would be substantially altered, particularly for the class of Weighted-Average Relative (WAR) exchange rate indices.

Another improvement I make in terms of the weights for the WARP and P-WARP indices is that instead of using the top 30 trading partners as the Federal Reserve does, I include weights for all 153 countries which have price level data in the PWT and trade data in the IMF’s DOTS.

A very prudent second critique is that the Fed’s trade-weights measure trade in goods rather than trade in value-added. Bems and Johnson (2012) show that for the US, the differences in trade shares using value-added measures of trade seems to make little difference.¹⁷ For example, they find that the US trade share with China shrinks by just -.2% in 2005 when they use value-added measures of trade as compared with aggregate trade (ex-oil).

A third critique was mounted by Ho (2012), who proposed using GDP weights instead of trade weights, and found some support that in many cases (although not for the US), the GDP weights do a better job of explaining real exports using cointegration analysis. Thus, following Ho (2012), I also provide GDP-weighted versions of my index, which actually differ more substantially than trade-weighted indices for the class of weighted-average relative indices proposed here.

A fourth critique I owe to an anonymous referee, who argued that theoretically, we should use physical trade *volumes* rather than aggregate trade values. However, there is a reason why the Federal Reserve, the IMF, the World Bank, the OECD, the BIS, and Thomas *et al.* (2008) all used trade values instead of volumes: you would still need some way to compare apples and oranges, and using prices is a very natural choice of

¹⁷While there is very good logic for why one would want to use trade in value-added rather than actual trade, the downside is that it effectively ignores trade in intermediate inputs. Thus the domestic value-added share of a highly fragmented production processes could see greater volatility in value-added for a given movement in the exchange rate since multiple decisions must be made whether to import or buy domestically.

weights, and would seem to make more sense than physical weight.

Lastly, the rationale for combining import, export, and third country weights is to create an aggregate measure of competitiveness, but in some cases, one might want to look only at the evolution of relative prices of only imports or exports separately.¹⁸ In Section 7 I consider such a case and derive an index only using import weights.

3.3 Post-War WARP for the United States

Weighted average relative prices (WARP) are computed as a geometric weighted average using trade-weights, $w_{j,t}$, of the nominal exchange rate, $e_{j,t}$, divided by purchasing power parity, $PPP_{j,t}$.¹⁹

$$I_t^{WARP} = \prod_{j=1}^{N(t)} \left(\frac{e_{j,t}}{PPP_{j,t}} \right)^{w_{j,t}} = \prod_{j=1}^{N(t)} \left(RER_{j,t} \right)^{w_{j,t}}. \quad (3.8)$$

The nominal exchange rate divided by PPP (the ratio of the local currency price of a basket of goods) yields the ratio of the price level in the US to the price level in the comparison country, also called the relative price. Thus if the nominal exchange rate between the US and China is 8 to 1, but the cost of a given basket of goods costs two Yuan in China yet only a dollar in the US, then the price level in the US will be four times the price level in China.

In Figure 4 I compare WARP using version 8.1 of the PWT to four alternatives: (1) WARP using version 7.1 of the PWT, (2) WARP using GDP weights as suggested by Ho (2012) (using a consistent sample of countries from 1970), (3) a divisia index using CPIs with the Federal Reserve’s Methodology, and (4) divisia using the PWT. WARP using PWT’s version 8.1 has lower US relative prices in every period up until the end of the 1990s compared to version 7.1, after which the indices are broadly similar (as might have been expected). Consequently, version 8.1 has a larger relative price increase for the US for the period 1990-2002, which registered as a much larger shock to relative prices than the 1980s dollar appreciation.

¹⁸Again, I thank an anonymous referee for suggesting I explain why I would want to combine import and export trade-weights.

¹⁹Older versions of the PWT include a PPP variable, but version 8.0 contains price levels (PPP divided by the nominal exchange rate). Thus it is easier to compute relative prices directly using the ratio of the price-level of output-side US GDP to the price level of output-side GDP in other countries (using expenditure-side GDP instead leads to very similar results). I thus use the same measure of relative prices at Thomas *et al.* (2008) Also note that the PWT creates estimates of the price level annually which are based on the ICP benchmark years with years in between interpolated using country-specific deflators.

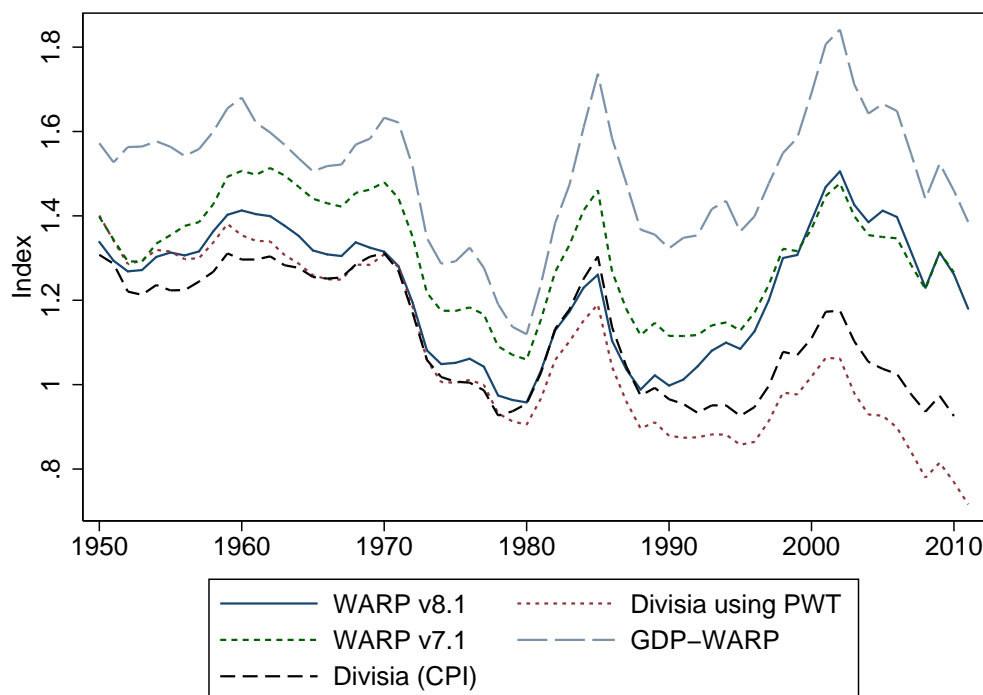


Figure 4: WARPs vs. Divisias

I also find that divisia using PWT implies a further dollar depreciation than divisia using country CPIs, matching the finding of Thomas *et al.* (2008). For ease of comparison the Divisia using the CPI (essentially the Fed's series) is multiplied by a scaling factor so that it begins at the same level as the WARP in 1973, which gives the Fed's series base year an intuitive economic meaning – in 1973, the U.S. price level was about 30% higher than a (geometric) weighted average of U.S. trading partners. WARP v8.0 approximates the Fed's index up until the dollar appreciation in the 1980s, when it shows less of an appreciation (this was much less apparent in version 7.1 of the PWT). Since the early 1990s, the WARP index reveals a much larger appreciation relative to the Fed's index, appreciating 26% more from 1990-2002. From 1990-2011 WARP appreciated by 12.9% versus a 9% *depreciation* according to divisia. The divisia index computed using the PPP of output from PWT v8.0 is very similar to that using expenditure-based PPP, and also very similar to the index which uses World Bank GDP deflators, as employed in the construction of value-added exchange rates (Bems and Johnson, 2012, and Bayoumi *et al.*, 2013).

3.4 A WARPed View of US Real Exchange Rate History

This paper is the first to plot weighted average relative prices for the U.S. before 1970, adding 150 years of data to the Thomas *et. al.* (2008) series. This series uses PWT v8.1 as the basis of all bilateral relative prices from 1950, and before that extrapolates historically using data sources detailed in Table A.3.

How does WARP change our view of history? The major difference is that in the WARP series, the price level was lower in the interwar period and in the period before World War I than *divisia* relative to the Bretton Woods period. In the *divisia* series, the interwar price level was generally higher. Another difference is that the *divisia* series exhibits a more dramatic depreciation in the period following the Civil War and just before the return of the US to the gold standard.

Outside of the World War II era, when trading partners rapidly evolved and the sample temporarily shrinks, the largest disparity between *divisia* and WARP comes after 1990. This implies that the recent rise of China and its impact on average relative prices is unprecedented in US history. During Japan's more methodical rise up until the 1980s (see the trade shares in Figure 7), the US did not experience the same large increase in relative prices. One small difference is that the WARP view of history reveals a slightly sharper appreciation during the Great Depression, with a 25.2% appreciation from 1929-1932 vs. a 22.2% appreciation using a *divisia*-based method from 1928-1932. In Figure 3 in the previous section, one of the striking features of the long run history of the dollar is that prices in the U.S. in 2010 were as low relative to trading partners as they had ever been. This is no longer the case with the WARP index, as prices in the US prior to the Civil War were much cheaper relative to trading partners than they are today, although extreme caution is warranted when interpreting historical data that result from using country-specific deflators over long periods of time.

The historical estimates of price-levels used to create the WARP index were created by starting with Penn World Table estimates and then extrapolating backward using the best historical estimates available. For the most important trading partners such as the UK, I then checked relative prices with studies conducted by economic historians (Devereux and Ward, 2003 and 2006) in benchmark years. Reassuringly, for most of this period, my relative price estimates never differ by more than 10%. However, there is a difference in the late 19th century that widens up until World War I, with Devereux and Ward (2003) finding higher US prices. For Japan, my price estimates (Table A.3 in the appendix) are very similar to Allen *et al.* (2011).

My estimate for the real exchange rate between the US and the UK in 1925 is also

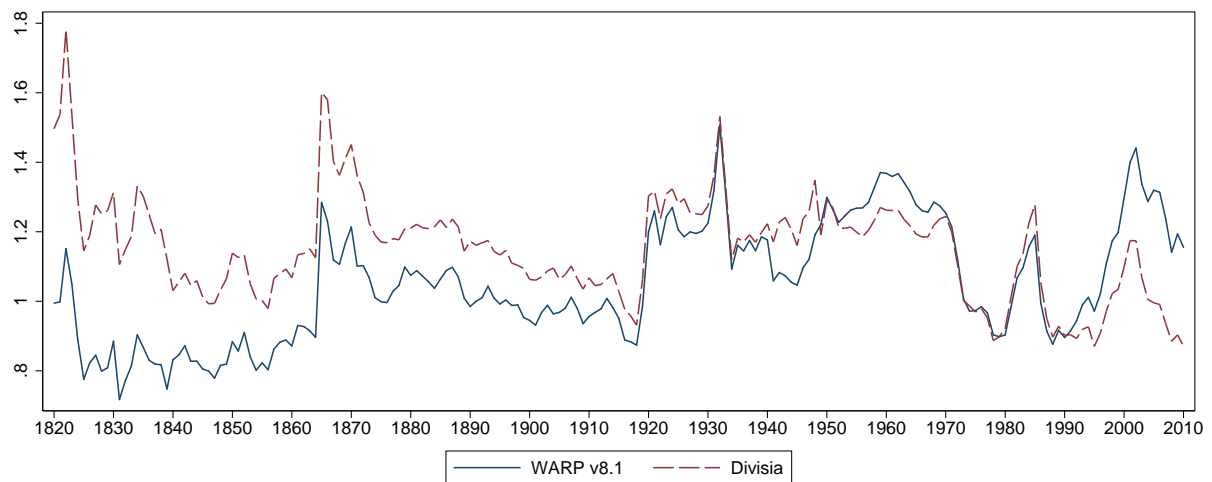


Figure 5: WARP vs. Divisia, 1820-2010

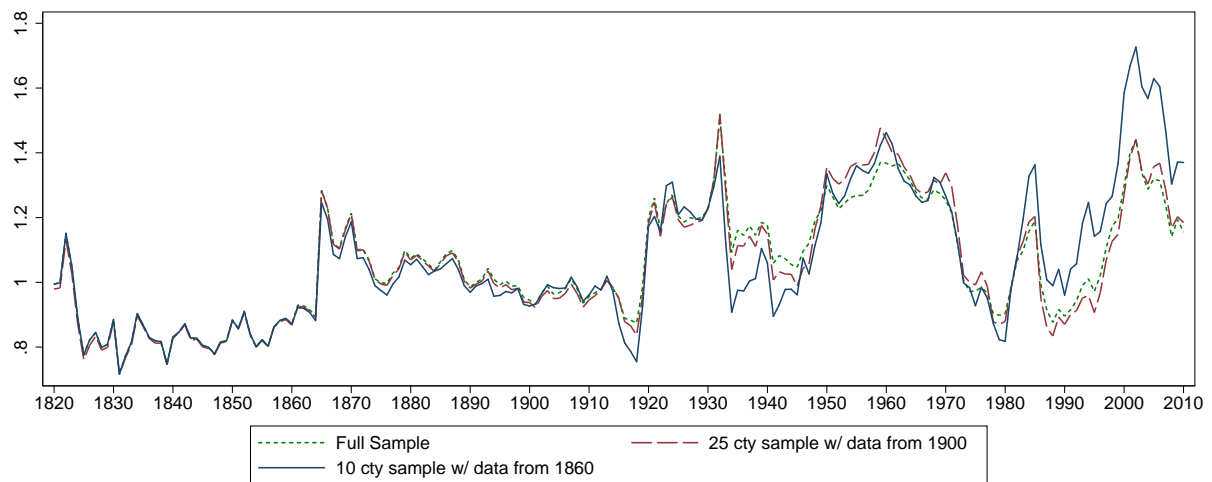


Figure 6: WARP, Full Sample vs. Smaller Balanced Samples

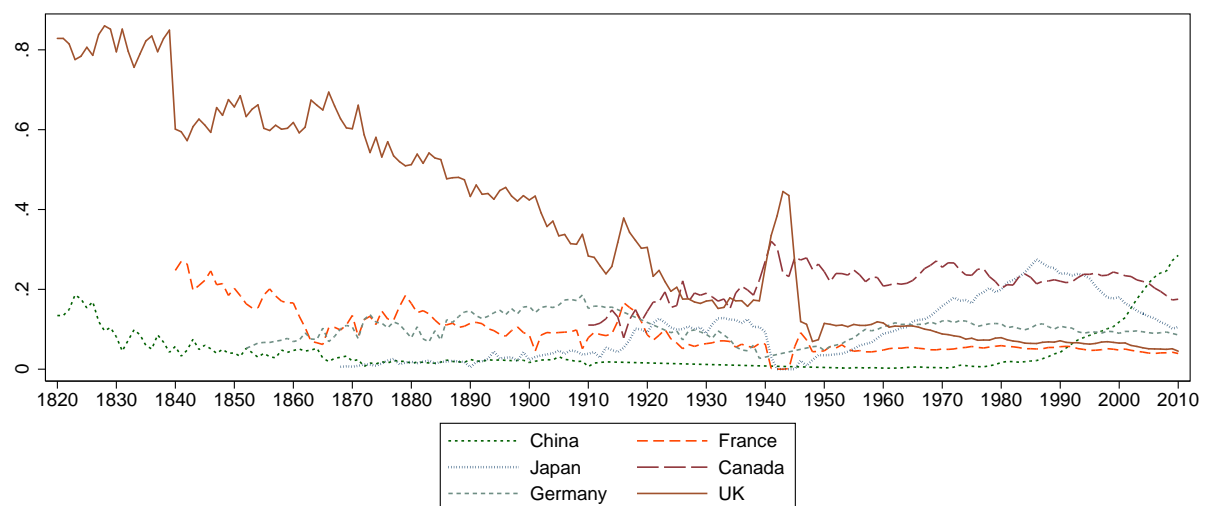


Figure 7: Trade Shares in Sample, 1820-2010

reassuringly very close to John Maynard Keynes’s estimate in that year. As Chancellor of the Exchequer, Winston Churchill decided to return to the gold standard at the prewar parity, citing wholesale PPP data which reportedly showed the price differentials between America and Britain were similar. Keynes’s alternative PPP series based on retail prices and wages showed that a return to gold at the prewar parity would result in a pound overvalued by about 12% (Skidelsky 1995). My series estimates that the pound was overvalued by just 9%.

4 Balassa-Samuelson Productivity Adjustment

The WARP index, while likely preferable to the Fed’s series for the purposes of measuring the competitiveness of U.S. goods and services in international trade, may not be the optimal method since it only factors in prices and not productivity. The Balassa-Samuelson (or the “Penn”) effect implies that if traded sectors have relatively fast productivity growth, then the productivity differentials between rich and poor countries will be greater in the tradable sectors. In each country wages in the non-traded sector will be bid to equality with wages in the traded sector, which depend on productivity, and hence non-traded prices in less developed countries will be lower, reflecting their lower productivity.

Intuitively, if Afghanistan has low wages (and prices), but also low productivity, then its low price level does not necessarily indicate a competitive threat. The original motivation for this paper was contemplating the case of Japan in 1985 vs. 1960. Japan was a much more productive economy in 1985, but the Japanese price level relative to the US had not appreciated much from the 1960s. Given its higher productivity, it is clear that Japan in 1985 should have been a larger competitive threat than Japan in 1960, which it was. Thus, what matters for competitiveness is the level of real exchange rates relative to some measure of productivity, such as per capita GDP.

Theoretically, the key result from the Balassa-Samuelson model (borrowing the setup from Obstfeld and Rogoff, 1996), a two sector model with tradable and non-tradable sectors, and with a tradable sector price which is the same everywhere due to the law of one price, is that the ratio of the price of non-tradables to tradables is the ratio of tradable to non-tradable productivity.

$$p_N/p_T = \theta_T/\theta_N \tag{4.1}$$

Thus, countries with relatively higher tradable-sector productivities will have higher

non-tradable prices. The overall price level includes both tradable and non-tradable goods, and so the relative price of US goods to foreign goods can be written as:

$$RP = \frac{p_T s_T + p_N (1 - s_T)}{e p_T^* s_T + e p_N^* (1 - s_T)}, \quad (4.2)$$

where e is the nominal exchange rate, or the dollar price of foreign currency, s_T is the share of tradables, and foreign prices are denoted with an asterisk. We can rewrite the equation for relative prices using equation (4.1):

$$RP = \frac{(s_T + \theta_T / \theta_N (1 - s_T))(p_T)}{(s_T + \theta_T^* / \theta_N^* (1 - s_T))(e p_T^*)} \quad (4.3)$$

In equilibrium, the price of tradable goods are the same in the US and abroad ($p_T = e p_T^*$), and thus the term on the right of equation 4.3 will cancel. This relative price, which depends on the ratio of relative tradables-to-nontradables productivity levels for the US vs. trading partners, has historically been very strongly correlated with relative per capita GDP. However, the nominal exchange rate moves much more sharply, even within the course of a day, than local retail prices. Thus, if the nominal exchange rate moves from its equilibrium value, but local-currency prices of goods are rigid, then the relative price could also deviate from its equilibrium level. Thus, a sudden dollar appreciation (or a decline in the dollar price of a foreign currency) would lead to US prices which are also high relative to their equilibrium value, indicating overvaluation.

This is essentially the logic behind a large body of research which studies the impact of exchange rates on the economy using the Balassa-Samuelson residual, including Frankel (2006), Rodrik (2008), Johnson, Ostry, and Subramanian (2010), Stahler and Subramanian (2014), Cheung, Chinn, and Fujii (2010), Nicita (2013), and Glüzmann *et al.* (2012, 2013). While Balassa-Samuelson residuals may be an imperfect measure of competitiveness, these studies generally find that BS residuals are related to economic outcomes. For example, Nicita (2013) finds a negative relation between overvaluation and bilateral exports, and also finds a positive relation between having an overvalued residual and protectionism (including both tariff policies and antidumping cases), while Rodrik (2008) finds a negative relation between overvaluation and economic growth. Glüzmann *et al.* (2012, 2013) argue that the link from undervaluation to growth runs through savings and investment in developing countries. Stahler and Subramanian (2014) offer a dynamic validation for the empirical Penn Effect, as they find that faster-growing countries also experience faster-growing price levels. Lastly, there is, in fact, a very tight empirical correlation between price levels and GDP per capita (see Figure

while the estimate will increase a bit if we exclude major oil producers (as suggested by Stahler and Subramanian, 2014). However, in the interests of conservatism, I use the estimate using the full sample. The residual $\epsilon_{US,t}$ has a simple economic meaning – it tells us how over- or under-valued the dollar is relative to where it should be based on US GDP per capita. This number is then adjusted based on the relative valuation of US trading partners. The advantages of this adjustment should be obvious – the US trades much more with countries such as Canada and Mexico, and thus the US’s valuation relative to its largest trading partners should matter more than the residual by itself. If the US and each of its trading partners were to lie on the Balassa-Samuelson regression line, then the average difference in the residuals would be zero, yielding an index value of one, indicating that the dollar is fairly valued.

One can see the relationship between divisia, WARP, and P-WARP by totally differentiating the log of the P-WARP index:

$$d \ln I_t^{PWARP} = \underbrace{- \sum_i w_{it} d \ln(RER_{j,t})}_{\text{Divisia}} + \underbrace{\sum_i d w_{i,t} \ln(RER_{j,t})}_{\text{WARP}} - \beta d \ln(RGDPPC_{US,t}) + \beta \sum_i w_{i,t} d \ln(RGDPPC_{i,t}) \quad (4.6)$$

Thus there are four distinct channels by which the P-WARP index can be changed. The first term tells us that when bilateral real exchange rates move, this affects the index. This term is equivalent to the changes in the divisia index. The second term indicates that shifts in trade patterns also affect the index. The first two terms combined yield the changes from WARP. The third and fourth terms are new—they indicate that when US real GDP per capita increases, this will reduce the P-WARP index holding the real exchange rate equal, and when real GDP per capita rises in US trading partners, the index will appreciate if prices do not also adjust as expected based on the Penn Effect.

The Balassa-Samuelson-adjusted weighted average relative price (P-WARP) index for the US is plotted below vs. WARP and the Fed’s benchmark in Figure 9. The magnitude of the dollar’s appreciation from 1990-2002 is much larger than the Fed’s index indicates, with an appreciation of 43% for the P-WARP index to just 21.7% for the Fed’s index (and 50.8% for WARP). Over the period 1990-2011, the P-WARP index appreciated about 11.5% vs. 18% for WARP, while the Fed’s index fell by 9.1%. The major difference between the P-WARP and WARP indices for the US is that the P-WARP index implies a much lower valuation for the dollar in all periods, with the dollar on average just 1.2% more richly valued than trading partners (well within the margin

of error of being fairly valued on average) vs. an average of 22.9% using WARP for the period 1950-2011. This, by itself, is one piece of evidence for why an adjustment for the Penn Effect is warranted. If one interprets the high value of the US WARP index as dollar overvaluation, then the question becomes how the dollar could have been overvalued for so much of the post-war period. Theoretically, long periods of overvaluation should have led to a shrinking of tradables sectors and thus a fall in the real exchange rate to parity. The Balassa-Samuelson model presented above, of course, can solve this apparent puzzle. The US's history of having a high price level is fully explained by its high level of GDP per capita.

Figure 11 decomposes the degree of over or undervaluation in exchange rates according to the Balassa-Samuelson regression between the US, US trading partners, and China. In the 1992-2002 period, the US residual and the average residual of US trading partners both appreciated significantly. China became much less undervalued in this period, but rapid GDP growth and a generally weak dollar meant that while the Renminbi appreciated significantly vs. the dollar after 2000, China's Balassa-Samuelson residual was essentially unchanged until 2008.

One alternative to the P-WARP index with a fixed Balassa-Samuelson coefficient would be to let the Balassa-Samuelson coefficient vary by year, as implemented by Johnson, Ostry, and Subramanian (2010). A compelling reason to discount this method *a priori* is that Feenstra *et al.* (2015) convincingly show that the apparent growing Balassa-Samuelson effect discovered by Bergin, Glick, and Taylor (2006) using version 6 of the PWT is a statistical artifact of extrapolated data. Data from benchmark years of the International Comparison Project always show a similarly-sized Balassa-Samuelson effect, although there is year-to-year variation. For this reason, and since the results are similar to the constant-coefficient P-WARP index since 1970, I would refer readers interested in the time-varying Balassa-Samuelson adjusted series to the Not-for-Publication Appendix.

An alternative to using indices based on PPP data from the Penn World Tables would be to use data on export prices relative to import prices, which is also called the "terms of trade".²¹ The BEA's terms of trade index for the US does suffer from the same index numbers problem as other RER indices, although this problem should theoretically be corrected for export and import price indices provided in PWT v8.0 and v8.1, at least for the period 1984 to 2007, as prices are computed in this period using "EKS aggregation" (Feenstra *et al* 2015). However, the goal of this paper is to create an index that summarizes the change in competitiveness due to shocks to the real

²¹I am indebted to an anonymous referee for suggesting this.

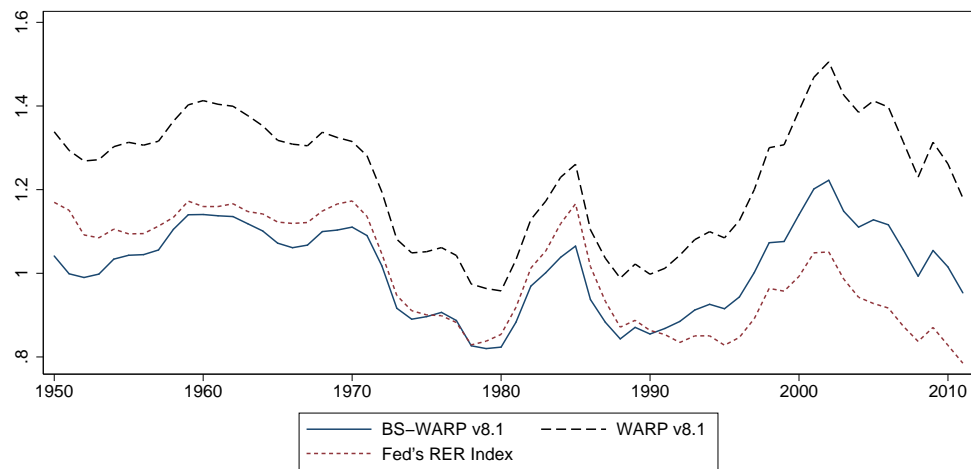


Figure 9: P-WARP vs. WARP vs. Divisia

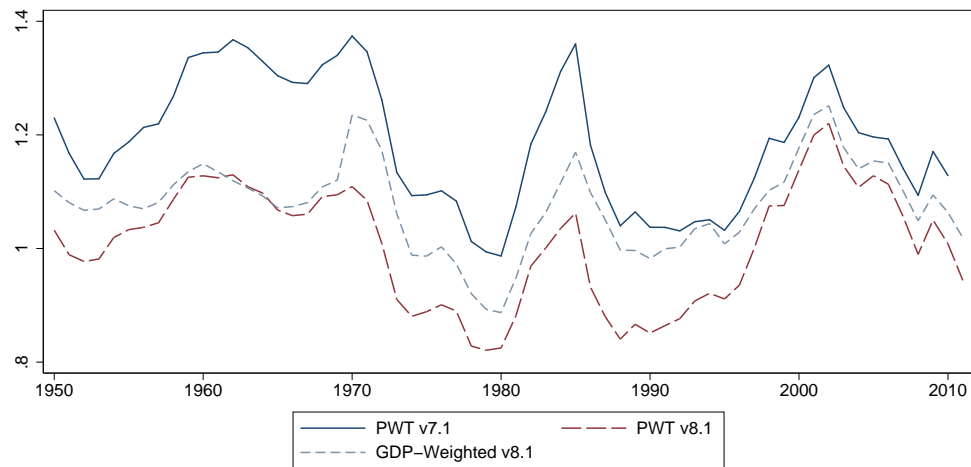
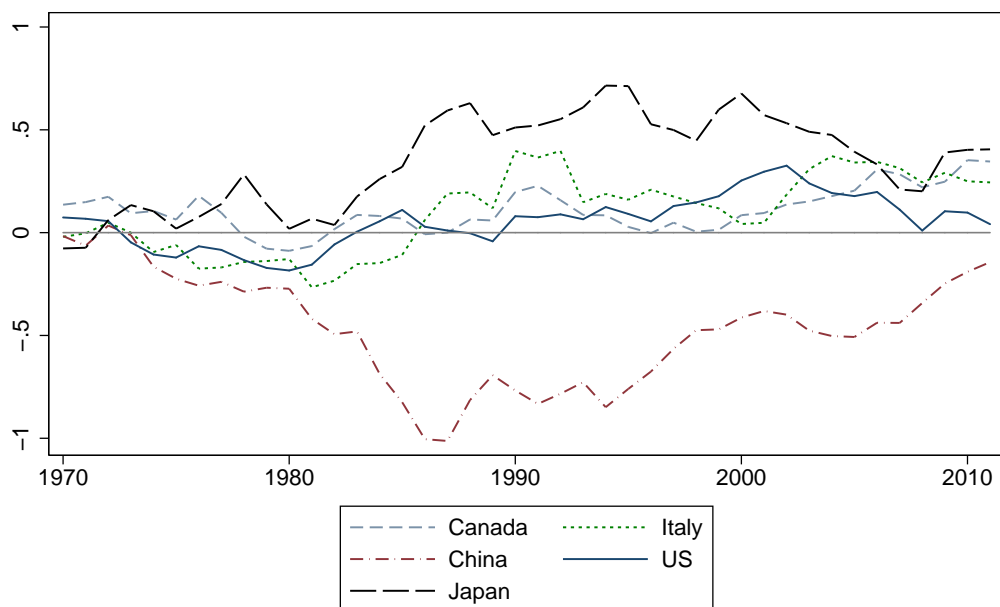


Figure 10: Variations of Balassa-Samuelson Adjusted WARP



exchange rate, which primarily arise from changes in the nominal exchange rate. Even sharp changes in the nominal exchange rate, which do tend to impact trade flows, do not necessarily affect the terms of trade, as a dollar appreciation could make imports and exports cheaper in dollars. Traditionally, for the US, it has been changes in oil prices which have had the largest affect on the terms of trade, while many users of RER indices are interested in the impact on trade ex-oil. For these reasons, the terms of trade are a less than ideal measure of competitiveness, while, for the US at least, there is generally no correlation between aggregate trade flows and the terms of trade.

Other possibilities would be to use producer or wholesale prices rather than consumer prices, which are what the Penn World Tables provides based on ICP data. Producer or wholesale prices, in fact, might be preferable *ceteris paribus* since, as Chinn (2006) notes, they exclude non-tradable retail services inputs, but the problem is comparable data availability. While one can get country-specific producer price indices for many countries, these series are indices and so do not contain necessary information on levels.²² And even if they did contain information on levels for a single year, when compared over long-time horizons with only a single benchmark year these indices will become biased over time, suffering from the same problem as versions of the Penn World Tables which predated v8.0. I am not aware of any large cross-country producer price comparison projects which seriously rival the ICP in terms of breadth and coverage.²³

Lastly, another possibility would be to focus on relative unit labor cost data for the manufacturing sector. This I explore in the next section.

5 Unit Labor Costs

5.1 Methodology

It is generally thought to be preferable on theoretical grounds to use wages rather than prices when measuring competitiveness, particularly for the manufacturing sector (Turner and Van't Dack, 1993). Intuitively, this is because labor is perhaps the largest non-traded input into the manufacturing sector, and labor tends to be at least somewhat immobile. Thus, the level of wages should naturally be one key determinate

²²I thank an anonymous referee for suggesting I use producer prices, and a discussant, Tadashii Ito, for suggesting using sectoral producer data collected by Sato *et al.* (2012). However, the Sato *et al* data covers just 7 years and 26 countries, and does not include data on relative price levels.

²³It is also unlikely that there are any other cross-country consumer price data which could seriously rival the ICP. For example, MIT's billion prices project does not contain information on quantities and thus also suffers from substitution bias.

for firms deciding where to produce. And, indeed, the levels of wages matter for competitiveness in almost all standard trade models, including Heckscher-Ohlin, the “Specific Factors” Model, the Melitz (2003) model, and in Ricardian Models of trade, including Eaton and Kortum (2002), among many others.

This explains the popularity of using unit labor costs in manufacturing to gauge competitiveness. Similar to the CPI-based REER produced using “divisia”, the ULC indices produced by the IMF and the OECD also are computed as indices of bilateral unit labor cost indices and thus also suffer from trading partner substitution bias. In addition, manufacturing PPP for each country is deflated using country-specific deflators, exactly as older versions of the Penn World Tables, which predated version 8.0, were computed. The series I propose is thus a simple Weighted Average of Relative Unit Labor Costs (WARULC) rather than of the unit labor cost *indices* – analogous to WARP in that it also solves the same index-numbers problem. In my series, I compute manufacturing PPP using PWT v8.0 methodology described in Feenstra *et al.* (2015). When I also expand the sample to include developing countries such as China and time-varying trade-weights, the differences in the underlying index become substantial. This is due to China’s systematically lower unit labor costs and growing weight in manufacturing trade over time.

The IMF’s RULC index, documented by Desruelle and Zanello (1997), is computed as:

$$I_{US,t}^{RULC} = \prod_{i=1} \left(\frac{C_{US}^I R_{US}^I}{C_i^I R_i^I} \right)^{w_i} \quad (5.1)$$

Where C_i^I is the normalized unit labor cost index for country i , computed as the ratio of nominal sectoral wages to real productivity, R_i is the nominal exchange rate index, and w_i are the time invariant trade weights. One intuitive proposed alternative specification would be to replace the unit labor cost indices with the same unindexed unit labor costs, and actual nominal exchange rates. However, the relative unit labor costs using deflated real productivity will depend crucially on the choice of the base year. To circumvent this problem, I convert nominal productivity into dollars using the PPP exchange rate conversion for the manufacturing sector, following the method Ceglowski and Golub (2007) implement for just the US and China, while converting nominal wages into dollars at the nominal exchange rate.²⁴

For this index, I used OECD data created specifically for the construction of ULC

²⁴Since the index is of *relative* unit labor costs, and as Rudiger Dornbusch used to say, two nominals make a real, the use of nominal wages converted at exchange rates is not problematic.

indices where available, and supplemented this with data from the BLS, the Chinese government, the World Bank's WDI, and UNIDOs. The manufacturing PPP data for benchmark years come from the relevant manufacturing ICP headings and were computed using PWT methodology (described in Section 3 of Feenstra *et al.* 2015), and were interpolated in the intervening years using country-specific deflators from either the OECD, or country-specific sources (in the case of China). The methodology and formulas for the PPP interpolation were also borrowed from Feenstra *et al.* (2015). *E.g.*, after the last benchmark year in 2005, the series are extended based on country growth rates for country i:

$$P_{i,2006} = P_{i,2005}^{ICP} \cdot \frac{P_{i,2006}^{i,deflator}}{P_{i,2005}^{i,deflator}}, \quad (5.2)$$

where $P_{i,t}^{i,deflator}$ is the country-specific deflator at time t.

For the years in between ICP benchmarks, a weighted average was used. For example, for the years between the ICP benchmarks in 1996 and 2005, the formula is:

$$P_{i,t} = P_{i,1996}^{ICP} \cdot \frac{P_{i,t}^{i,deflator}}{P_{i,1996}^{i,deflator}} \left(\frac{2005 - t}{2005 - 1996} \right) + P_{i,2005}^{ICP} \cdot \frac{P_{i,t}^{i,deflator}}{P_{i,2005}^{i,deflator}} \left(\frac{t - 1996}{2005 - 1996} \right). \quad (5.3)$$

Data on manufacturing trade to create trade weights $\omega_{i,t}$ is computed from bilateral manufacturing data at the SITC 4 level from Feenstra *et al.* (2005), and with updated data through 2008 via direct communication with Feenstra. Manufacturing trade data from 2009 and 2010 were taken from the OECD.

The Weighted Average Relative Unit Labor Cost index (WARULC) is computed as:

$$I_{US,t}^{WARULC} = \prod_{i=1} RULC_{US,i,t}^{\omega_{i,t}} = \prod_{i=1} \left(\frac{C_{US,t}}{C_{i,t}} \right)^{\omega_{i,t}} = \prod_{i=1} \left(\frac{\frac{w_{US,t}}{e_{US,t}} \frac{Y_{US,t}}{PPP_{US,t}}}{\frac{w_{i,t}}{e_{i,t}} \frac{Y_{i,t}}{PPP_{i,t}}} \right)^{\omega_{i,t}}. \quad (5.4)$$

Where $w_{i,t}$ are manufacturing wages of country i at time t, $e_{i,t}$ is the nominal exchange rate to convert to dollars, $Y_{i,t}$ is manufacturing production, and is divided by PPP for the manufacturing sector. Thus the C's in this equation are actual unit labor costs rather than indexes of unit labor costs.

5.2 Data

When the Weighted Average Relative Unit Labor Cost (WARULC) index is compared with the official IMF RULC index (indexed to start at the same value in 1975) and an index using my data but the IMF's index-of-indices method, the results are strikingly different, with the difference much larger than the disparity between WARP and divisia computed with CPIs. The series are roughly similar until the late 1980s, but by 2001, the WARULC index is 32% higher than the IMF's index, and 44% higher in 2008. The IMF benchmark index constructed here using the IMF's index-of-index method is similar to the IMF's index, despite the fact that I used time-varying manufacturing trade weights, a larger sample of countries (including China), and I compute manufacturing value-added using data on PPP. The IMF instead uses an index of real output measured in the home currency, and so it is striking that the benchmark is similar to both the IMF and the OECD indexes (the latter is not shown but also similar). I have also plotted a WARULC series which uses manufacturing PPP computed using only a single benchmark year and country deflators (short maroon dashes in Figure 12). The series without multiple benchmarks displays a downward trend relative to my preferred series with multiple benchmarks.

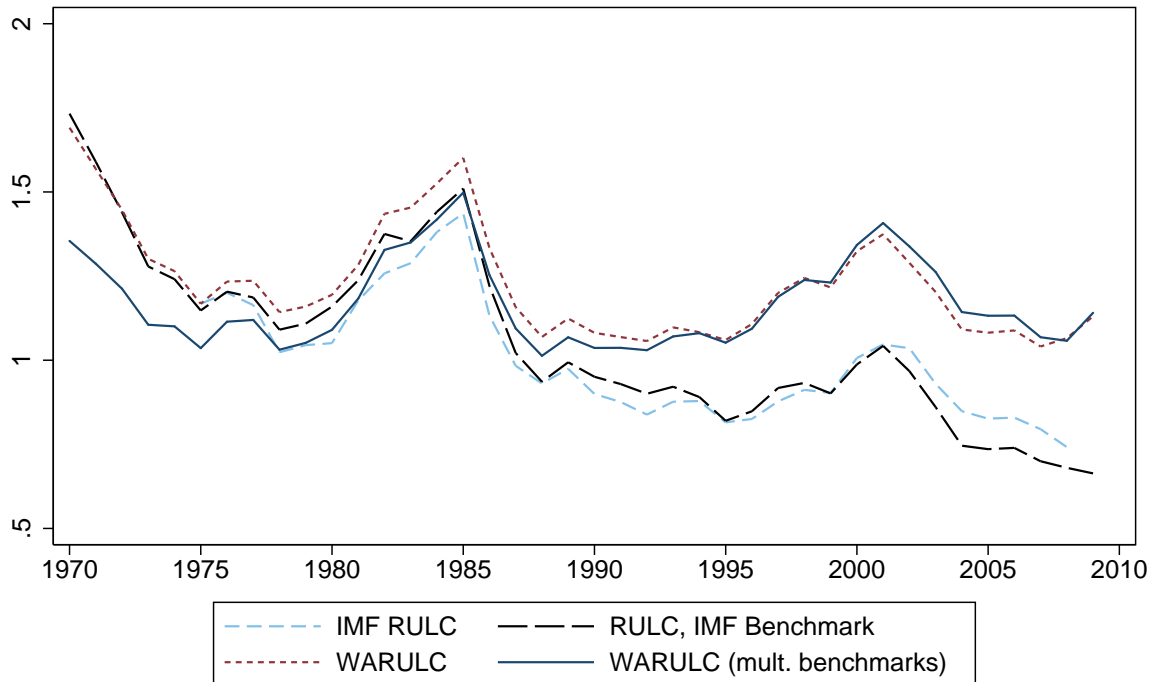


Figure 12: IMF Method vs. WARULC

And, just as with WARP, the difference between WARULC and the IMF's index is largely accounted for by the inclusion of China, as evidenced in Figure 13(a), as when China is dropped from WARULC, we arrive at something approximating the IMF's index. In 13(b) I compare WARP, P-WARP, and WARULC, and find that they are all broadly similar, with the exception being that the WARULC index displays a sharper depreciation after 2001, and that the P-WARP index implies that the dollar has generally been less over-valued as compared to WARP. Note that there aren't necessarily any pure theoretical grounds to prefer any of these three indices, as they all have their own strengths and weaknesses. Comparing WARP to P-WARP, one might prefer P-WARP on the grounds that it controls for productivity, but WARP on the grounds that it is more elegant since it does not require any faith in Balassa-Samuelson residuals. WARULC controls for productivity explicitly, and may be better designed for the manufacturing sector itself (though perhaps not for other tradable sectors), but given that it does not include other non-traded inputs, including services inputs, into manufacturing aside from labor, it can not be said to be a better measure than P-WARP *a priori*, even for the manufacturing sector.²⁵

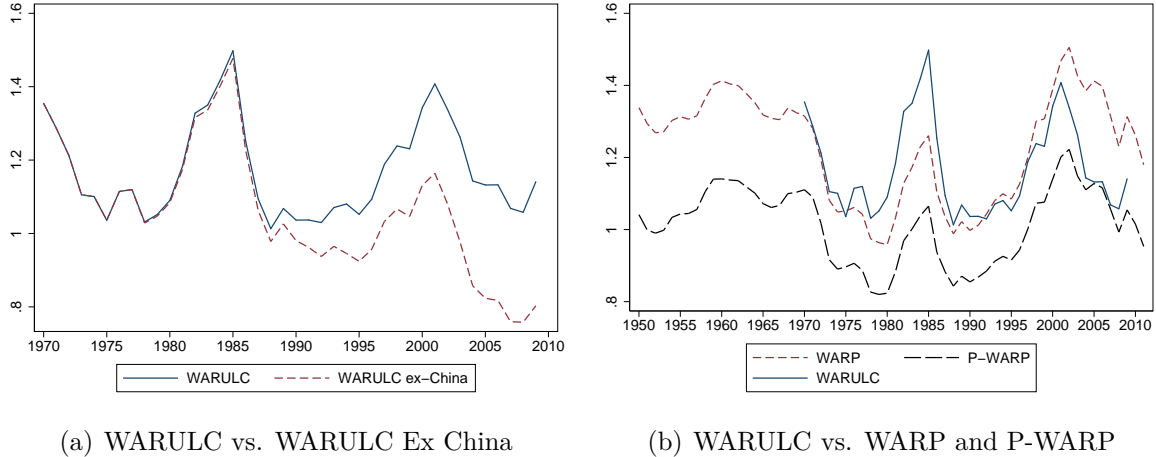


Figure 13: Comparing WARULC

Figure 14 details estimates of relative hourly productivity, wages, and ULCs for the US vs. China. These results are very similar to Ceglowski and Golub (2012) for the 1998-2009 period. The ratio of hourly wages has indeed fallen dramatically since the early 1990s, but not much more quickly than the convergence in productivity. Relative unit labor costs spiked in the late 1990s and early 2000s, during the collapse of US

²⁵A further discussion of the theoretical similarities between WARULC, P-WARP, and WARP is reserved for the additional Appendix, Section 10.2.

manufacturing employment concentrated heavily in China-competing industries. As of 2009, US manufacturing wages were still approximately 20 times larger than Chinese manufacturing wages, while unit labor costs in the US were about 2.6 times higher. In 2010, using data on just output and wages, with the strong assumption that hours worked stayed constant, I can estimate that the RULC did narrow substantially.

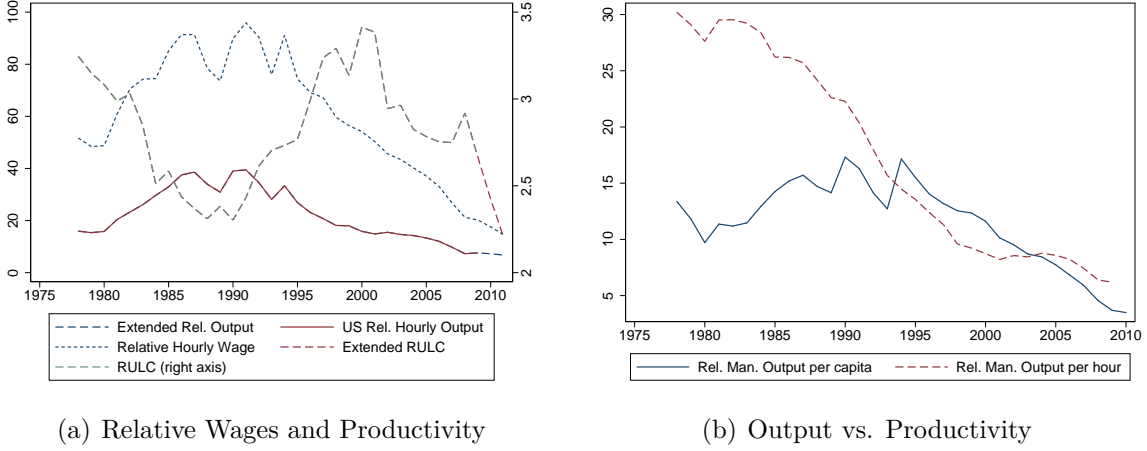


Figure 14: US vs. China

It is worth noting that the hiatus in the downward trend in relative productivity between China and the US during the 2000-2004 period (Figure 14(b)) came amidst a collapse in employment in manufacturing in the US biased toward low-productivity firms and sectors and was accompanied by a large growth in relative Chinese manufacturing output. This is apparent in Figure 14(b), where it is clear that relative manufacturing output *per capita* for the US vs. China fell at a steady pace during the period in which hourly relative output was stable. One solution to this problem would be to do ULCs by sector, but this would require at the very least sectoral output and wage data for China, and also PPP data and sectoral deflators for specific manufacturing sectors. In addition, it is likely that even within narrowly defined manufacturing sectors, low-productivity firms were more likely to lose workers, and even within firms, low-productivity workers were more likely to lose their jobs.

6 International Extensions

A key contribution of this paper is that I have also computed WARP and P-WARP indices for a total of 154 countries that had accurate PWT data, and using trade weights computed from each of the other 153 countries using trade data from the IMF DOTS,

and I have also computed WARULC indices for a handful of major economies (using the same data as for the US).²⁶ In this section I display some of the highlights.

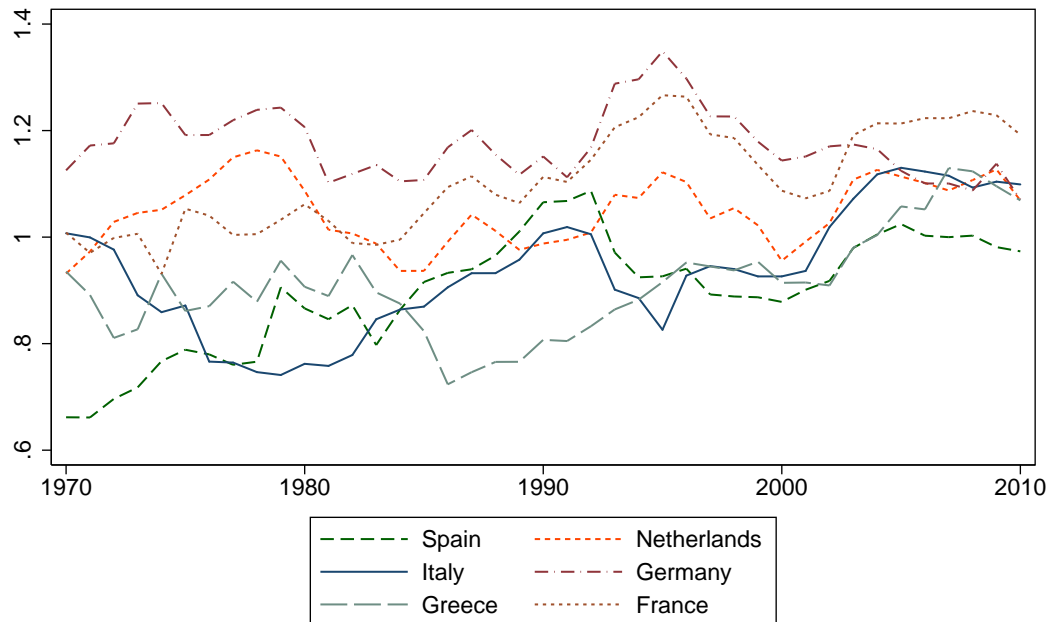


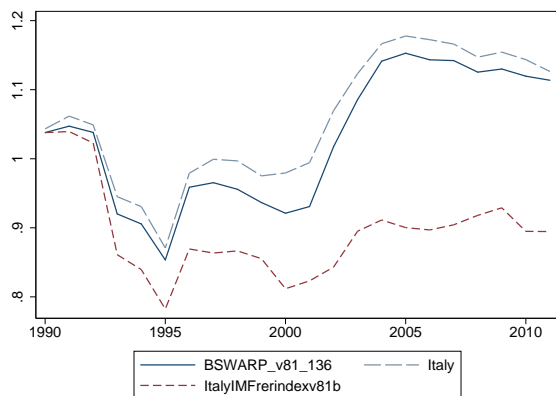
Figure 15: P-WARP, Major Euro Area Countries

First, in Figure 15, I show the results for the Penn-adjusted WARP for France, Germany, Greece, Italy, Spain and the Netherlands. This paints a picture of a relatively steady appreciation of the southern European countries vs. France and Germany since the mid-1990s, with Italy and Greece now having as strong an exchange rate as Germany does.²⁷

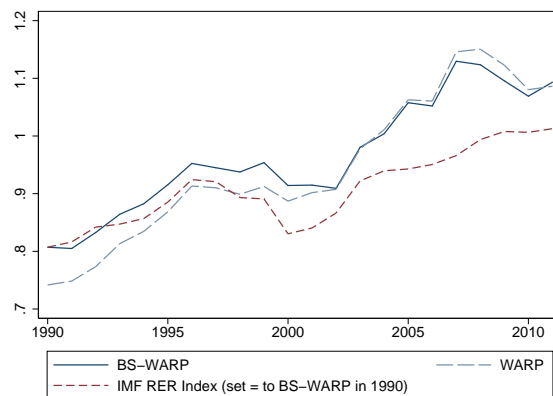
In Figure 30, I plot both the WARP and the Balassa-Samuelson adjusted series (which are similar to WARP for Europe) vs. the divisia-based indices produced by the IMF for a number of countries. The IMF and OECD indices are indexed to be equal to the P-WARP series in 1990. For Italy, Greece, the UK, Iceland, and China, the WARP and P-WARP indices reveal substantially greater appreciations than the IMF's RER

²⁶Note: originally, I computed WARP and P-WARP for 157 countries that have PWT v8.1 data and IMF DOTs data, but these indices revealed problems in some of the adjusted price levels in the PWT v8.1 and v8.0, and so I have excluded these observations. This indices have all been made publicly available for downloading on my webpage, www.dougcampbell.weebly.com. Additional graphs of comparisons between different vintages of the PWTs and different measures of WAR indices for many of these countries are available in the appendix.

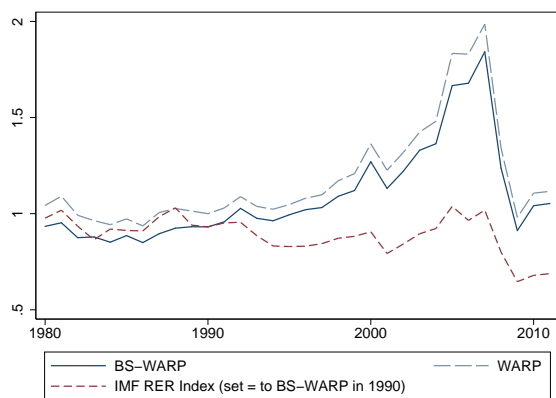
²⁷In the Additional Appendix Figure 28 I show that WARP implies a broadly similar picture, while the WARULC indices in Figure 29 suggest more dramatic appreciations for Spain and Italy relative to Germany, and also suggest that the Euro area as a whole had high relative ULCs going into the Financial Crisis and later Euro debt crisis.



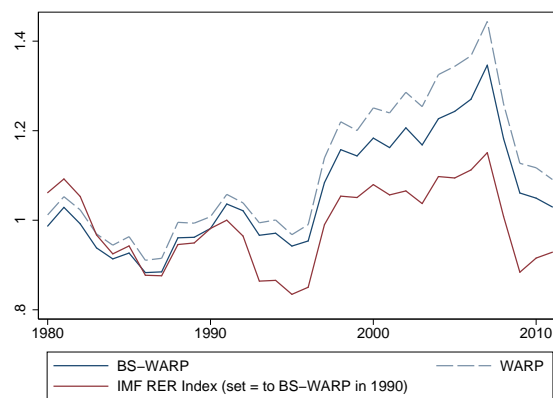
(a) Italy



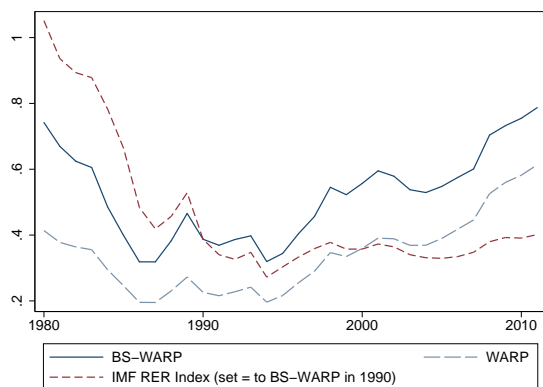
(b) Greece



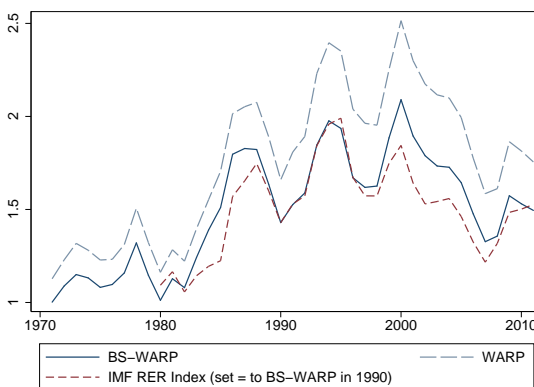
(c) Iceland



(d) UK



(e) People's Republic of China



(f) Japan

Figure 16: P-WARP and WARP vs. IMF Divisia RER Index

series, particularly going into the financial crisis of 2008-9.

As of 2011, the price level in the People's Republic of China was just 35% less than prices in China's trading partners, and only 21% less on a Balassa-Samuelson adjusted basis. Quite surprisingly, China's P-WARP has appreciated just as fast as its WARP index since the mid-1990s despite the fact that China's Balassa-Samuelson residual itself has been flat since around 2000. Hence, China's appreciation in this measure was driven instead by a shrinking in the Balassa-Samuelson residual of China's trading partners – most notably the United States.

Japan's WARP more than doubled from the early 1980s into the mid-1990s, outpacing even the swift appreciation of the IMF's trade-weighted REER index. For much of the past two decades, just as Japan has been immersed in a deflationary liquidity trap, its price level has been on average more than twice that of its trading partners. Japan's P-WARP index has mimicked the IMF's REER series much more closely, and implies that as of 2010, Japan's Balassa-Samuelson residual was still about 50% higher than the residuals of her trading partners.²⁸

7 Empirical Testing

The main point of this paper is to introduce new and improved measures of relative prices. I would argue that theoretical and intuitive concerns should dictate the choice of RER index, and that the choice should not necessarily depend on which index happens to predict trade flows the most closely, given that real exchange rate movements are not exogenous. Growing productivity should theoretically lead to an increase in exports and an appreciation of the real exchange rate, while events such as the Asian Financial Crisis might lead to sharp declines in both RERs and exports. However, in such cases there is unlikely to be a causal relationship. Such events make it unwise to try to use exchange rate indices to predict trade flows generally, and is probably why some observers, such as Rose (1991), have found that exchange rates do not effect trade for certain samples and time periods.²⁹

Thus, correcting the “trading partner substitution bias” problem and the problems arising from country-specific deflators, and the improvements in the trade weights –

²⁸These indices also rewrite the price history of a number of other countries in an interesting way. For example, in the period leading up to the Asian Financial Crisis, 1990-1996, the WARP index for Korea appreciated roughly 8% more than the OECD's index (although, as in Chinn 2000, these indices still imply that Korea was undervalued before the crisis).

²⁹In addition, Aizenman (2008), finds that current account changes have an impact on future relative prices, highlighting the likelihood of reverse causality.

adding in time-varying trade weights to the IMF’s RULC index, including China, and greatly increasing the sample of trading partners in the WARP and P-WARP indices – are the more persuasive reasons to prefer the WAR indices presented in this paper to previous indices produced by the IMF, the BIS, and the Federal Reserve. However, for WARP, BSWARP, and WARULC, I would view each series as having different strengths and weaknesses as they are different measures, and thus there is no one “best” index, particularly since the problems with endogeneity make any test vs. the data inherently a heroic task. Nevertheless, there appears to be interest in whether these indices can do a better job predicting trade flows, and so in this section I explore the issue.

While the problem of endogeneity generally makes the task of testing the indices a difficult and perhaps fruitless task, in the US case, it seems likely that much of the large RER movements since the end of the Bretton Woods period have happened due to budget deficits driven by political events.³⁰ Thus it can be argued that, to a large extent, the RER movements really have been driven by factors which could be plausibly argued to be exogenous from the perspective of trade flows ex-oil (or, perhaps, relatively more plausibly exogenous than for many other countries).

In the US case, RER movements and trade balances do seem to go together.³¹ Thomas *et al.* (2008) have already shown that relative price indices based on WARP do a better job predicting US trade imbalances than do divisia-based indices using CPIs. That this result holds up for WARP computed with more recent vintages of the PWT is clear from Figure 1(a).³² Indeed, this period is the strongest evidence and the original motivation for using WAR exchange rates. For these reasons, I first consider the US and then look at additional international evidence.

7.1 Testing with US Data

First, in Figure 27 I show that the level of WARULC seems to do a reasonably good job of predicting changes in the import share of manufacturing trade (imports divided

³⁰Chinn and Ito (2008) detail the importance of budget deficits for the current account.

³¹For example, Chinn (2004) finds that exchange rates have an impact on US exports, though not imports. Another example of sharp, plausibly exogenous movements in RERs is Canada in the past 10 years, as oil prices increased and the US dollar weakened in the mid-2000s. Due to space constraints, I consider Canada in the Additional Appendix, in Table VII.

³²While it would be nice to extend this evidence using WARP for the period 1820 to 1970, most of the large movements in relative prices during this earlier period as well were associated with major wars or the Great Depression, and so exchange rates were unlikely to be the major determinate of trade flows as they arguably were in the post-Bretton Woods period.

by total trade) not due to changes in GDP.³³ The IMF's RULC index, by contrast, implies a steadily more competitive US manufacturing sector over time, which seems to be at odds with the large import share of trade the US has experienced since the late 1990s, and at odds with the realized collapse in US manufacturing in the early 2000s concentrated in sectors more exposed to international trade (Campbell 2015) and Chinese import competition (Autor *et al.* 2013).

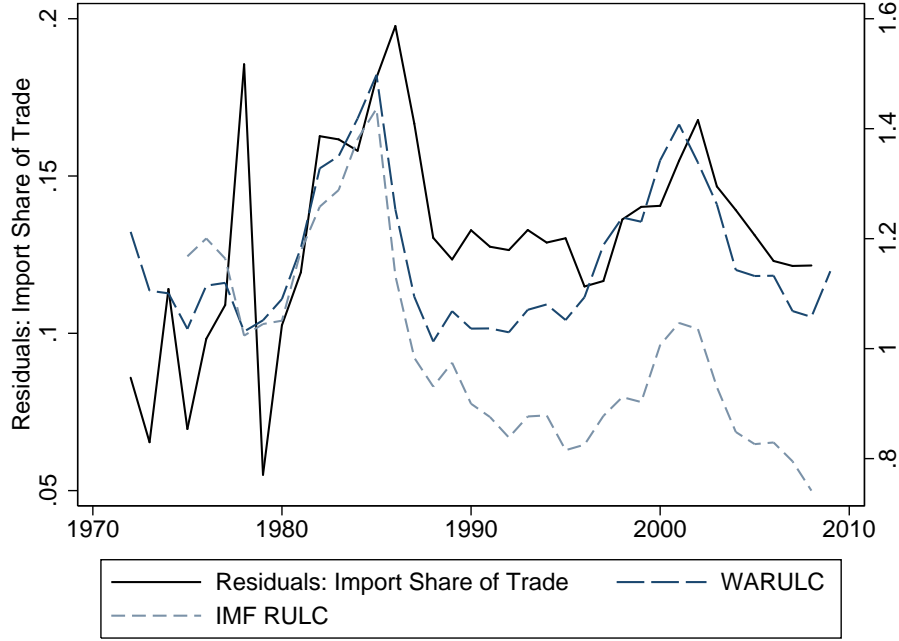


Figure 17: RULC Indices vs. Import Share of Manufacturing Trade

Related evidence is offered in a companion paper, by Campbell (2015), which argued that the class of Weighted-Average Relative (WAR) price indices developed in this paper can predict declines in more open manufacturing sectors relative to less open manufacturing sectors for the US and Canada, which, he argues, are less exposed to international trade. I confirm and strengthen this evidence, by showing (Figure 18) that an import-Weighted Average Relative Unit Labor Cost (iWARULC) index does a remarkably good job predicting years when manufacturing sectors with a higher share of import-penetration suffered declines in employment.³⁴ The coefficient on import penetration, plotted with two standard deviation error bounds, came from the regression:

³³*I.e.*, they do a good job of explaining the residuals of regression: $M_{US,t} = \alpha + \rho M_{US,t-1} + \beta_0 \ln \Delta RGDPO_t + \beta_1 \ln \Delta RGDPO_t^* + \beta_2 I_{t-1}^j + \epsilon_t$, with the lagged value of WARULC times the coefficient on the lagged value of WARULC added back in.

³⁴This index is constructed in the same manor as WARULC, only using only import weights instead of overall trade weights.

$$\ln(L_{ht}/L_{h,t-1}) = \alpha_t + \beta_0 MPPen_{h,t-1} + \beta_2 \ln(D_{h,t}/D_{h,t-1}) + \beta_3 \ln((TFP)_{h,t}/(TFP)_{h,t-1}) + \epsilon_{ht}, \quad (7.1)$$

where h is one of 353 balanced 4-digit SIC manufacturing industries (data from the BEA's Annual Survey of Manufactures), and the years span 1973 to 2009. $MPPen$ is import penetration, defined as imports divided by "domestic consumption", which is equal to domestic production plus imports minus exports. D_{ht} is defined as demand in sector h at time t , and TFP is a measure of 4-factor productivity (the details of this measure, which come from the NBER-CES manufacturing database, are described by Bartelsman and Gray (1996)).

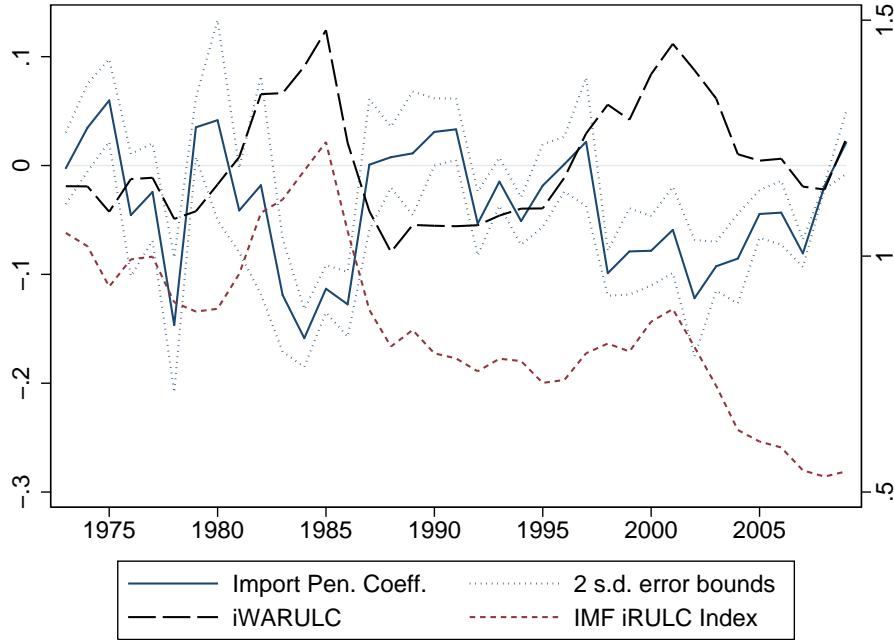


Figure 18: The Coefficient on Import Penetration vs. Two Measures of RULCs

A relative unit labor cost (RULC) index derived using the IMF's data and methods, except using time-varying import weights (dotted red line in Figure 18), implies that US RULCs in manufacturing were no higher in the early 2000s than they had been in the late-1980s or late-1970s, periods which were not associated with adverse performance of import-competing manufacturing sectors. By contrast, the import-Weighted Average Relative Unit Labor Cost (iWARULC) implies that US RULCs were very high from the late-1990s to the mid-2000s, a period associated with a relatively worse performance of more open manufacturing sectors (this graph shows employment, but this is also true for

value-added and labor productivity). In addition, since RER movements are observed to impact trade with a lag of about one year, the apex of the iWARULC index in 2001 correctly predicts the year of the worst employment performance in import-competing manufacturing employment.

Next, I postulate a simple model based on intuition and a standard theoretical model to predict the log of the ratio of imports to exports, ex-oil.³⁵

$$\ln(X/M)_{US,t} = \alpha + \rho \ln(X/M)_{US,t-1} + \beta_0 \ln \Delta RGDP O_t + \beta_1 \ln \Delta TPMD_t^* + \beta_2 I_{t-1}^j + \epsilon_t, \quad (7.2)$$

j = Various RER indices.³⁶

This equation conjectures that the log ratio of exports to imports ($X/M_{US,t}$) grows when home GDP growth is higher ($RGDP O_t$), falls when a weighted average of trading partner import demand is faster ($TPMD_t^*$), and rises when the lagged level of the real exchange rate is higher (I_{t-1}^j). Trading partner import demand was computed from using the predicted changes in trade from a regression of the log change of (total) imports ex-oil on the log changes in real GDP and a constant for 154 countries separately (since imports are generally procyclical), and then taking a trade-weighted average.³⁷

I estimate this equation for the US for the period 1976 to 1996, and then use the coefficients to test this model out of sample for the 1997 to 2008 period. The results for the in-sample period are displayed in Table I, which illustrates how various measures of the RER compare in predicting US trade flows. All are statistically significant predictors with the exception of the terms of trade, which is not significant and of the wrong sign. The t-scores and the R-squareds were the largest for the P-WARP index, followed by WARP, which are both much larger than for the Fed's index. Of the RULC indices, the WARULC index has a higher R-squared than the IMF's RULC index. I would consider this as weak suggestive evidence that, at least for the US, the WAR indices seem to do a better job of predicting trade flows in sample.³⁸

In Table II and Figure 19, I extrapolate the predictions from the regressions in

³⁵The solution to a CES expenditure function (*e.g.*, from Feenstra (2004) used to derive a gravity equation: $x_{ij} = \frac{Y_j p_i^{1-\sigma} \tau_{ij}^{1-\sigma}}{P_j^{1-\sigma}}$, where x_{ij} are exports from i to j, Y_j is GDP, τ_{ij} are trade costs, and p_i is the price level in country i, while P_j is the CES price index for country j. This equation tells us that prices and GDP matter for trade flows. Thus, home and foreign GDP, and the relative price levels would matter for the import share of total trade.

³⁶Several Notes: (1) I start in 1976 for comparison purposes because this is when the IMF's index begins. (2) There are not enough observations to do an error correction model here. (3) Trading partner import demand could be replaced by a weighted average of trading partner GDP growth and nothing would change here.

³⁷*I.e.*, the regression was $\ln(M_t) = \alpha + \beta \ln(RGDP_t/RGDP_{t-1}) + \epsilon_t$ run separately for each of 154

Table I: US Log Exports/Imports, 1976-1996, In-Sample Predictions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
L.ln(Exports/Imports)	0.75*** (0.14)	0.53*** (0.090)	0.67*** (0.15)	0.52*** (0.061)	0.53*** (0.060)	0.71*** (0.085)	0.55*** (0.076)
ln Δ GDP	-2.04*** (0.50)	-1.31*** (0.35)	-1.84*** (0.62)	-0.79** (0.28)	-0.74** (0.29)	-1.64*** (0.39)	-1.36*** (0.33)
Foreign Import Dem.	0.80 (0.80)	0.14 (0.48)	0.75 (0.66)	0.14 (0.27)	0.014 (0.28)	0.62 (0.52)	-0.036 (0.46)
L.Terms of Trade	0.049 (0.38)						
L.Fed's RER Index		-0.54*** (0.088)					
L.Balassa-Sam. Resid.			-0.19 (0.15)				
L.WARP				-0.80*** (0.074)			
L.P-WARP					-0.98*** (0.092)		
L.IMF RULC						-0.21*** (0.058)	
L.WARULC							-0.39*** (0.059)
Constant	-0.13 (0.31)	0.42*** (0.092)	-0.11* (0.056)	0.73*** (0.075)	0.78*** (0.081)	0.14* (0.077)	0.36*** (0.078)
Observations	21	21	21	21	21	21	21
R^2	0.778	0.915	0.793	0.958	0.959	0.868	0.915

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parenthesis. The dependent variable is the log exports/imports. Foreign import demand is the trade-weighted average of predictions of import demand growth due to changes in GDP.

Table I to the out-of-sample period (1997-2008), and compute the mean absolute error, computed as $MAE_i = (1/T) \sum_t ||\epsilon_{it}||$. I also regress the predictions on the actual $\ln(\text{exports}/\text{imports})$, and report the R-squared from each of these regressions while suppressing the coefficients. Thus, the R-squared in Column (1) of Table II reports the R-squared from a regression of the out-of-sample predictions from Column (1) of Table I, which uses the Terms of Trade as the RER index of choice.

I find that the model with the P-WARP index (the WARP index adjusted for the Penn Effect) validates the best in the sense that this model has the lowest mean absolute error in the out-of-sample period. The model run with the Balassa-Samuelson residual also happens to do well, although the R-squared from these predictions do much worse. WARP does very well in terms of predicting changes in the trade balance, but it seems to overpredict the level of the trade deficit (Figure 19). Once again, WARULC has a much lower mean absolute error than the IMF's RULC index, although, admittedly, the R-squared is lower. It is also worth noting that these indices are designed for the manufacturing sector, and not trade as a whole, so it might be expected that they would perform less well than the others.

Table II: US Imports/Exports, 1997-2008, Out-of-Sample Statistics

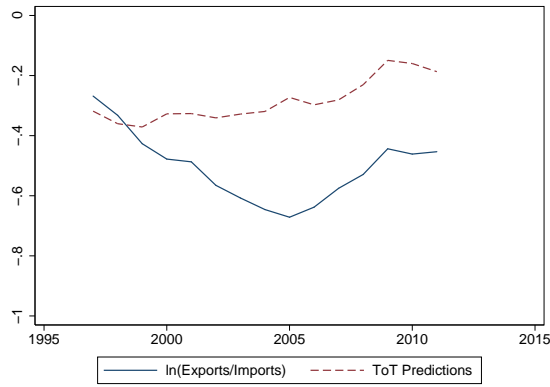
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	ToT	Fed	BS Resid.	WARP	PWARP	IMF	WARULC	None
R^2	0.327	0.505	0.534	0.950	0.932	0.307	0.210	0.337
Mean Absolute Error	0.210	0.190	0.133	0.198	0.131	0.292	0.162	0.211

The dependent variable is the log of exports/imports. The explanatory variables are the predicted values extrapolated for the period 1997-2008 from Table I, using the predictions as the lagged dependent variable in the subsequent year. The mean absolute error is simply the average absolute difference between the actual and the predicted values. The last column contains the predictions from a model with no real exchange rate.

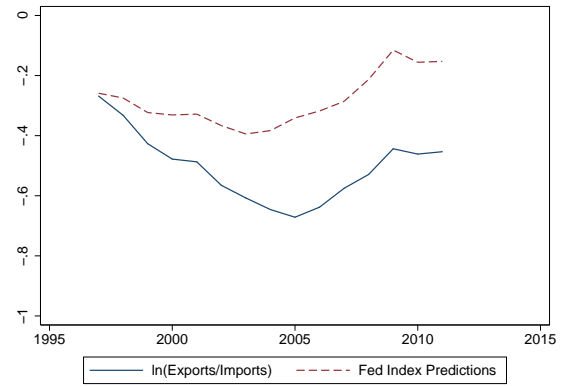
Overall, this subsection provides strong evidence that, for the US at least, the WAR-indices seem to be better predictors of the trade balance and the collapse in import-competing manufacturing employment than do their indices-of-indices counterparts. However, as mentioned before, given the problems with identification, theoretical and intuitive considerations should generally dictate the choice of a RER index.

countries.

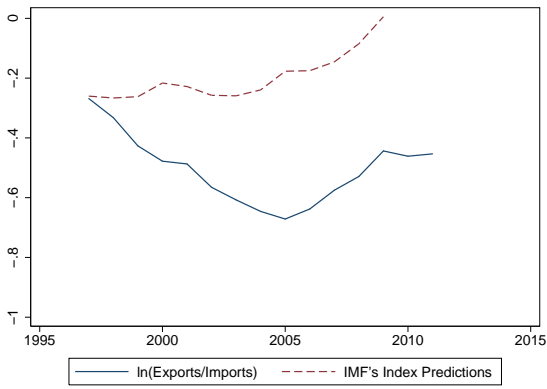
³⁸This simple estimation strategy based on intuitive priors is not without trouble, as, for example, the coefficients on the constant and occasionally the foreign import demand are quite unlikely to hold up out of sample and are unlikely to be economically meaningful. An additional problem with using a lagged dependent variable is Nickell (1981) bias, where the bias on the lagged dependent variable will be equal to $(1+\rho)/(T-1)$, which should be on the order of 10% with $t=20$ in this case, which is a bit worrisome. However, the results are unchanged if we expand the sample to 1970-2008, in which case the Nickell bias would shrink to less than 5%.



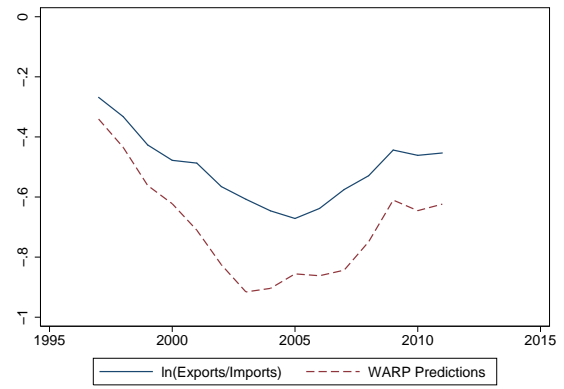
(a) Terms of Trade



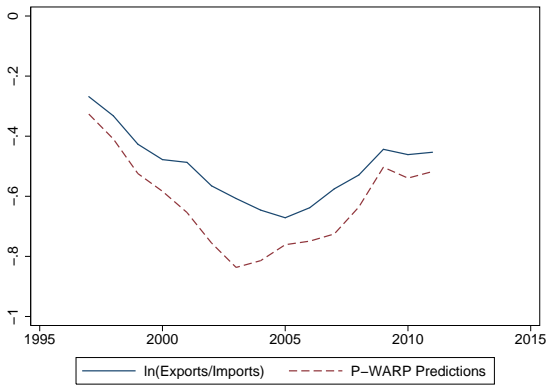
(b) Fed's RER Index



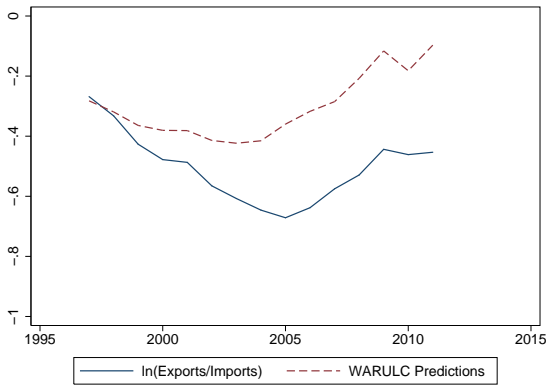
(c) IMF RULC Index



(d) WARP



(e) P-WARP



(f) WARULC

Figure 19: Out-of-Sample Testing

7.2 International Testing

The US case looks fairly straightforward, but how well do WAR indices, or RER indices in general, predict trade flows internationally? As mentioned, the US, with its relatively sharp exchange rate movements driven by factors likely to be largely exogenous from the perspective of tradable sectors of the economy, appears to be anomalous. I find that, in general, there does not appear to be any tight, consistent correlation between any real exchange rate indices and trade flows.

As we now have panel data, I fit a random coefficients model internationally to the sample of countries for which I computed WARULC indices during the period in which there exist comparison IMF indices. I estimate:

$$(X/M)_{i,t} = \beta_{0,i} + \beta_{1,i}(X/M)_{i,t-1} + \beta_{2,i}\ln\Delta RGDPO_t + \beta_{3,i}\ln\Delta TPMD_t^* + \beta_{4,i}I_{t-1}^j + \epsilon_t, \quad (7.3)$$

as before, with the difference being that now each coefficient is comprised of both a constant and of a stationary mean zero random variable: $\beta_{ik} = \beta_k + \xi_{ik}$, $i = 1, \dots, N$, $k = 0, 1, 2$, where N is the number of countries, in this case seven, and k is the number of regressors. The rationale behind the random coefficients model is essentially that countries are likely to be different, and so it is desirable to let the coefficients vary, but that data on one set of countries likely contains useful information for others, in which case a random coefficients model would be superior to running the simple model for each country separately. Otherwise, this setup is roughly similar to the panel data approaches in Chinn and Prasad (2004), Gruber and Kamin (2007), and Aizenman *et al.* (2014).

The results can be seen in Table III, where the country-specific random effects are omitted for space. None of the RER indices appear to be correlated with trade flows, while “Foreign Import Demand” is significant and has the wrong sign, while domestic GDP growth – a powerful predictor of import demand in the US – is not even significant. These results also hold if we use the full sample of 154 countries for which I created terms of trade and WAR indices for (see Table V in the Appendix), or if we employed a panel error-correction model instead.

Table III: Predicting Imports/Exports, 1976-2008 (6 countries), Random Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
L1.ln(Exports/Imports)	0.91*** (0.027)	0.91*** (0.027)	0.92*** (0.026)	0.91*** (0.029)	0.91*** (0.028)	0.87*** (0.034)	0.91*** (0.027)
ln Δ GDP	-0.17 (0.22)	-0.17 (0.21)	-0.19 (0.23)	-0.16 (0.21)	-0.16 (0.21)	-0.15 (0.20)	-0.16 (0.24)
Foreign Import Demand	-0.33* (0.17)	-0.33** (0.17)	-0.32* (0.16)	-0.32* (0.17)	-0.33** (0.17)	-0.33* (0.17)	-0.34** (0.17)
L1.ToT	-0.030 (0.089)						
L1.IMF RER Index (CPI)		-0.0049 (0.042)					
Balassa-Sam. Resid.			0.0029 (0.023)				
L1.WARP v8.1				-0.014 (0.030)			
L1.P-WARP v8.1					-0.0083 (0.033)		
L1.IMF RULC Index						-0.025 (0.022)	
L1.WARULC							-0.0052 (0.028)
Observations	182	182	182	182	182	182	182
Log Likelihood	306.5	305.7	305.3	305.3	305.4	305.5	306.5

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log of exports/imports. Foreign import demand is the trade-weighted average of predictions of import demand growth due to changes in GDP. The six countries in the sample are France, Germany, Italy, Netherlands, Canada and Spain. The random part of the coefficients is suppressed for space.

8 Conclusion

In this paper, I sought to improve the methodology used to create trade-weighted real exchange rate indices, proposing several new series, including a Penn adjusted Weighted Average Relative Price index (P-WARP), and a Weighted Average Relative Unit Labor Cost index (WARULC), which are designed to solve the trading partner substitution bias problem and control for productivity as per the Penn Effect. I provide the WARP and P-WARP indices for 154 countries, and WARULC for 7 major economies, and find that these indices indicate a seriously warped view of economic history for many nations. I demonstrate that, for several countries currently mired in liquidity traps, including Japan, the US, and for the Euro area, prices and wages had grown to high levels relative to trading partners heading into the financial crisis of 2008. For the US, I demonstrate that in 2002 – during the surprisingly swift collapse of US manufacturing – relative prices had not been that high since the worst year of the Great Depression, when the golden straitjacket led to an appreciated currency, collapsing commodity prices, rising real debt levels and a financial and economic crisis. Anyone wishing to understand the past 20 years of economic history would do well to start with an investigation of relative prices.

References

- AIZENMAN, J., “Relative Price Levels and Current Accounts: An Exploration,” *Department of Economics, UCSC* (2008).
- AIZENMAN, J., Y. JINJARAK AND N. P. MARION, “The People’s Republic of China’s Growth, Stability, and Use of International Reserves,” ADBI Working Papers 455, Asian Development Bank Institute, January 2014.
- ALLEN, R. C., J.-P. BASSINO, D. MA, C. MOLL-MURATA AND J. L. VAN ZANDEN, “Wages, Prices, and Living Standards in China, 1738–1925: In Comparison with Europe, Japan, and India,” *The Economic History Review* 64 (2011), 8–38.
- AUTOR, D., D. DORN AND G. H. HANSON, “The China Syndrome: The Local Labor Market Effects of Import Competition in the US,” *American Economic Review* 103 (2013), 2121–68.
- BARBIERI, K., *The Liberal Illusion: Does Trade Promote Peace?* (University of Michigan Press, 2002).
- BAYOUMI, T., J. LEE AND S. JAYANTHI, “New Rates From New Weights,” *IMF Staff Papers* (2006), 272–305.

- BAYOUMI, T., M. SAITO AND J. TURUNEN, “Measuring Competitiveness: Trade in Goods or Tasks?,” (2013).
- BEMS, R. AND R. C. JOHNSON, “Value-Added Exchange Rates,” Working Paper 18498, National Bureau of Economic Research, October 2012.
- BERGIN, P. R., R. GLICK AND A. M. TAYLOR, “Productivity, Tradability, and the Long-Run Price Puzzle,” *Journal of Monetary Economics* 53 (2006), 2041–2066.
- BERNANKE, B. S., “The Global Saving Glut and the US Current Account Deficit,” Technical Report, 2005.
- BORDO, M., B. EICHENGREEN, D. KLINGEBIEL AND M. S. MARTINEZ-PERIA, “Is the Crisis Problem Growing More Severe?,” *Economic Policy* 16 (04 2001), 51–82.
- BORDO, M. D., E. U. CHOUDHRI, G. FAZIO AND R. MACDONALD, “The Real Exchange Rate in the Long Run: Balassa-Samuelson Effects Reconsidered,” NBER Working Papers 20228, National Bureau of Economic Research, Inc, June 2014.
- CAMPBELL, D. L., “Relative Prices, Hysteresis, and the Decline of American Manufacturing,” Working Papers w0212, Center for Economic and Financial Research (CEFIR), July 2015.
- CATAO, L. AND S. SOLOMOU, “Effective Exchange Rates 1879–1913,” *European Review of Economic History* 4 (2000), 361–382.
- CEGLOWSKI, J. AND S. GOLUB, “Just How Low are China’s Labour Costs?,” *The World Economy* 30 (2007), 597–617.
- CEGLOWSKI, J. AND S. S. GOLUB, “Does China Still Have a Labor Cost Advantage?,” *Global Economy Journal* 12 (2011).
- , “Does China Still Have a Labor Cost Advantage?,” *Global Economy Journal* 12 (2012).
- CHEUNG, Y.-W., M. D. CHINN AND E. FUJII, “China’s Current Account and Exchange Rate,” in *NBER Volume: China’s Growing Role in World Trade* (University of Chicago Press, 2010), 231–271.
- CHINN, M. D., “Before the Fall: Were East Asian Currencies Overvalued?,” *Emerging Markets Review* 1 (2000), 101–126.
- , “Incomes, Exchange Rates and the US Trade Deficit, Once Again*,” *International Finance* 7 (2004), 451–469.
- , “A Primer on Real Effective Exchange Rates: Determinants, Overvaluation, Trade Flows and Competitive Devaluation,” *Open Economies Review* 17 (2006), 115–143.

- CHINN, M. D. AND H. ITO, "Global Current Account Imbalances: American Fiscal Policy versus East Asian Savings*," *Review of International Economics* 16 (2008), 479–498.
- CHINN, M. D. AND E. S. PRASAD, "Medium-Term Determinants of Current Accounts in Industrial and Developing Countries: An Empirical Exploration," *Journal of International Economics* 59 (2003), 47–76.
- DESRUELLE, D. AND A. ZANELLO, "A Primer on the IMFs Information Notice System," *IMF Working Paper No. 97/71* (1997).
- DEVEREUX, J. AND M. WARD, "Measuring British Decline: Direct Versus Long-span Income Measures," *The Journal of Economic History* 63 (2003), 826–851.
- , "Relative British and American Income Levels During the First Industrial Revolution," *Research in Economic History* 23 (2006), 249–286.
- DIEWERT, W. E., A. O. NAKAMURA, J. S. GREENLEES, L. I. NAKAMURA AND M. B. REINSORF, "Sourcing Substitution and Related Price Index Biases," (2014).
- EATON, J. AND S. KORTUM, "Technology, Geography, and Trade," *Econometrica* 70 (2002), 1741–1779.
- FAHLE, S., J. MARQUEZ AND C. THOMAS, "Measures of International Relative Prices for China and the USA," *Pacific Economic Review* 14 (2009), 376–397.
- FEENSTRA, R. C., *Advanced International Trade: Theory and Evidence* (Princeton University Press, 2003).
- FEENSTRA, R. C., R. INKLAAR AND M. TIMMER, "The Next Generation of the Penn World Table," *American Economic Review, Forthcoming* (2015).
- FEENSTRA, R. C., R. E. LIPSEY, H. DENG, A. C. MA AND H. MO, "World Trade Flows: 1962-2000," Technical Report, National Bureau of Economic Research, 2005.
- FEENSTRA, R. C., H. MA, J. PETER NEARY AND D. PRASADA RAO, "Who Shrunk China? Puzzles in the Measurement of Real GDP," *The Economic Journal* 123 (2013), 1100–1129.
- FRANKEL, J., "On the Yuan: The Choice Between Adjustment Under a Fixed Exchange Rate and Adjustment Under a Flexible Rate," *CESifo Economic Studies* 52 (2006), 246–275.
- FRATZSCHER, M., A. MEHL AND I. VANSTEENKISTE, "130 Years of Fiscal Deficits and Currency Crashes in Advanced Economies," (2011).
- GLÜZMANN, P. A., E. LEVY-YEYATI AND F. STURZENEGGER, "Exchange Rate Undervaluation and Economic Growth: Díaz Alejandro (1965) Revisited," *Economics Letters* 117 (2012), 666–672.

- GOLUB, S. S. AND A. G. TURNER, "Towards a System of Multilateral Unit Labor Cost-Based Competitiveness Indicators for Advanced, Developing, and Transition Countries," Technical Report, International Monetary Fund, 1997.
- GRUBER, J. W. AND S. B. KAMIN, "Explaining the Global Pattern of Current Account Imbalances," *Journal of International Money and Finance* 26 (2007), 500–522.
- HO, L. S., "Globalization, Exports, and Effective Exchange Rate Indices," *Journal of International Money and Finance* 31 (2012), 996–1007.
- HOUSEMAN, S., C. KURZ, P. LENGERMANN AND B. MANDEL, "Offshoring Bias in U.S. Manufacturing," *Journal of Economic Perspectives* 25 (Spring 2011), 111–32.
- INKLAAR, R., "Import Sourcing Bias in Manufacturing Productivity Growth: Evidence Across Advanced and Emerging Economies," In *S. N. Houseman, and M. Mandel (Eds.), Measuring globalization: Better Trade Statistics for Better Policy. (Vol. 1, pp. 195-217). Kalamazoo MI: W.E. Upjohn Institute for Employment Research* (2015).
- JACKS, D., C. MEISSNER AND D. NOVY, "Trade Booms, Trade Busts, and Trade Costs," *Journal of International Economics* (2006).
- JOHNSON, S., J. D. OSTRY AND A. SUBRAMANIAN, "Prospects for Sustained Growth in Africa: Benchmarking the Constraints," *IMF Staff Papers* 57 (2010), 119–171.
- KLAU, M. AND S. S. FUNG, "The New BIS Effective Exchange Rate Indices," *BIS Quarterly Review*, March (2006).
- LEVY-YEYATI, E., F. STURZENEGGER AND P. A. GLUZMANN, "Fear of Appreciation," *Journal of Development Economics* 101 (2013), 233–247.
- LORETAN, M., "Indexes of the Foreign Exchange Value of the Dollar," *Fed. Res. Bull.* 91 (2005), 1.
- MELITZ, M. J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica* 71 (November 2003), 1695–1725.
- NICITA, A., "Exchange Rates, International Trade and Trade Policies," *International Economics* 135 (2013), 47–61.
- NICKELL, S., "Biases in Dynamic Models with Fixed Effects," *Econometrica: Journal of the Econometric Society* (1981), 1417–1426.
- OBSTFELD, M., K. S. ROGOFF AND S. WREN-LEWIS, *Foundations of International Macroeconomics*, volume 30 (MIT press Cambridge, MA, 1996).
- PIERCE, J. R. AND P. K. SCHOTT, "The Surprisingly Swift Decline of U.S. Manufacturing Employment," Working Paper 18655, National Bureau of Economic Research, April 2015.

- REINSdorf, M. AND R. YUSKAVAGE, “Offshoring, Sourcing Substitution Bias and the Measurement of US Import Prices, GDP and Productivity,” (2014).
- REINSdorf, M. B., “Formula Bias and Within-Stratum Substitution Bias in the US CPI,” *Review of Economics and Statistics* 80 (1998), 175–187.
- RODRIK, D., “The Real Exchange Rate and Economic Growth,” *Brookings Papers on Economic Activity* 2008 (2008), 365–412.
- ROSE, A. K., “The Role of Exchange Rates in a Popular Model of International Trade: Does the ‘Marshall–Lerner’ Condition Hold?,” *Journal of International Economics* 30 (1991), 301–316.
- SATO, K., J. SHIMIZU, N. SHRESTHA AND S. ZHANG, “Industry-Specific Real Effective Exchange Rates for Japan,” Technical Report, RIETI Discussion Paper, 12-E-044, 2012.
- SCHMITZ, M., M. DE CLERCQ, M. FIDORA, B. LAURO AND C. PINHEIRO, “Revisiting the Effective Exchange Rates of the Euro,” *ECB Occasional Paper* (2011).
- SKIDELSKY, R., *John Maynard Keynes: Volume 2: The Economist as Savior, 1920–1937*, volume 2 (Penguin Books, 1995).
- STAHLER, K. AND A. SUBRAMANIAN, “Versailles Redux? Eurozone Competitiveness in a Dynamic Balassa-Samuelson-Penn Framework,” *Journal of Globalization and Development* (2014).
- TAYLOR, A. M., “A Century of Purchasing-Power Parity,” *Review of Economics and Statistics* 84 (2002), 139–150.
- THOMAS, C., J. MARQUEZ AND S. FAHLE, “Measuring US International Relative Prices: A WARP View of the World,” *FRB International Finance Discussion Paper* (2008).
- THOMAS, C., J. MARQUEZ ET AL., “Weighted Average Relative Price (WARP),” *IFC Bulletins chapters* 36 (2013), 340–350.
- TURNER, P. AND J. VAN’T DACK, “Measuring International Price and Cost Competitiveness,” *Bank for International Settlements, Monetary and Economic Department* (1993).

9 Data Appendix

Table A.1: Historical Relative Price Levels (vs. US)

Year	<u>UK</u>		<u>Japan</u>	
	Ward and Devereux	Our Estimates	Allen <i>et al.</i>	Our Estimates
1831	156	143.9		
1839	122	134.6		
1849	123	133.6		
1859	112	124.7		
1869	110	105.6		
1872	108	105.8		
1874	113	112.5		
1878	112	110.3		
1884*	94	106.6	200.5	194.2
1891*	99	113.0	147.3	147.3
1905	83	121.2	121.5	112.7
1930	83	95.5		
1950	71	56.4		
1955	82	64.5		
1967	83	75.9		
1970	72	67.2		
1973	84	76.4		

*For Japan, the comparisons are for 1886 and 1892.

Table A.2: Historical US vs. Federal Reserve Samples

Country	Historical Sample Start	Post-war Sample Start
1. China	1820	1950
2. Netherlands	1820	1950
3. Norway	1820	1950
4. Spain	1820	1950
5. Sweden	1820	1950
6. UK	1820	1950
7. Belgium	1830	1950
8. Brazil	1830	1950
9. France	1840	1950
10. Germany	1852	1950
11. Australia	1861	1950
12. Italy	1861	1950
13. Denmark	1864	
14. Japan	1868	1950
15. India	1873	1950
16. Chile	1880	1950
17. Austria	1884	1950
18. Switzerland	1885	1950
19. Uruguay	1885	
20. Egypt	1895	
21. South Africa	1895	
22. Peru	1900	
23. Venezuela	1900	1950
24. Philippines	1902	1950
25. Canada	1910	1950
26. New Zealand	1914	
27. Colombia	1919	1950
28. Finland	1920	1950
29. Greece	1922	2001
30. Turkey	1922	
31. Hong Kong		1950
32. South Korea		1950
33. Singapore		1950
34. Russia		1993
35. Taiwan		1950
36. Malaysia		1950
37. Mexico		1950
38. Euro Area		1999
39. Israel		1950
40. Indonesia		1950
41. Saudi Arabia		1950

Table A.3
Data Sources

Data	Source	URL
Bilateral Trade Data (total flows)	IMF Dots	http://elibrary-data.imf.org/
Bilateral Trade in Oil	UN Comtrade	http://comtrade.un.org/
Federal Reserve Trade Weights	Federal Reserve Board of Governors	http://www.federalreserve.gov/releases/h10/weights/
Fed's Broad Trade-Weighted Real Exchange Rate, Korean Inflation Index	Federal Reserve Economic Data	http://research.stlouisfed.org/fred2/
PPP, Real GDP, Exchange Rates	Penn World Tables, v7.1, v8.0, v8.1	https://pwt.sas.upenn.edu/
Historical Exchange Rates	Bordo	https://sites.google.com/site/michaelbordo/home4
Historical Inflation, Trade Data	Historical Statistics of the United States	http://hsus.cambridge.org/
Historical Trade Data	Barbieri (2002)	http://people.cas.sc.edu/barbierk/databases.html
Historical Trade and Tariff Data	Jacks, Meissner, Novy (2009)	N/A
Average MFN Applied Tariff Rates in Developing and Industrial Countries	World Bank	http://econ.worldbank.org
Historical Tariff Data	Irwin (2007)	http://www.nber.org/papers/w13450
Post-war Taiwan Inflation	Taiwanese Government	http://eng.stat.gov.tw/mp.asp?mp=5
Post-War Tariff Data	UN Comtrade	http://comtrade.un.org/
Manufacturing Data (Employment, etc.)	BEA, Annual Survey of Manufactures	www.census.gov/manufacturing/asm
Historical Exchange Rates	Historical Volumes of Mitchell	N/A
Historical Exchange Rates	<i>Wahrungen der Welt</i> , Vol. 1-10	N/A
Historical Price Indices (Table A.1 data)	Devereux and Ward (2006), via Lindert & Arroyo Abad, Global Price and Income History Group	gpih.ucdavis.edu
British Prices From 1750	Office for National Statistics (ONS)	http://www.ons.gov.uk
Chinese Historical Prices	Allen et al. (2011)	via gpih.ucdavis.edu
IMF Trade Weights	Bayoumi <i>et al.</i> (2006)	http://www.imf.org/external/pubs/ft/wp/wp9771.pdf
Tariff Data, 1984-1995	UNCTAD -- TRAINS	www.unctad.info/en
Trade Data by SITC4, 1962-2008	Feenstra, Lipsey et al. 2005	cid.econ.ucdavis.edu
Manufacturing Labor Costs, Productivity	UNIDO's INDSTAT	http://data.un.org/
Manufacturing Labor Costs, Productivity	BLS	http://www.bls.gov/fls/
Relative Price Levels in Manufacturing	World Bank, International Comparison Project 2005	http://databank.worldbank.org/data/databases.aspx
Relative Price Levels in Manufacturing	Groningen Growth and Development Centre Productivity Level Database	http://www.rug.nl/research/ggdc/data/
Manufacturing Labor Costs, Productivity, Trade	OECD	http://stats.oecd.org/
Canada, Manufacturing Deflator	Statistics Canada	http://www.statcan.gc.ca

10 Appendix: Not for Publication

10.1 Theoretical Appendix: Armington Model

Assume a utility function for country j of the form:

$$U_j = \sum_i N_i \left(c_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \quad (10.1)$$

where c_{ij} is a variety of good produced in country i and consumed in country j , and N_i is the number of identical varieties produced in country i (this setup is identical to the model presented in Chapter 5 of Feenstra's textbook), and where $\sigma > 1$. This is subject to a budget constraint: $Y_j = \sum_i p_{ij} c_{ij}$, with iceberg transport costs τ_{ij} , such that $p_{ij} = p_i \tau_{ij}$. The solution for revenue of total exports from country k to country j will thus be:

$$x_{kj} = \frac{N_k (\tau_{kj} p_k)^{1-\sigma}}{\sum_i N_i (p_i \tau_{ij})^{1-\sigma}} Y_j \quad (10.2)$$

This equation says that exports from k to j are determined by income in country j , and by the relative price level of goods exported from k to j vs. the prices of goods exported from other countries to country j , weighted by the number of varieties in each country. Notice that this relative price can certainly change even if none of the p_i change, if, for example, a country increases the number of varieties it offers or opens up to free trade. This is precisely the problem with an index-of-indices setup, which assumes that only changes in prices can change the relative price index. In the real world, the liberalization of trade and the fast growth of India and China have led to a decline in the first term in equation 10.2.

Differentiating x_{jj} , the shipments of country j to itself, with respect to τ_{kj} , I have:

$$\frac{\partial x_{jj}}{\partial \tau_{kj}} = \frac{N_j (\tau_{jj} p_j)^{1-\sigma} Y_j}{(\sum_i (N_i p_i \tau_{ij})^{1-\sigma})^2} N_k p_k^{1-\sigma} (1-\sigma) \tau_{kj}^{-\sigma} > 0. \quad (10.3)$$

Thus, the impact of a decline in trade costs or an increase in the number of varieties will reduce country j 's shipments to itself, in a way that depends on the price of goods in country k relative to the price of goods in other countries. Note also that this expression is almost identical to the impact of a change in prices (only, flipping the exponents for p_k and τ_{kj}). The impact of a 10% decline in unilateral trade costs or a 10% change in relative prices have very similar theoretical implications. For any trade model in which relative prices matter, the Weighted-Average Relative (WAR) series will be the proper price index. However, the theoretical rationale for using WAR series is more easily

intuited from the numerical example in Section 3.1, and thus this section is relegated to the appendix.

10.2 Theoretical Similarity of WARP, P-WARP and WARULC

As mentioned, the P-WARP and WARP indices for the US are similar directionally, but have quite different implications for the level of the valuation of US real exchange rates. Economists have generally preferred real exchange rate measures which include productivity adjustments, and on theoretical grounds, given the Balassa-Samuelson Effect, I argue that the Balassa-Samuelson adjusted series might potentially have more explanatory power in some situations, particularly for fast-growing economies. For example, Japan's price level relative to the US was the same in 1973 and 1986, but in 1986 Japanese GDP per capita (PPP) had grown from 67% of the US total to close to 80%. Thus on a Balassa-Samuelson adjusted basis, relative to productive output, Japan's currency actually appears more competitive vs. the US dollar in 1986.

Although economists have generally preferred RER indices computed from unit labor costs in manufacturing, these series do have some obvious drawbacks, as mentioned above. These include a myriad of data comparability issues, the fact that wages give an incomplete picture of competitiveness (there are also non-traded services inputs), and, in addition, the fast-shrinking manufacturing sector might not yield a complete measure of the competitiveness of all tradable sectors. A potential solution to these problems would be to use the labor share of GDP per capita as a proxy for manufacturing wages (the two are highly correlated), evaluated at exchange rates, and to use manufacturing output per capita for productivity (which will not rise when employment falls), evaluated at PPP. The problem then becomes that some countries are endowed with oil, large quantities of natural resources, or larger tradable services sectors. Since we are also actually more interested in all tradable industries rather than just manufacturing specifically, an obvious solution would be to use the labor share of GDP per capita evaluated at PPP as a proxy for tradable sector output. A further problem is that the labor share of income is often thought to be heavily influenced by tax changes and reporting standards and accuracy. For example, Bai and Chen (2010) argue that the official estimates of the labor share in China are wrong, and suggest a vastly different alternative series.

A solution to this problem would be to use GDP per capita in the place of the labor share of GDP per capita. Then I have:

$$\frac{GDP_{pc}/e}{GDP_{pc}/PPP} = \frac{PPP}{e} = RER_{dollar/fx} \quad (10.4)$$

In other words, this index would be equivalent to WARP (derived previously as the inverse of equation (10.4), in local currency per dollar). However, the problem with this method is that it uses PPP for all industries when PPP for tradable sectors is preferred. Tradable sector prices are higher in poor countries relative to the price of nontraded services, which means that this measure will bias up the estimates of the size of developing countries tradables sectors, and bias down the estimates of their ULCs. RULCs for rich countries will then be biased up for rich countries. One solution might be to use tradable sector PPPs from the ICP, but one additional problem is that this will be an incomplete measure of competitiveness, since many non-traded services are also important inputs into production. Thus we would ideally want to use something in between overall PPP and PPP in tradable sectors. Such a measure would be correlated with GDP per capita according to the Balassa-Samuelson effect. Thus we would arrive at something approximating P-WARP as the ideal measure of unit input costs for all tradable sectors.

10.3 Adjusting for Domestic Competition

In trying to take the indices computed in this paper to the data, or in order to make comparisons over different countries or over different epochs, another problem emerges. Any given producer of a tradable good would theoretically be more exposed to exchange rate movements if located in a small open economy versus a large economy with less trade exposure. A US manufacturer in 1950 mostly competed with other manufacturers located in the US, was likely to export little, and on average would not have been much affected by a 10% appreciation of the dollar. Thus, an alternative index operating from the perspective of an individual firm should include a weight for home-country competition, for which the real exchange rate is always one. Firms located in countries with less trade as a share of output would thereby systematically experience less variation in their real exchange rate indices. Thus since Italy trades much more than the US as a share of GDP, in large part because Italy is much smaller, we would expect its real exchange rate to rise by less due to increased trade with China, since even if China traded as much with Italy as with the US as a share of GDP, its share in Italian trade would be much smaller.

Ideal trade-weights would require information on domestic shipments and domestic consumption for all trading partners, which is not readily available. A simple approximation would be to assume that all exports compete primarily with foreign goods, and that all domestic shipments compete with foreign competition based on the share of

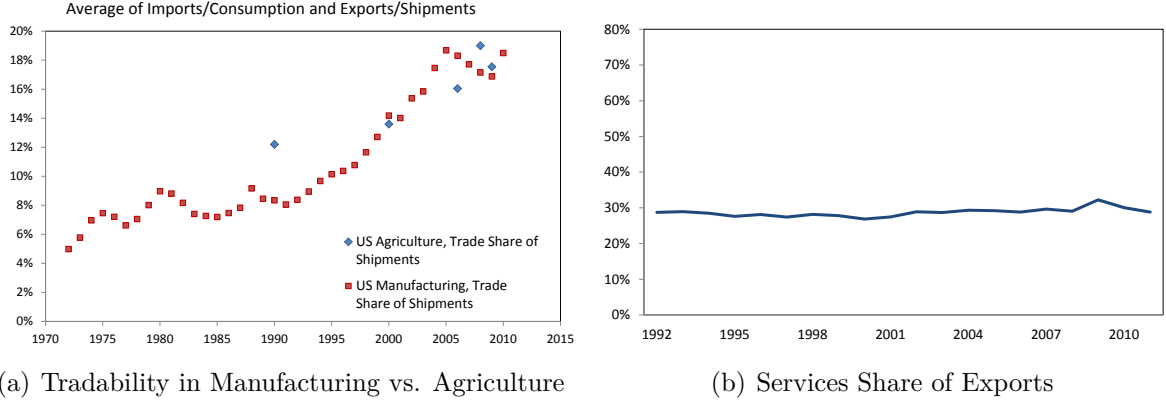


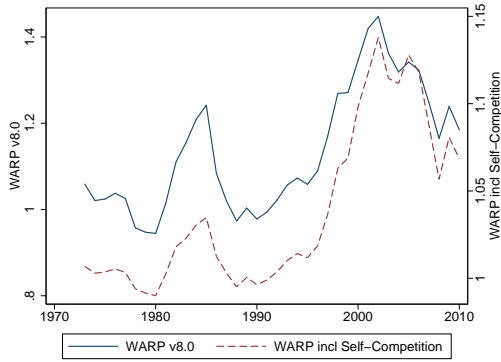
Figure 20: Tradability: Manufacturing, Agriculture, and Services

imports in home consumption. The export share of shipments and the import share of home consumption in all tradables sectors can be proxied using manufacturing sector data. Manufacturing accounts for a large share of US trade, having been 63% of total US trade in 1992, declining only slowly to about 60% by 2006, although it did decline sharply after 2007. Another large component of traded goods is agriculture, which Figure 20(a) shows has tended to have a roughly similar share of traded goods in production and consumption as manufacturing over the past several decades. This implies that using data on manufacturing, which still accounts for a majority of US trade, as a proxy for traded shares of tradable goods overall is reasonable. A third major category of trade is services, which economists frequently stress has witnessed impressive growth in trade. However, contrary to many economists' priors, the BEA data presented in Figure 20(b) show that services exports as a share of total US exports have remained relatively constant over the past few decades.

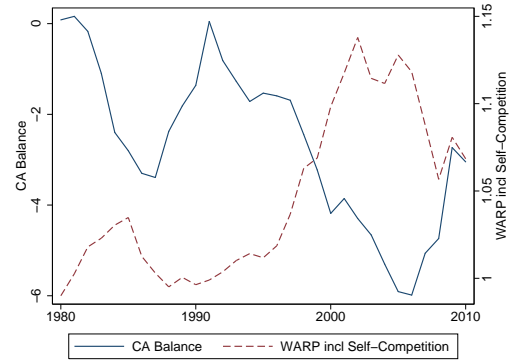
Thus, in terms of the indices computed ignoring the role of domestic market competition, an index adjusted for domestic competition could be written as:

$$I_{j,t}^{new} = I_{j,t}^{old} \cdot s_t + 1 \cdot (1 - s_t). \quad (10.5)$$

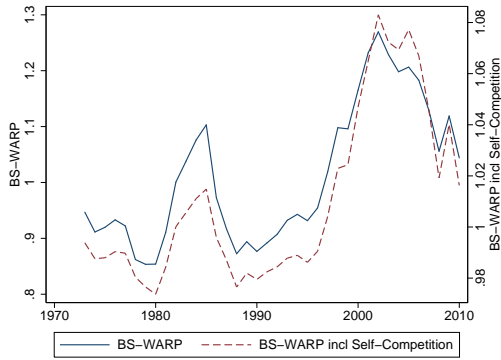
Therefore the modified index would simply be a weighted average of the real exchange rate index with trading partners time the trade share s_t plus the real exchange rate



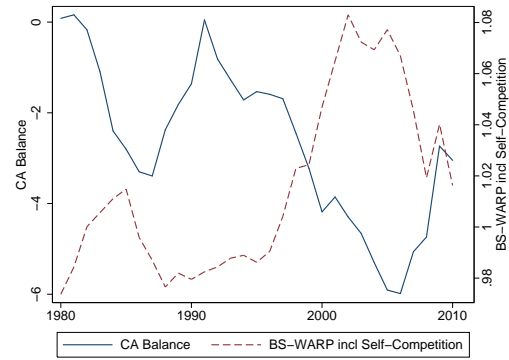
(a) vs. WARP



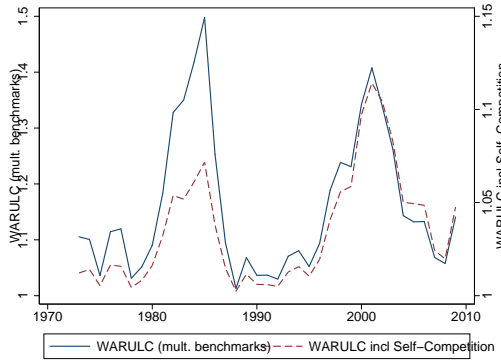
(b) vs. the US CA Balance



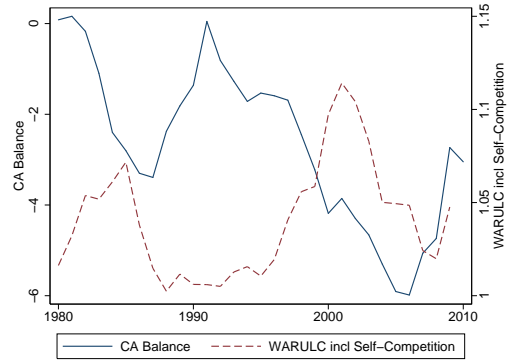
(c) vs. P-WARP



(d) vs. the US CA Balance



(e) vs. WARULC



(f) vs. the US CA Balance

Figure 21: Comparing Indices with Weights for Self-Competition

with domestic competitors (one) times the share of domestic competition out of total competitors. The trade share is an average of imports (m_t) over domestic consumption (c_t) and exports (x_t) over shipments (s_t), which will be equal when trade is balanced. Domestic consumption is defined as shipments minus exports plus imports.

$$s_t = \frac{1}{2} \left(\frac{m_t}{c_t} + \frac{x_t}{sh_t} \right) \quad (10.6)$$

The results in Figure 21 are striking. The WARP and P-WARP indices appear to be close to mirror images of the CA balance. In each case, it is now much more apparent that the late 1990s and early 2000s exchange rate appreciation was a much larger shock to firms in tradable sectors than was the dollar appreciation episode in the 1980s. In both the WARP and the P-WARP indices, the shock in the early 2000s looks to have been about twice as large, which seems to match the behavior of the CA deficit. In addition, for all indices, the dollar depreciation after 2002 is now considerably slower. The WARULC index looks good up until the early 2000s. Then, it shows a large depreciation just before the CA deficit begins to worsen (its peak is in 2001, vs. 2002 for WARP and P-WARP). The need to adjust for cyclical factors is also apparent in these graphs, as the US economy was in a recession in 2001, which would have improved the CA balance, while the cyclical peak in 2005 would have substantially worsened the CA deficit.

10.4 Time-Varying Balassa-Samuelson Coefficient

The Balassa-Samuelson effect was generally small and insignificant during the entire Bretton Woods period when many developing countries had high tariff barriers relative to developed countries, and when there were no benchmark years of the ICP. Thereafter, the effect grows until reaching a peak in the mid-1990s after which time it fluctuates. Given that the Balassa-Samuelson coefficient on the full sample of PWT countries evolves over time, a second approach would be to run the Balassa-Samuelson regression by year letting the coefficient vary (implemented by Johnson, Ostry, and Subramanian, 2010):

$$\ln RER_{j,t} = \alpha_t + \beta_t * \ln RGDP_{j,t} + \epsilon_{j,t}, \forall t = 1950, \dots, 2010. \quad (10.7)$$

The index using the residuals from this regression in equation (4.4) is plotted in Figure 22 versus the benchmark of the Fed's Broad Trade-Weighted real exchange rate from Figure 2. Since 1970, these two indices for the US happen to line up rather closely, although they diverge before that.

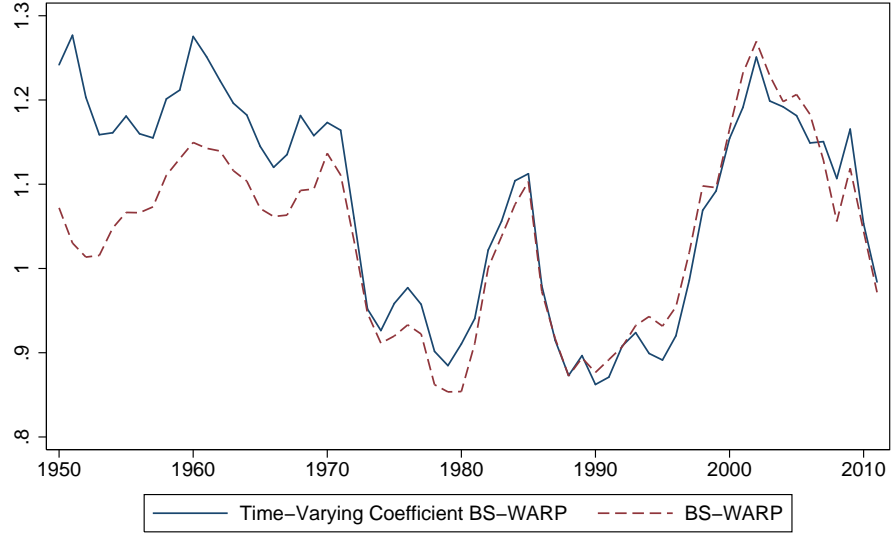


Figure 22: P-WARP: Time Varying vs. Constant Coefficient

As mentioned in the main text, a compelling reason to discount this method *a priori* is that Feenstra, Inklaar, and Timmer *et al.* (2015) convincingly show that the apparent growing Balassa-Samuelson effect discovered by Bergin, Glick, and Taylor (2006) using version 6 of the PWT is a statistical artifact of extrapolated data. Data from benchmark years of the International Comparison Project always show a similarly-sized Balassa-Samuelson effect, although there is year-to-year variation.

10.5 Weighted Average Relative Wages

One alternative to ULCs which does not rely on Value-Added or PPP estimates in manufacturing, are simply to plot a weighted-average of relative hourly wages (WARW) in manufacturing (Fig. 23). Interestingly, the relative wage differentials were just as large under Bretton Woods as they are today, although the 23 country sample becomes slightly unbalanced prior to 1980.

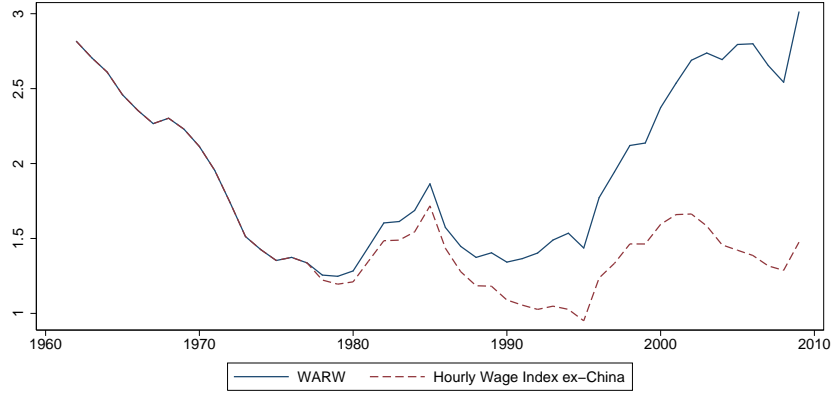
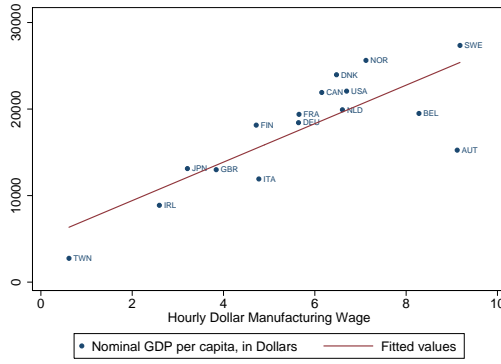
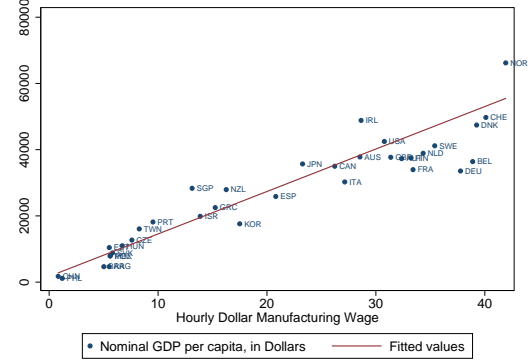


Figure 23: Trade-Weighted Relative Wages in Manufacturing

One drawback is that most of this data comes from the OECD, and so China and Mexico are the only two large developing countries in the sample. Since there appears to be a very strong correlation between nominal GDP per capita converted into dollars at exchange rates, and the hourly manufacturing wage in dollars (Figure 24), one obvious way to expand the sample of countries is by using GDP per capita as a proxy for hourly dollar wages in tradable sectors. The weighted average of relative GDP per capita (WARG) and WARW, plotted in Figure 25, track each other closely, with the difference being that the series using GDP per capita shows a generally higher gap between the US and trading partners. This can be expected since it includes a much larger sample of developing countries.



(a) 1975



(b) 2005

Figure 24: GDP per capita vs. Hourly Manufacturing Wage (at exchange rates)

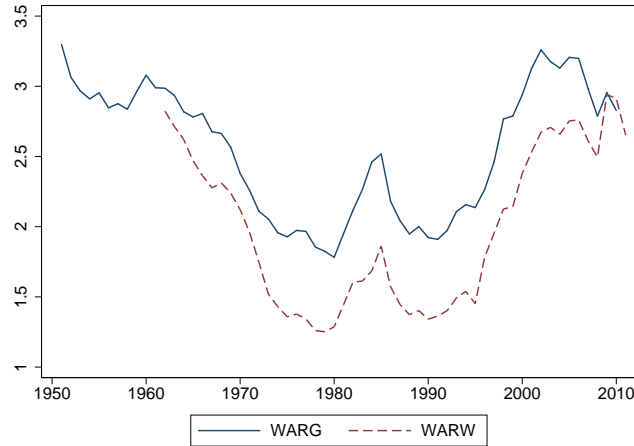


Figure 25: Weighted-Average Relative GDP per capita vs. Hourly Manufacturing Wages

11 Additional Appendix: ULC's in China vs. Mexico

In June of 2013, the Boston Consulting Group released a report on relative unit labor costs in manufacturing between China and Mexico, finding that unit labor costs in China exceeded unit labor costs in Mexico beginning in 2011.³⁹ However, as of 2009, I find that ULCs in Mexico were still about 50% larger than the ULCs in China (Figure 26). While I do not have access to complete data on manufacturing employment and hours worked through 2011, hourly dollar wages in China were just \$2.52 vs. \$6.48 in Mexico in 2011, and total Chinese manufacturing production rose 27% vs. just 17% for Mexico from

³⁹BCG press release, 6/28/2013, available here: <http://www.bcg.com/media/PressReleaseDetails.aspx?id=tcm:12-139022>.

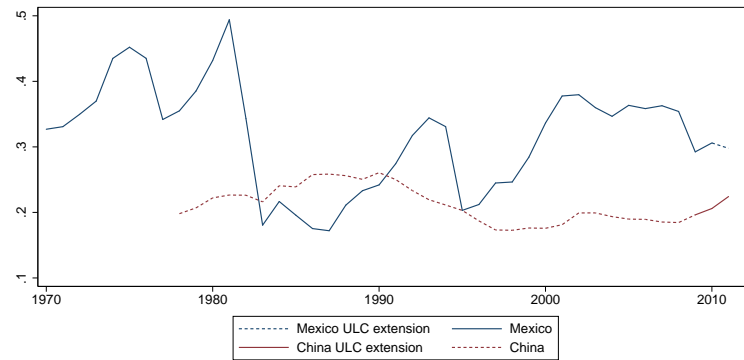


Figure 26: ULCs, Mexico vs. China

2009-2011. If there was no change in relative hours worked, admittedly a very strong assumption, then ULCs did converge a bit between 2009 and 2011, but Mexican ULCs were still roughly 33% higher than Chinese ULCs in 2011. Thus Mexican manufacturing workers, in the aggregate, are still substantially more productive than their Chinese counterparts, although also better paid relative to productivity.

11.1 Additional Appendix: Testing

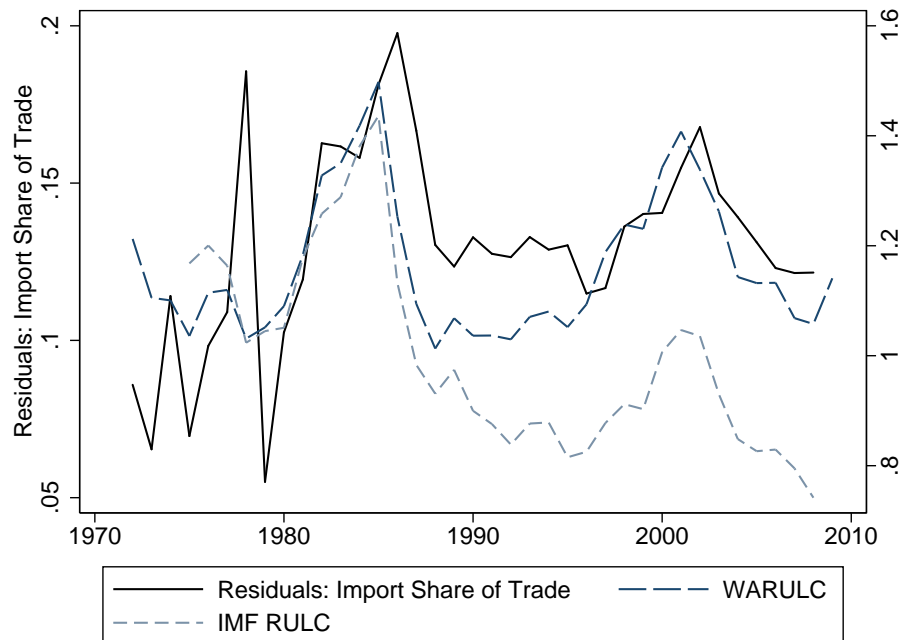


Figure 27: RULC Indices vs. Import Share of Manufacturing Trade

Table IV: Predicting US Log Exports/Imports, 1976-2008

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
L.ln(Exports/Imports)	0.93*** (0.066)	0.80*** (0.054)	0.72*** (0.100)	0.52*** (0.056)	0.53*** (0.049)	0.93*** (0.054)	0.80*** (0.060)
ln Δ GDP	-2.06*** (0.40)	-1.29*** (0.40)	-1.42* (0.69)	-0.99*** (0.33)	-0.84** (0.32)	-1.66*** (0.42)	-1.23*** (0.41)
Foreign Import Dem.	0.51 (0.33)	0.33 (0.31)	0.28 (0.37)	-0.20 (0.20)	-0.28 (0.21)	0.56 (0.37)	0.12 (0.30)
L.Terms of Trade	0.45 (0.33)						
L.Fed's RER Index		-0.44*** (0.077)					
L.Balassa-Sam. Resid.			-0.32** (0.12)				
L.WARP				-0.52*** (0.060)			
L.P-WARP					-0.73*** (0.081)		
L.IMF RULC						-0.13*** (0.036)	
L.WARULC							-0.33*** (0.063)
Constant	-0.45 (0.30)	0.36*** (0.077)	-0.074* (0.038)	0.47*** (0.061)	0.58*** (0.072)	0.094* (0.047)	0.33*** (0.078)
Observations	33	33	33	33	33	33	33
R^2	0.897	0.927	0.909	0.967	0.972	0.901	0.932

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in parenthesis. The dependent variable is the log exports/imports. Foreign import demand is the trade-weighted average of predictions of import demand growth due to changes in GDP.

Table V: Predicting Imports/Exports, 1976-2008, 154 Countries, Random Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)
L1.ln(Exports/Imports)	0.81*** (0.0092)	0.80*** (0.0093)	0.80*** (0.0095)	0.80*** (0.0094)	0.82*** (0.0091)	0.82*** (0.0091)
ln Δ GDP	-0.10 (0.083)	-0.11 (0.083)	-0.12 (0.082)	-0.12 (0.083)	-0.12 (0.084)	-0.12 (0.084)
Foreign Import Demand	0.32** (0.14)	0.30** (0.14)	0.33** (0.14)	0.32** (0.14)	0.31** (0.15)	0.31** (0.15)
L.ToT	-0.17*** (0.043)					
Balassa-Sam. Resid.		-0.033* (0.017)				
L.WARP v8.1			0.026 (0.027)			
L.P-WARP v8.1				-0.018 (0.019)		
L1.WARP v7.1					0.0099 (0.0095)	
L.P-WARP v7.1						0.0013 (0.0074)
Observations	4318	4318	4318	4318	4318	4318
Log Likelihood	-341.5	-347.2	-344.9	-343.7	-360.8	-361.6

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log of exports/imports. Foreign import demand is the trade-weighted average of predictions of import demand growth due to changes in GDP. The random part of the coefficients is suppressed for space.

Table VI: Predicting Imports/Exports, 1976-2008, 95 Countries, Random Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)
L1.ln(Exports/Imports)	0.76*** (0.014)	0.78*** (0.013)	0.77*** (0.014)	0.78*** (0.014)	0.78*** (0.014)	0.81*** (0.013)
ln Δ GDP	-0.045 (0.13)	-0.080 (0.13)	-0.037 (0.12)	-0.069 (0.13)	-0.062 (0.12)	-0.076 (0.13)
Foreign Import Demand	0.19 (0.17)	0.17 (0.18)	0.12 (0.18)	0.15 (0.18)	0.13 (0.17)	0.16 (0.18)
L1.ToT	-0.28*** (0.086)					
L1.IMF RER Index (CPI)		0.013 (0.011)				
Balassa-Sam. Resid.			-0.015 (0.026)			
L1.WARP v8.1				0.021 (0.033)		
L1.P-WARP v8.1					-0.0017 (0.028)	
L1.WARP v7.1						0.019* (0.011)
Observations	2195	2195	2195	2195	2195	2195
Log Likelihood	154.3	132.3	137.9	142.6	139.4	128.1

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the import share of manufacturing trade.

Note that this measure includes exports in the denominator, and so it functions as a normalized version of the trade balance. Foreign GDP is the average of the G7 economies.

Table VII: Predicting Canadian Exports/Imports, 1973-2011

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
L.ln(Exports/Imports)	0.75*** (0.11)	0.85*** (0.13)	0.61*** (0.13)	0.68*** (0.14)	0.72*** (0.14)	0.60*** (0.14)	0.55*** (0.12)
ln Δ GDP	0.60 (0.37)	0.58 (0.40)	0.36 (0.37)	0.44 (0.41)	0.47 (0.42)	0.28 (0.39)	0.27 (0.35)
Foreign Import Dem.	-0.56 (0.35)	-0.59 (0.37)	-0.41 (0.33)	-0.40 (0.36)	-0.41 (0.37)	-0.48 (0.33)	-0.48 (0.31)
L.ToT	0.35** (0.16)						
L.IMF CPI REER		-0.020 (0.070)					
L.Balassa-Sam. Resid.			-0.23** (0.087)				
L.WARP v8.1				-0.097 (0.075)			
L.P-WARP v8.1					-0.080 (0.084)		
L.IMF RULC Index						-0.21** (0.082)	
L.WARULC							-0.14*** (0.041)
Constant	-0.35** (0.17)	0.025 (0.029)	0.030 (0.026)	0.11 (0.080)	0.090 (0.086)	0.077** (0.036)	0.15*** (0.045)
Observations	36	35	36	36	36	34	36
R^2	0.648	0.616	0.669	0.616	0.606	0.530	0.710

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. The dependent variable is the log of exports/imports. Foreign import demand is the trade-weighted average of predictions of import demand growth due to changes in GDP.

12 Additional Appendix: International Indices

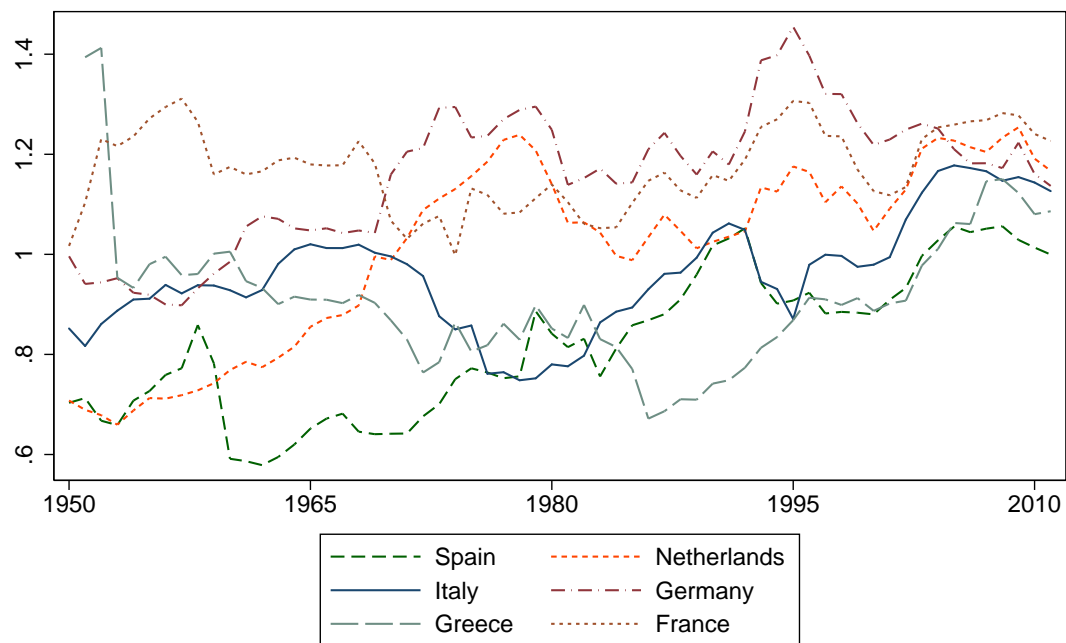


Figure 28: WARP, Major Euro Area Countries

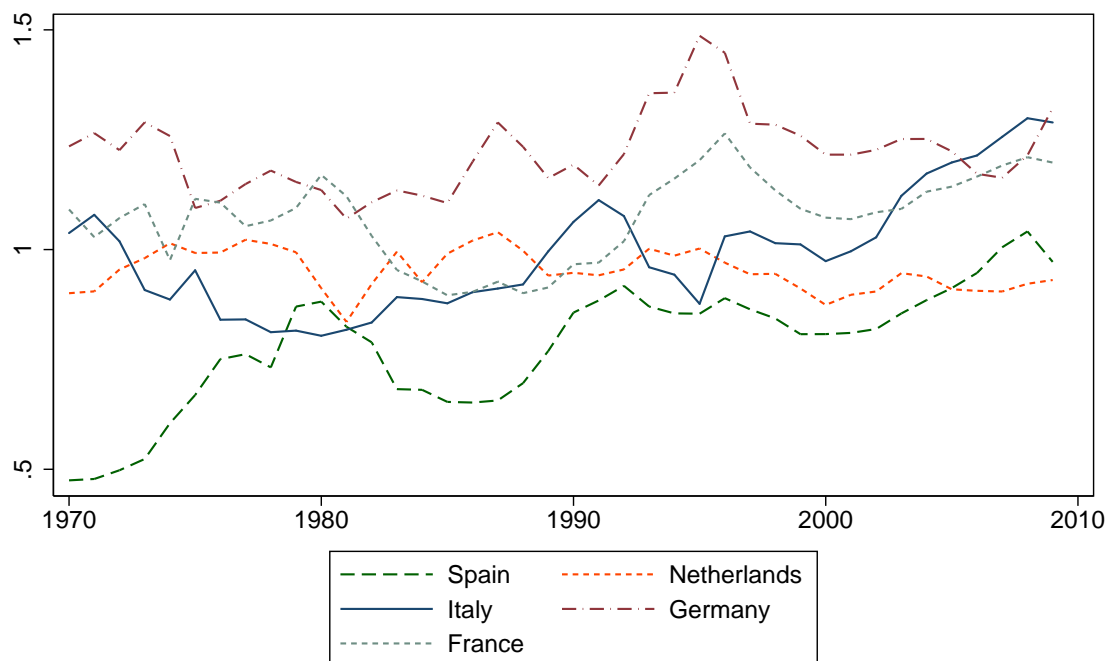
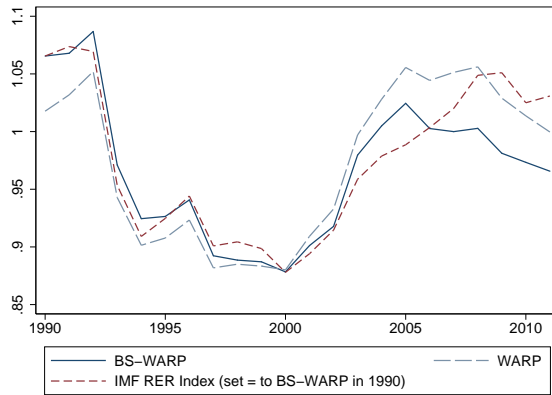
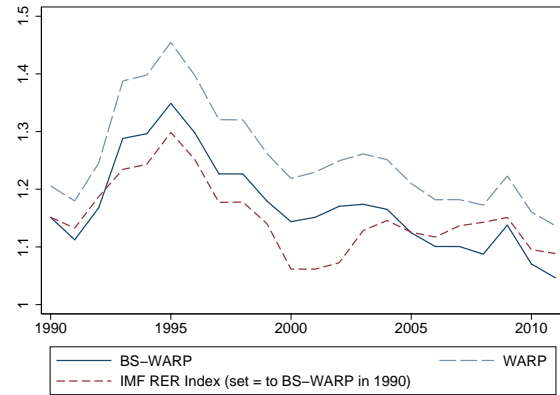


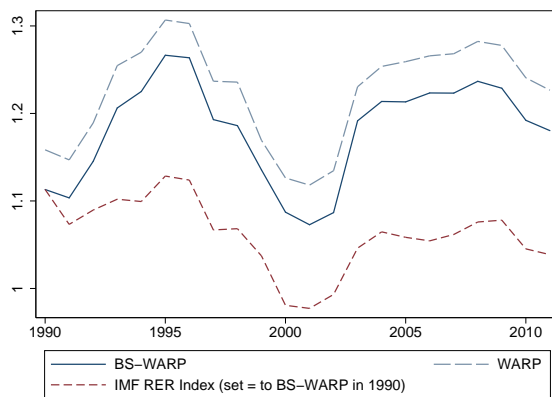
Figure 29: WARULC, Major Euro Area Countries



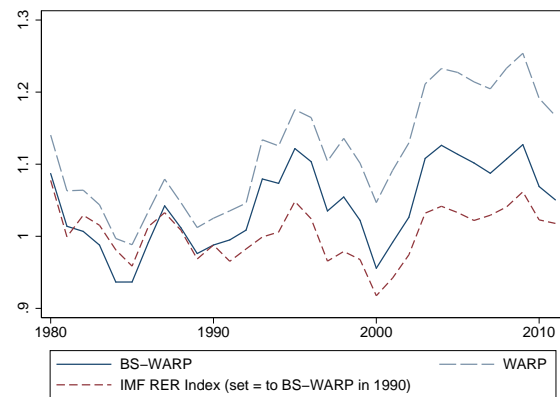
(a) Spain



(b) Germany

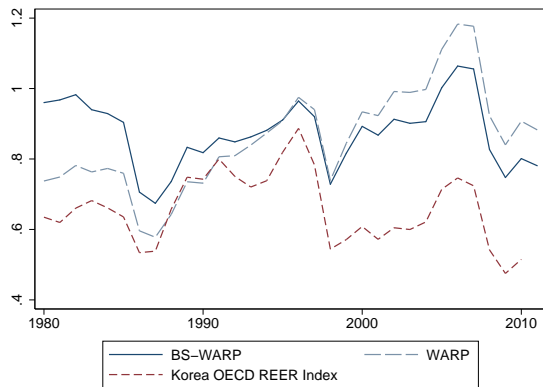


(c) France

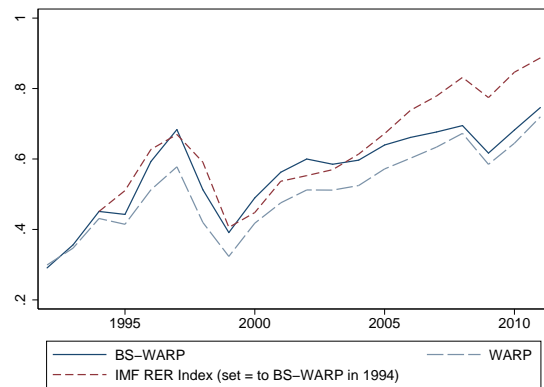


(d) Netherlands

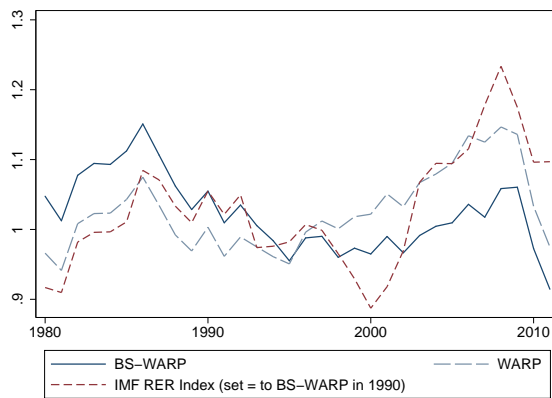
Figure 30: P-WARP and WARP vs. IMF Divisia RER Index



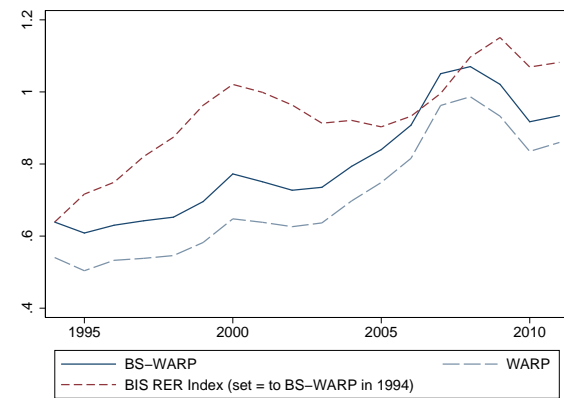
(a) Korea



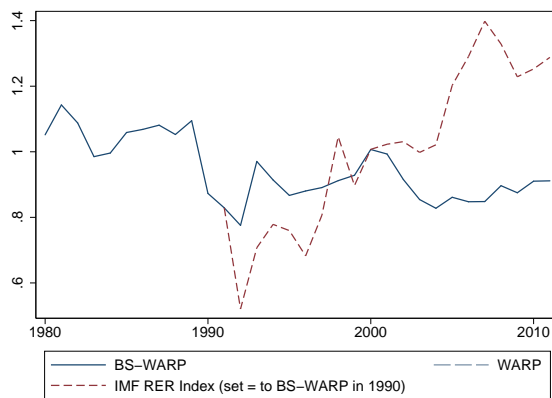
(b) Russian Federation



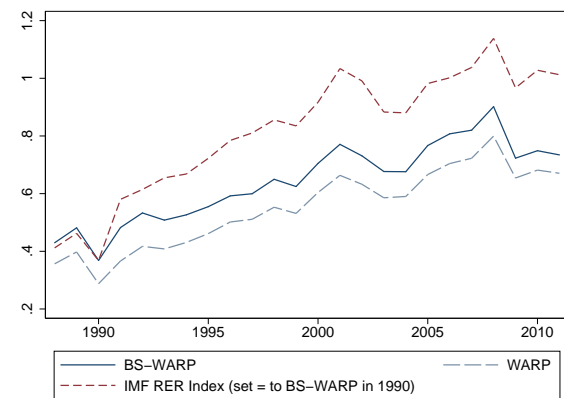
(c) Ireland



(d) Latvia

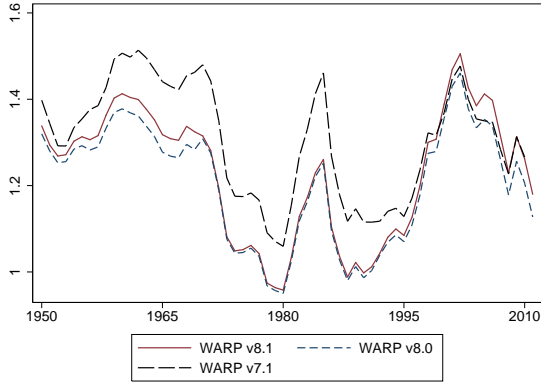


(e) Romania

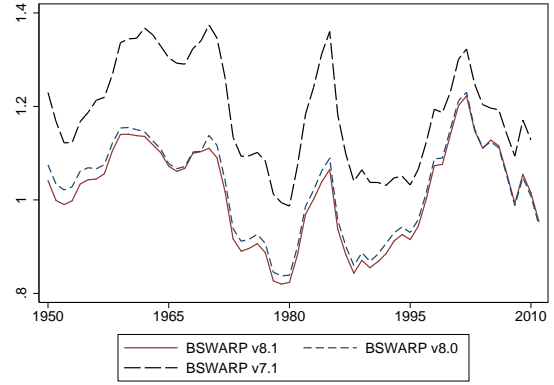


(f) Poland

Figure 31: P-WARP and WARP vs. IMF Divisia RER Index

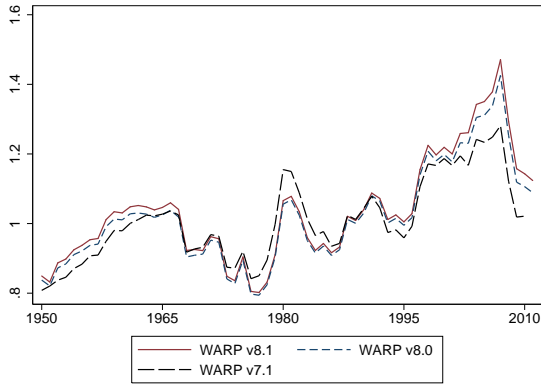


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

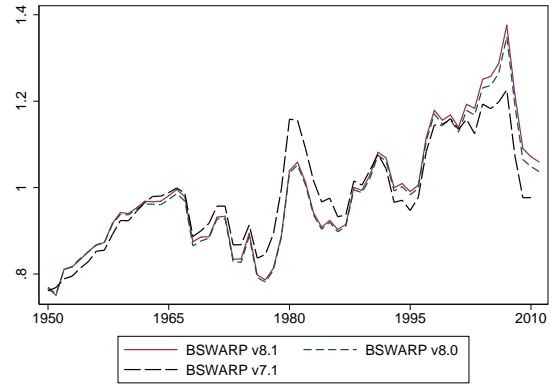


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 32: US

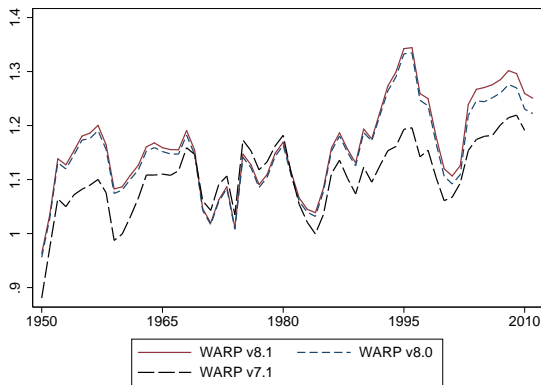


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

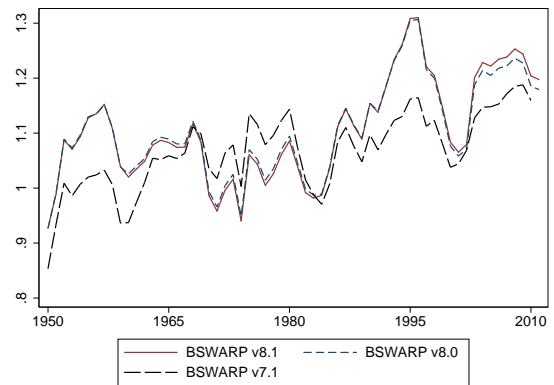


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 33: UK

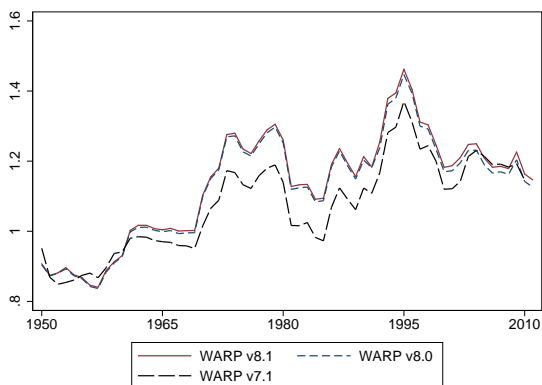


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

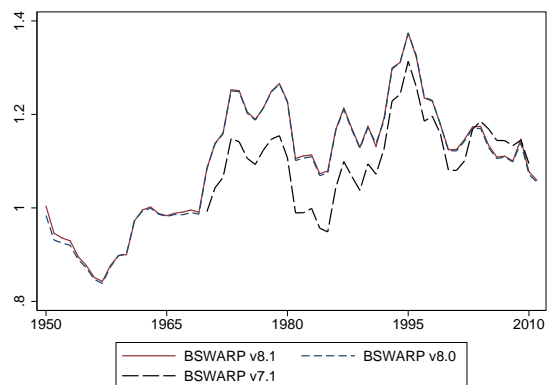


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 34: France

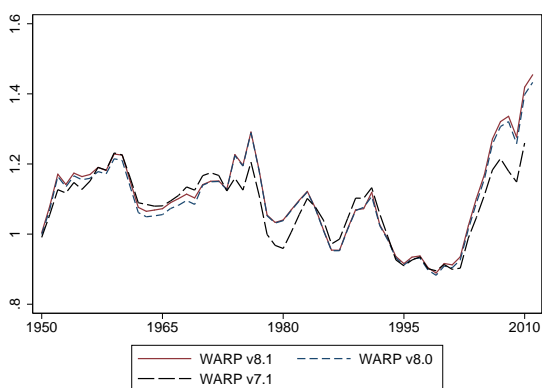


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

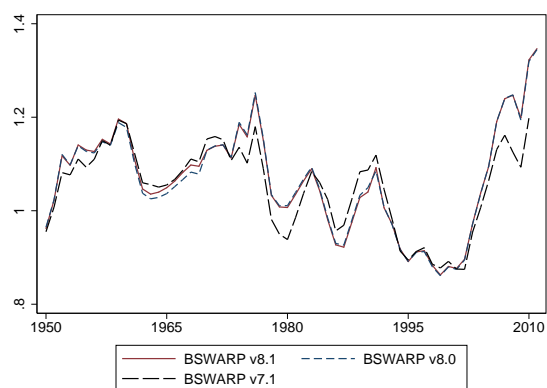


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 35: Germany

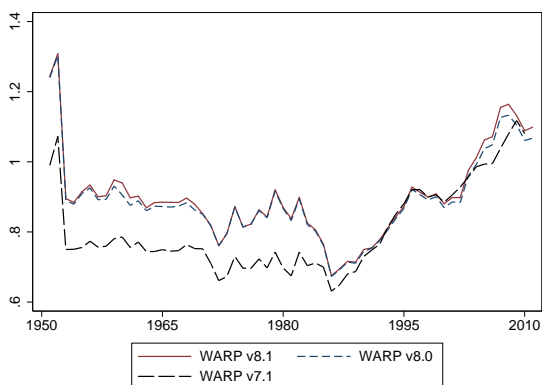


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

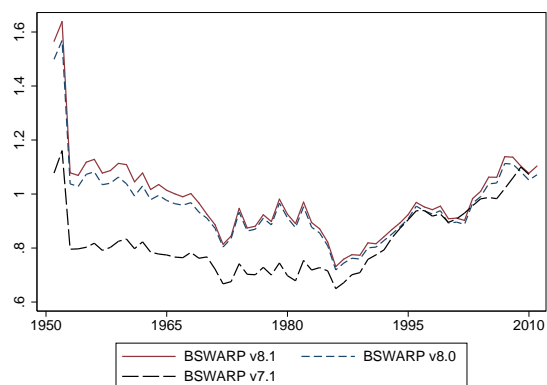


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 36: Canada

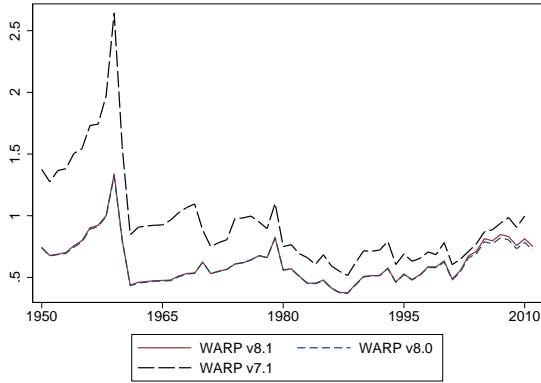


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

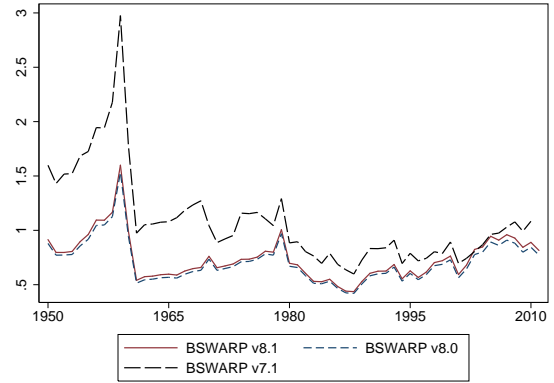


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 37: Greece

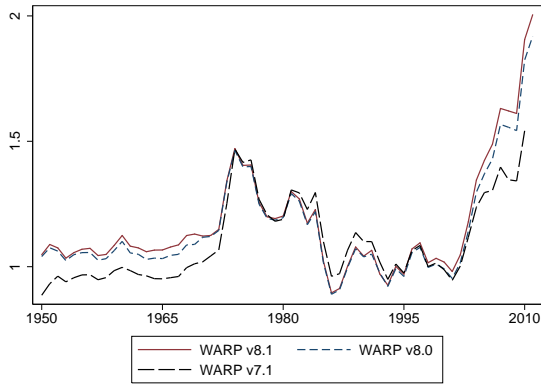


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

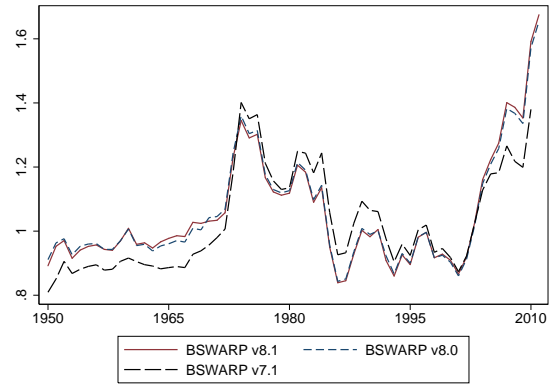


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 38: Turkey

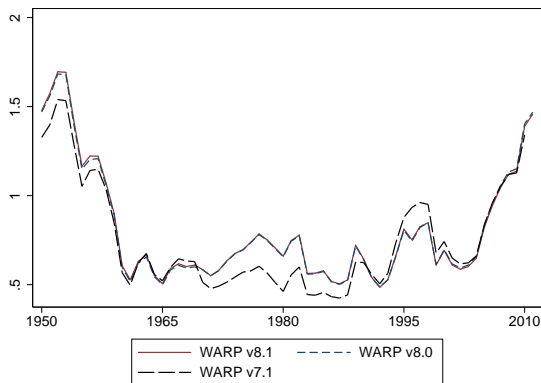


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

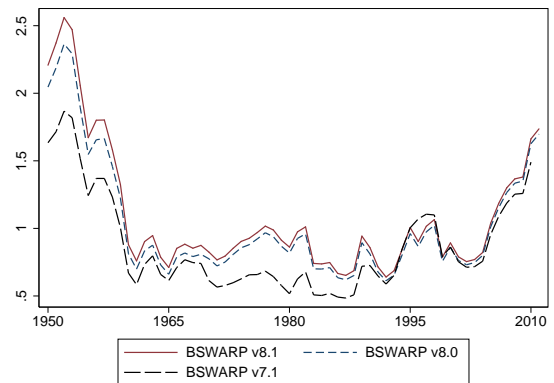


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 39: Australia

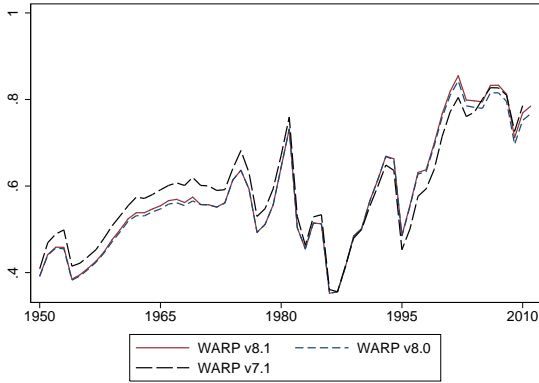


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

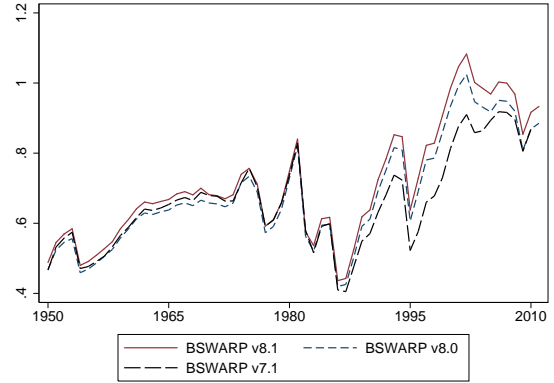


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 40: Brazil

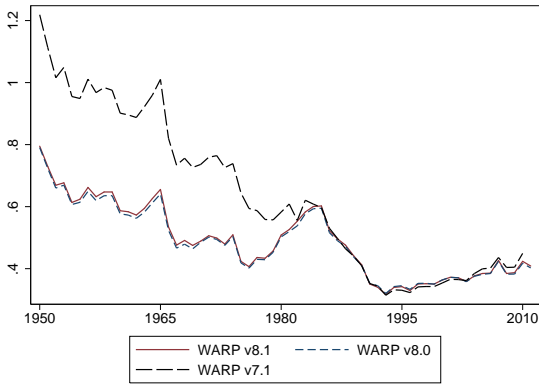


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

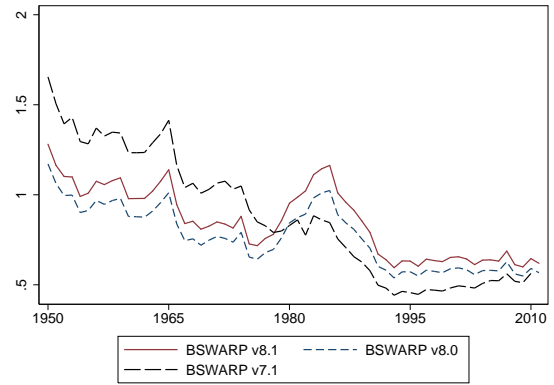


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 41: Mexico

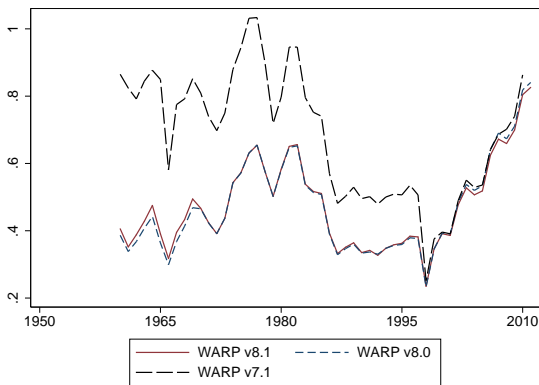


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

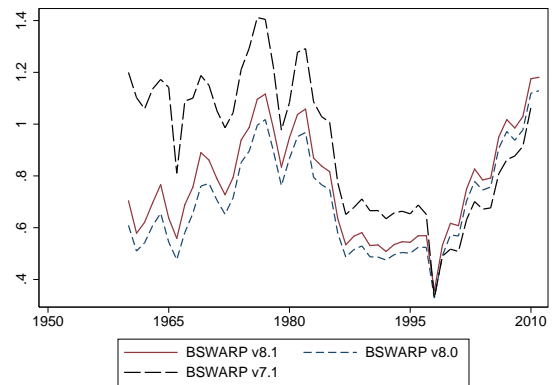


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 42: India

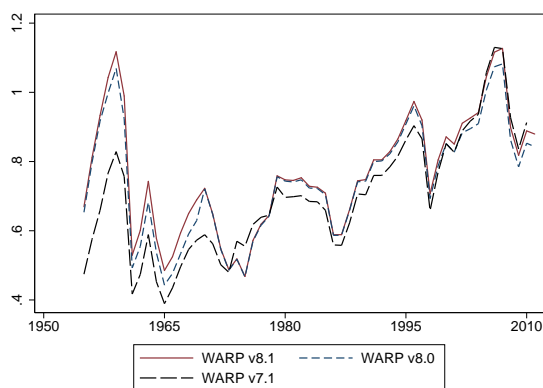


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

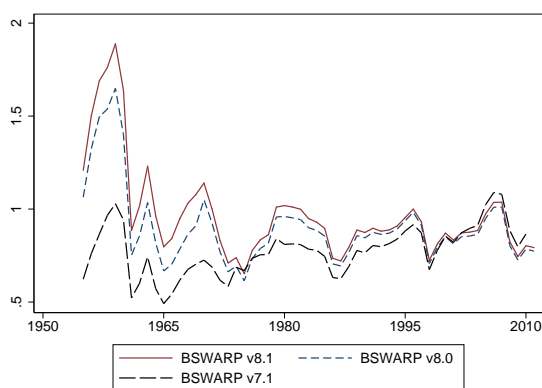


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 43: Indonesia

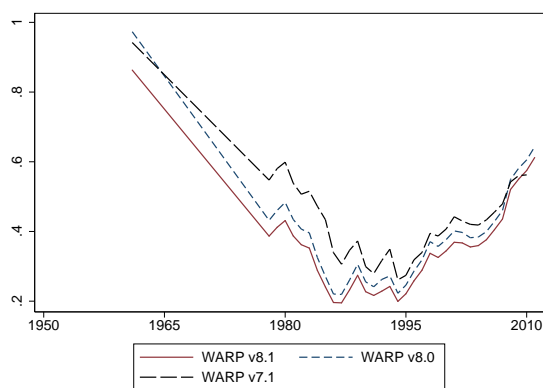


(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1

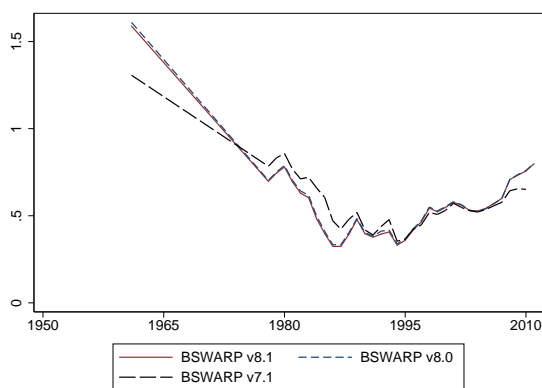


(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 44: South Korea



(a) WARP: PWT v7.1 vs. v8.0 vs. v8.1



(b) BSWARP: PWT v7.1 vs. v8.0 vs. v8.1

Figure 45: China