Balassa–Samuelson in the Long Run: Empirical Success, Quantitative Limits*

Shinnosuke Kikuchi
UCSD

November 16, 2025

Preliminary and do not circulate

Please click HERE for the most recent version.

Abstract

We empirically and quantitatively revisit the Balassa–Samuelson (BS) mechanism in the long run. Traditional specifications are fragile, but adding time fixed effects yields a stable, positive BS elasticity across samples and frequencies—evidence that the data support BS qualitatively. Quantitatively, however, a standard multi-country trade model fed only observed sectoral productivity cannot match country paths and delivers too-small magnitudes; for Japan, it predicts appreciation while the data show a large depreciation since 1995. These quantitative failures persist with costly trade, multi-country settings, input–output linkages, and time-varying trade costs.

*We thank Daron Acemoglu, Arnaud Costinot, Doireann Fitzgerald, and Ippei Fujiwara for their helpful comments.

1 Introduction

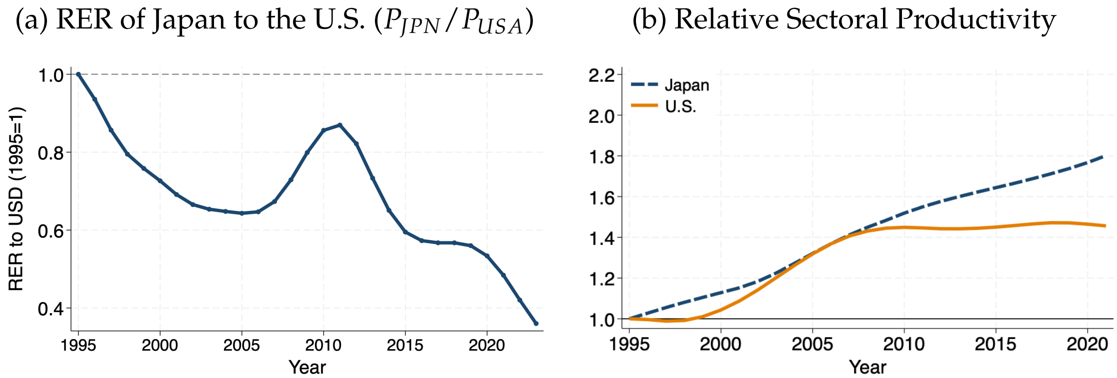
One of the most influential ideas in international economics, developed by Roy Harrod and later formalized by Béla Balassa and Paul Samuelson, is that countries experiencing faster productivity growth in tradables relative to non-tradables should see their price levels appreciate in real terms. Despite its intuitive appeal, the empirical record is mixed.

As an illustration, Figure 1 plots the time series of the real exchange rate (RER) of the Japanese yen (JPY) against the U.S. dollar (USD) and relative sectoral labor productivity (goods relative to services) for Japan and the United States since 1995. Over this period, Japan's RER has depreciated by approximately 60 percent. At the same time, relative to services, labor productivity in Japan's goods sectors has risen by about 80 percent, compared with roughly 50 percent in the United States. According to the Balassa–Samuelson mechanism, such differential productivity growth should have led to an appreciation of JPY, not a depreciation. This divergence already indicates that the standard Balassa–Samuelson framework cannot account for JPY's sustained RER depreciation against USD.

This paper empirically and quantitatively analyzes how the Balassa–Samuelson mechanism explains the evolution of RERs across countries. Empirically, we show that traditional specifications are not robust. This helps rationalize the mixed evidence in the literature across settings. We keep a single reference and add time fixed effects that absorb period-common movements, yielding consistency under standard exogeneity and a $\sqrt{NT}$ (rather than $\sqrt{T}$) rate, which stabilizes the estimates. We show that, on average, the Balassa–Samuelson mechanism is qualitatively active across different countries, sample periods, time frequencies, filtering methods, specifications, and labor-productivity measures. Quantitatively, we develop a standard trade model and evaluate how much sectoral productivity growth can explain RERs. Feeding observed productivity growth does not quantitatively replicate RER paths, and it fails even qualitatively in some countries. This quantitative failure persists after introducing costly trade, multiple countries, input–output linkages, and time-varying trade costs.

Empirical Analysis In the first half of the paper, we sharpen the standard test of the Balassa–Samuelson (BS) mechanism. We first show that traditional single-reference panels *without* time effects perform poorly, even on average. Estimates are fragile to the chosen numeraire: switching the reference country moves the coefficients substantially—only a few benchmarks yield significant positive estimates, while others are imprecise or even

Figure 1: RER and Sectoral Productivity of Japan and the U.S.



Notes: This figure plots Japan's real exchange rate against the United States and the relative sectoral productivity of the two countries. All series are normalized to one in 1995. To isolate low-frequency movements, the Hodrick–Prescott filter with $\lambda = 6.25$ is applied to the logarithmic series. The RER data are constructed from the CPI and nominal exchange rates in the Penn World Table (PWT) data. Sectoral productivity is constructed from real value added (in national currency) and hours worked from EU KLEMS. The agricultural, mining, and manufacturing sectors are classified as goods, and the figure shows productivity in goods relative to services.

wrong-signed. On net, this traditional approach fails to deliver robust qualitative support for BS in panel averages.

We then retain a single reference but add time fixed effects. Time effects absorb period-common movements and shift identification to within-year cross-sectional differences; under standard exogeneity the estimator is consistent and converges at the $\sqrt{NT}$ rate rather than $\sqrt{T}$. With time effects, the BS pattern becomes visible: the relative tradable–nontradable productivity ratio is positively associated with real exchange rates, and the corresponding results are even tighter. These findings are robust across countries, sample periods, time frequencies, filtering choices, specifications, and labor-productivity measures.

Quantitative Analysis In the second half of the paper, we move from reduced form to a general equilibrium environment. We feed the same sectoral series $\{A_{i,s,t}\}$—and, when relevant, bilateral iceberg costs $\{\tau_{i,j,s,t}\}$—into three settings: (i) a 2×2 free-trade case to isolate the textbook BS channel, (ii) a 2×2 costly-trade case to quantify attenuation from frictions, and (iii) a full $N \times 2$ costly-trade case that allows geography and partner reallocation. In all experiments, we solve the static equilibrium year by year, compute sectoral price indices and CPI, and compare model-implied REER to the data.

Feeding only sectoral productivity shocks—the classical BS channel—performs poorly. Across all three environments, model-implied REER movements are a fraction of those in the data, and for several countries even the sign is wrong. The quantitative shortfall persists when moving from the 2×2 to the full $N \times 2$ setting.

Allowing time-varying trade costs to move along their inferred low-frequency paths does not change the conclusion. Adding $\{\tau_{i,j,s,t}\}$ on top of $\{A_{i,s,t}\}$ leaves the slope of fitted versus actual REER close to zero in long differences, with only marginal improvements in a few cases. Incorporating input–output linkages yields the same result: the BS mechanism remains quantitatively too weak.

At the sectoral level, service-sector prices comove with sectoral productivity in the expected direction and align somewhat better than aggregate CPI, but magnitudes are still far too small. Goods-sector prices fare no better. These patterns remain unchanged when both productivity and trade-cost shocks are combined.

Related Literature This paper contributes to two strands of the literature. First, we contribute to the empirical literature that investigates whether the Balassa–Samuelson mechanism—that countries with faster productivity growth in tradable industries experience real appreciation—holds in the data. The empirical record is mixed across specifications, time periods, and country samples.[1] Some studies find a positive association between relative productivity and real exchange rates (Officer, 1976; Hsieh, 1982; Lee and Tang, 2007; Lothian and Taylor, 2008; Cardi and Restout, 2015), while others do not (Canzoneri et al., 1999; Berka et al., 2018; Berka and Steenkamp, 2018). Notably, Berka et al. (2018); Berka and Steenkamp (2018); Devereux et al. (2025) include labor costs or labor wedges as covariates to restore the theory-consistent sign. Other studies shift focus from aggregate real exchange rates to sectoral prices and show that relative sectoral productivity can empirically explain differences in service (or non-tradable) prices (De Gregorio et al., 1994; Canzoneri et al., 1999; Devereux et al., 2025).[2]

Our contribution is to show that, contrary to the existing literature, the Balassa–Samuelson mechanism is empirically and qualitatively valid without these adjustments and robust across specifications, time periods, and countries in the sample. We demonstrate that previous regressions using a single reference country suffer from an over-weighting of that reference country. Using bilateral, pairwise regressions, we show that the Balassa–Samuelson mechanism is present and that the results are robust.

Second, we contribute to the quantitative literature that explores structural drivers behind the time paths of real exchange rates. Most quantitative studies focus on short- or medium-run dynamics of the RER (Berka et al., 2018; Chahrour et al., 2024; Gornemann

[1]See Froot and Rogoff (1995); Tica and Družić (2006) for reviews.

[2]See Engel (1999), who shows that real exchange rates and non-tradable prices are disconnected.

et al., 2025).[3] A few exceptions include Irwin and Obstfeld (2024) and Devereux et al. (2025). Irwin and Obstfeld (2024) decompose the real exchange rate into several price components and show that the relative sectoral price is important in explaining the depreciation of South Korea.[4] A closely related paper is Devereux et al. (2025), which examines the drivers behind stable real exchange rates of Eastern European countries against the average European country between 1999 and 2020.

Our paper differs from Devereux et al. (2025) in two major ways. First, we study a wider set of countries, including those outside Eastern Europe. In fact, fits of our model for several Eastern European countries, such as Slovakia, are better than those for other countries, which is consistent with Devereux et al. (2025). Second, we use a richer model. We employ N-country models and allow countries to trade in both sectors as in the data, while Devereux et al. (2025) focus on the classical two-country, tradable–non-tradable dichotomy with some cross-sector intermediate input structure. We show that using 2×2 models can overstate the variability of counterfactual values, though the difference is not quantitatively large.

Roadmap. Section 2 presents the theoretical motivation, data, and empirical specifications. We first replicate traditional benchmark-based regressions and show why results hinge on the chosen reference. We then estimate the pair-level model for RER and REER and examine lower-frequency panels. Section 3 develops the quantitative model. Section 4 feeds observed productivity and trade costs into 2×2 and $N \times 2$ environments and compares model-implied REER to the data. Section 5 concludes.

2 Empirical Analysis

2.1 Basic Specification for Testing Balassa–Samuelson Effects

Theoretical Motivation Consider a two-country (i and j) and two-sector ($s = T, NT$) setup following the classical Balassa–Samuelson framework. Each country produces a tradable good (T) and a non-tradable good (NT) using labor as the only input. Labor is perfectly mobile across sectors within each country, implying a single wage w_i. Production in sector s is characterized by productivity $A_{i,s}$, so that unit cost pricing implies

$$P_{i,s} = \frac{w_i}{A_{i,s}}. \tag{1}$$

[3] See Itskhoki (2021) for more comprehensive reviews.

[4] Note that they do not use sectoral productivity.

Under free trade in tradables, the law of one price holds:

$$P_{i,T} = P_{j,T}, \tag{2}$$

which in turn implies that relative wages are pinned down by relative tradable-sector productivity:

$$\frac{w_i}{w_j} = \frac{A_{i,T}}{A_{j,T}}. \tag{3}$$

Given this wage ratio, the relative price of non-tradables between countries i and j is determined by

$$\frac{P_{i,NT}}{P_{j,NT}} = \frac{(w_i/A_{i,NT})}{(w_j/A_{j,NT})} = \frac{(A_{i,T}/A_{j,T})}{(A_{i,NT}/A_{j,NT})}. \tag{4}$$

The aggregate price level in each country is a Cobb–Douglas composite of tradable and non-tradable prices:

$$P_i = P_{i,T}^{\alpha_{i,T}} P_{i,NT}^{\alpha_{i,NT}}, \qquad \text{where } \alpha_{i,T} + \alpha_{i,NT} = 1. \tag{5}$$

The real exchange rate between countries i and j is defined as

$$RER_{i,j} \equiv \frac{P_i}{P_j}. \tag{6}$$

Substituting the expressions above, the real exchange rate becomes

$$RER_{i,j} = \frac{\left(\dfrac{A_{i,T}}{A_{i,NT}}\right)^{\alpha_{i,NT}}}{\left(\dfrac{A_{j,T}}{A_{j,NT}}\right)^{\alpha_{j,NT}}}. \tag{7}$$

Hence, a country experiencing faster productivity growth in tradables relative to non-tradables—compared with its trading partner—will experience a real appreciation. This provides the theoretical foundation for our empirical specification testing the Balassa–Samuelson effects.

2.2 Data Sources and Variable Construction

We construct a panel dataset combining information on real exchange rates and sectoral productivity across countries. The real exchange rate ($RER_{i,t}$) is obtained from the Penn World Table (PWT 11.0) as the ratio of the consumer price index (pl_c) to the nominal exchange rate (xr) relative to the United States (Feenstra et al., 2015).

For sectoral productivity, we draw on multiple harmonized production databases. For Europe, the United States, and Japan, we use the *EU KLEMS* dataset (2008 and 2023 releases).[5] For China, we use the *China Industrial Productivity (CIP)* database from RIETI.[6] For Korea, India, and Taiwan, we use *Asia KLEMS*.[7] Sectoral productivity in each country and year is measured as real value added divided by total hours worked. We refer to this as sectoral Average Labor Productivity (ALP). As a robustness check, we alternatively use a composition-adjusted labor input index, and the results remain similar.

We classify agriculture, mining, and manufacturing as tradable sectors, while services are treated as non-tradable. For each country and year, we aggregate real value added and labor input across tradable (or non-tradable) industries. Aggregate sectoral productivity is then computed as total real value added divided by total labor input within each group.

Country-by-country coverage (start and end years) and data sources are summarized in Table A1 in the Appendix.

To isolate long-run movements consistent with the Balassa–Samuelson mechanism, we remove cyclical components from all productivity and price series using the Hodrick–Prescott filter with a smoothing parameter of $\lambda = 6.25$, following the scaling rule of Ravn and Uhlig (2002). Specifically, we first take logarithms, apply the filter, and then convert the series back to levels. Using alternative filters (e.g., the Christiano–Fitzgerald band-pass filter) yields nearly identical results.

2.3 Traditional Empirical Design and Its Problem

We start with the standard benchmark used in the literature: regressing the log real exchange rate *relative to a fixed reference country U* on the home tradable–non-tradable productivity differential, controlling for country fixed effects (and year effects where noted). Formally,

$$\ln RER_{i,t} = \beta \ln\left(\frac{A_{i,T,t}/A_{i,NT,t}}{A_{U,T,t}/A_{U,NT,t}}\right) + \mu_i + \varepsilon_{i,t}, \tag{8}$$

where $RER_{i,t} = \ln(P_{i,t}/P_{U,t})$, U denotes the reference country, and μ_i are country fixed effects.

[5] EU KLEMS 2008 covers 1970–2005; EU KLEMS 2023 covers 1995–2021. We take 1970–1994 from the 2008 release and 1995–2021 from the 2023 release, and splice at 1995 by multiplicatively normalizing the 1970–1994 series so that the 1995 level matches the 2023 release.

[6] CIP3 covers 1981–2010; CIP4 covers 1987–2017. We take 1981–1986 from CIP3 and 1987–2017 from CIP4, splicing at 1987 by multiplicatively normalizing the 1981–1986 segment to match the 1987 level in CIP4.

[7] Asia KLEMS covers 1980–2012. The data for India do not include hours worked in any period. We instead use value added per worker (composition-adjusted). Excluding India does not change any of the results.

Table 1: Real Exchange Rate and Relative Productivity

	(1)	(2)	(3)	(4)
Log Rel. ALP	-0.15	-0.03	0.39	0.49
	(0.06)	(0.14)	(0.04)	(0.08)
Observations	1,307	849	849	849
Num of Countries	33	33	33	33
Num of Years	52	27	27	27
Sample Years	1970-2021	1995-2021	1995-2021	1995-2021
Ref. Country	U.S.	U.S.	Germany	U.K.

Notes: This table reports panel regressions of the log bilateral real exchange rate against the log relative labor productivity (ALP) differential with the reference country, as specified in equation (8). The dependent variable is $\ln RER_{i,t} = \ln(P_{i,t}/P_{U,t})$, and the explanatory variable is the log difference between tradable and non-tradable productivity in country i relative to reference country U. All regressions include country fixed effects, and standard errors are clustered by country (shown in parentheses). Column (1) uses the United States as the reference over 1970–2021. Column (2) repeats the U.S.-normalized regression for the post-1995 sample to match EU KLEMS 2023 coverage. Column (3) switches the reference to Germany over 1995–2021, and column (4) uses the United Kingdom as an alternative reference country over the same period.

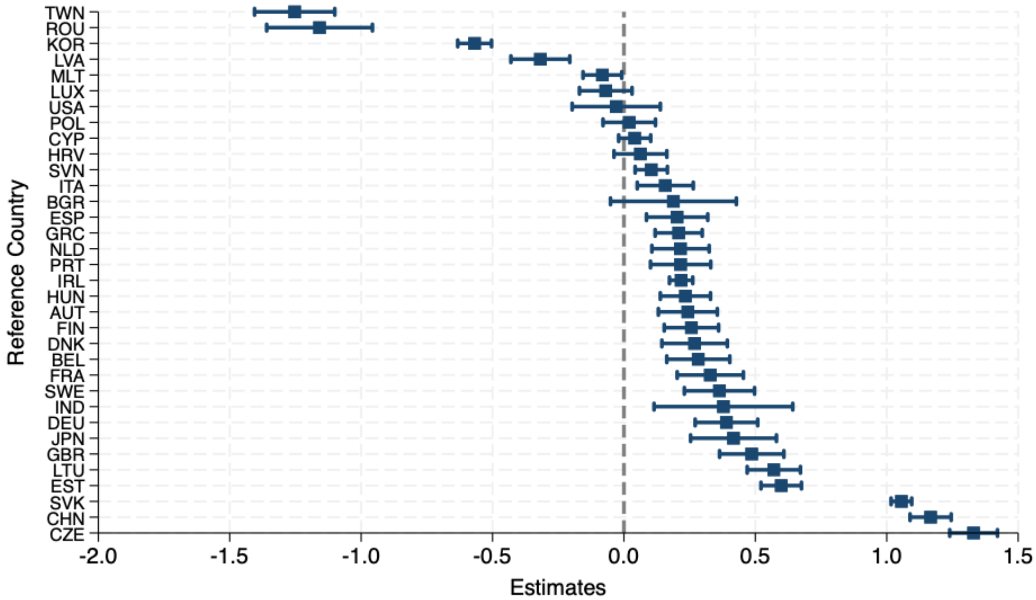
Column (1) of Table 1 implements equation (8) with the United States as the reference over 1970–2021. The estimate is small and statistically indistinguishable from zero, consistent with the mixed findings in U.S.-anchored panels. Column (2) repeats the U.S.-normalized regression for the post-1995 sample (to align with EU KLEMS 2023 coverage); the coefficient remains imprecise (-0.03, s.e. 0.32). This motivates exploring alternative references—both because BS effects may be clearer when nominal exchange-rate noise is muted (as in the euro area; Berka et al. 2018) and because dollar movements can reflect forces unrelated to productivity (Canzoneri et al., 1999).

Columns (3) and (4) switch the reference to Germany and the United Kingdom, respectively, over 1995–2021. The German benchmark yields a larger, positive coefficient (0.39, s.e. 0.27), though it remains statistically indistinguishable from zero.[8] The U.K. benchmark produces a positive and more precisely estimated elasticity (0.49, s.e. 0.23).

To further understand how the choice of reference country influences the results, we repeat the regression in equation (8) separately for each potential benchmark country. This exercise makes explicit how sensitive the estimated Balassa–Samuelson elasticity is to the numeraire adopted in conventional specifications. Earlier studies typically fix the United States as the reference, but there is no reason the theoretical relationship should depend on one country alone. By systematically varying the benchmark, we can assess

[8] While not significant, this is broadly consistent with the view that productivity–RER comovement may be easier to detect in settings closer to a common-currency environment and when sectoral wedges are accounted for (Berka et al., 2018; Devereux et al., 2025).

Figure 2: Balassa–Samuelson Elasticity by Reference Country (Post-1995)



Notes: This figure reports the estimated coefficients from panel regressions of the log bilateral real exchange rate against the log relative labor productivity (ALP) differential with each possible reference country, based on the specification in equation (8). The dependent variable is $\ln RER_{i,t} = \ln(P_{i,t}/P_{U,t})$, and the explanatory variable is the log difference between tradable and non-tradable productivity in country i relative to reference country U. Each horizontal line represents the coefficient and its 95% confidence interval from a separate regression using a different reference country. All regressions include country fixed effects, use data for 1995–2021, and cluster standard errors by country. The vertical line at zero corresponds to the null of no Balassa–Samuelson effect.

whether the lack of support in Table 1 reflects model failure or benchmark-specific noise.

Figure 2 summarizes the estimated coefficients from these regressions. Each horizontal line shows the coefficient on the home tradable–non-tradable productivity differential from a separate regression using a different reference country, with 95% confidence intervals clustered by country pair. The vertical line at zero corresponds to the null of no Balassa–Samuelson effect. Countries are sorted by the magnitude of the estimated elasticity.

The results reveal substantial heterogeneity across reference countries. Only six benchmarks yield positive and statistically significant coefficients consistent with the Balassa–Samuelson prediction, with the U.K. (GBR) among the clearest cases. Two reference countries, Taiwan and Romania, produce negative and significant estimates, while the remaining cases, including the United States and Germany used in Table 1, are statistically indistinguishable from zero. This dispersion highlights how sensitive the estimated elasticity is to the choice of benchmark. Even when the underlying productivity–price relationship exists, its measured strength varies with the reference country, reinforcing the need for a bilateral, pairwise approach that avoids imposing a common numeraire.

2.4 More Robust Specification with Time Fixed Effects

The previous subsection showed that estimates from the traditional single-reference regressions vary widely across benchmarks. To diagnose where this instability comes from and to determine how to fix it, we first make explicit the statistical environment that maps observables to the Balassa–Samuelson elasticity. We therefore state a minimal data-generating process and then study the properties of the conventional estimator under that process:

$$r_{i,t} = \beta\, a_{i,t} + \alpha_i + g_t + u_{i,t}, \qquad a_{i,t} \equiv \ln\left(\tfrac{A_{i,T,t}}{A_{i,NT,t}}\right), \qquad r_{i,t} \equiv \ln RER_{i,t},$$

where β is the elasticity of interest, α_i are country effects, g_t are period-common shocks, and $u_{i,t}$ are idiosyncratic errors. With this DGP in hand, we can compare the traditional estimator to a specification with time fixed effects and show precisely which features of the DGP each estimator does or does not absorb.

Traditional estimator. Fix a reference country U and define

$$y_{i,t} \equiv r_{i,t} - r_{U,t}, \qquad x_{i,t} \equiv a_{i,t} - a_{U,t}.$$

Then

$$y_{i,t} = \beta\, x_{i,t} + (\alpha_i - \alpha_U) + (u_{i,t} - u_{U,t}).$$

Estimating with country fixed effects only gives the within equation

$$\tilde{y}_{i,t} = \beta\, \tilde{x}_{i,t} + \tilde{\varepsilon}_{i,t},$$

with

$$\tilde{x}_{i,t} = (a_{i,t} - \bar{a}_i) - (a_{U,t} - \bar{a}_U), \qquad \tilde{\varepsilon}_{i,t} = (u_{i,t} - \bar{u}_i) - (u_{U,t} - \bar{u}_U).$$

The estimator is

$$\hat{\beta} = \beta + \frac{S_{x\varepsilon}}{S_{xx}}, \qquad S_{x\varepsilon} \equiv \sum_{i,t} \tilde{x}_{i,t}\tilde{\varepsilon}_{i,t}, \quad S_{xx} \equiv \sum_{i,t} \tilde{x}_{i,t}^2.$$

Properties of the traditional estimator. Expand the leading pieces of the score and the quadratic form:

$$S_{x\varepsilon} = \sum_{i,t}(a_{i,t} - \bar{a}_i)(u_{i,t} - \bar{u}_i) \; - \; \sum_{i,t}(a_{i,t} - \bar{a}_i)(u_{U,t} - \bar{u}_U)$$

$$- \; \sum_{i,t}(a_{U,t} - \bar{a}_U)(u_{i,t} - \bar{u}_i) \; + \; N\sum_{t}(a_{U,t} - \bar{a}_U)(u_{U,t} - \bar{u}_U),$$

$$S_{xx} = \sum_{i,t}(a_{i,t} - \bar{a}_i)^2 \; + \; N\sum_{t}(a_{U,t} - \bar{a}_U)^2 \; + \; \text{cross terms.}$$

Bias. There are two regimes. First, if strict exogeneity holds for all countries including U, then each summand has zero expectation conditional on the full history of a, so $E[\hat{\beta}] = \beta$. Second, if the benchmark U fails strict exogeneity so that $\sum_t (a_{U,t} - \bar{a}_U)(u_{U,t} - \bar{u}_U) \neq 0$, then the last term in $S_{x\varepsilon}$ has nonzero expectation and scales with N. Because the same factor appears in S_{xx}, the probability limit shifts to a benchmark-driven value that differs from β and the estimator is asymptotically biased.

Consistency. Even under strict exogeneity for all countries including U, the dominant random variation in $S_{x\varepsilon}$ comes from period-common components that repeat across all i. These parts do not wash out by adding countries. With $N, T \to \infty$,

$$\hat{\beta} - \beta = \frac{O_p(N\sqrt{T})}{O_p(NT)} + o_p(1) = O_p\Big(\tfrac{1}{\sqrt{T}}\Big),$$

so consistency requires $T \to \infty$, and increasing N alone does not help.

Efficiency. Because period-common pieces dominate the score, information accumulates in the time dimension. Under weak dependence,

$$\sqrt{T}\,(\hat{\beta} - \beta) \;\Rightarrow\; \mathcal{N}(0, \sigma^2_{\text{bench}}),$$

so the rate is $\sqrt{T}$ rather than $\sqrt{NT}$.

Time fixed effects: specification, identification, and rate. Include time effects while keeping the single-reference structure:

$$r_{i,t} - r_{U,t} = \beta\big(a_{i,t} - a_{U,t}\big) + \mu_i + \tau_t + \varepsilon_{i,t}. \tag{9}$$

Let $\ddot{z}_{i,t}$ be residuals after removing μ_i and τ_t. Then

$$\ddot{r}_{i,t} = \beta\,\ddot{a}_{i,t} + \ddot{\varepsilon}_{i,t}.$$

Time effects remove all period-common movements from both the regressor and the disturbance. This eliminates the benchmark-common channels in the score and restores the orthogonality $E[\ddot{a}_{i,t}\ddot{\varepsilon}_{i,t}] = 0$ under the same strict exogeneity stated for the double-demeaned variables. Hence

$$\operatorname{plim} \hat{\beta} = \beta.$$

With period-common components purged, identification uses within-year cross-sectional deviations and accumulates over countries and years. Under weak dependence,

$$\sqrt{NT}\,(\hat{\beta} - \beta) \;\Rightarrow\; \mathcal{N}(0, \sigma^2_{\text{timeFE}}),$$

so time effects resolve the benchmark-driven sensitivity and deliver the panel rate of convergence.

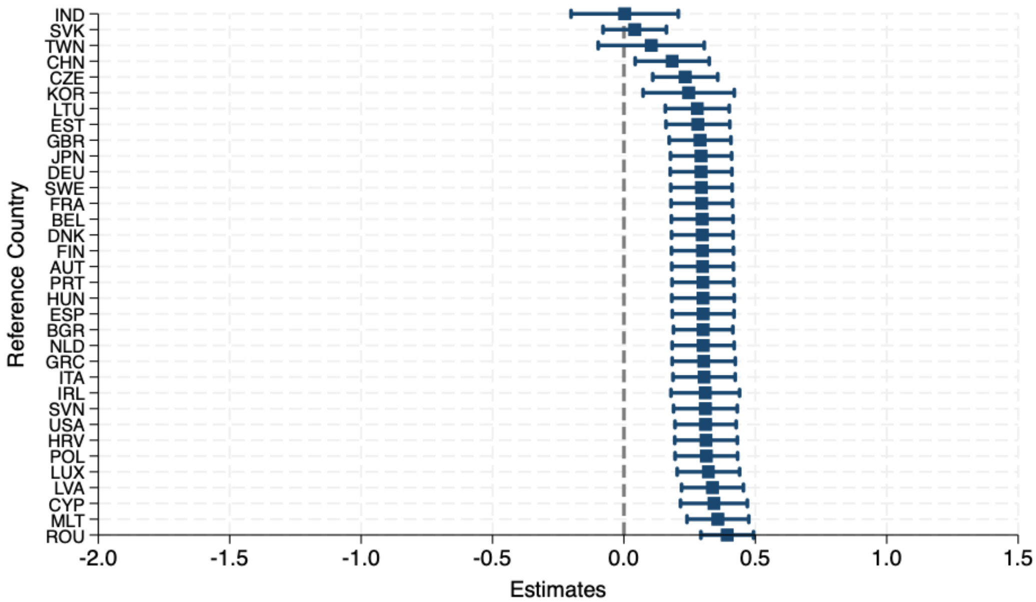
Main result. Figure 3 plots the elasticity from equation (9) for each benchmark over 1995–2021. The dispersion across benchmarks collapses once time effects are included. Coefficients are stable in sign and magnitude and no longer hinge on the numeraire.

2.5 Robustness for the Main Time-FE Specification

Alternative samples. We test whether the main time-FE result depends on sample composition. We first extend the window to 1970–2021 to use all available observations. We then restrict to the post-1995 period to align with EU KLEMS 2023 coverage. We next focus on advanced economies where measurement and institutions are more comparable across countries. Finally, we impose a balanced panel from 1995 to 2021 to hold composition fixed over time. Table 2 shows a stable and positive elasticity across all four choices. Magnitudes vary modestly but the confidence intervals overlap widely, indicating that the time-FE result is not driven by sample selection.

Penn effect. We ask whether the single-reference result is absorbed by aggregate productivity. We augment the regression with the relative GDP per worker from the Penn World Table and re-estimate on the post-1995 period as well as the extended 1970–2021 window. Table 3 shows that the Balassa–Samuelson coefficient remains positive and precisely estimated after adding the Penn effect control. The Penn coefficient weakens and can change sign when extending the window back to 1970, whereas the BS term remains stable.

Figure 3: Balassa–Samuelson elasticity by reference country, 1995–2021, with time fixed effects



Notes: Each line reports the coefficient from a separate panel regression of the log bilateral real exchange rate relative to a given benchmark on the log tradables–nontradables productivity differential relative to that benchmark, as in equation (9). The dependent variable is $\ln RER_{i,t} - \ln RER_{U,t}$. The regressor is $\ln(A_{i,T,t}/A_{i,NT,t}) - \ln(A_{U,T,t}/A_{U,NT,t})$. The sample is 1995–2021. All regressions include country fixed effects and time fixed effects. Standard errors are panel-corrected to allow correlation across countries within a year using the Beck–Katz method with period correlation. Countries are sorted by the point estimate. The vertical line at zero marks the null of no Balassa–Samuelson effect.

Table 2: Single-reference time-FE regressions: alternative samples

	(1)	(2)	(3)	(4)
Log Rel. ALP	0.10	0.31	0.32	0.34
	(0.04)	(0.06)	(0.06)	(0.06)
Observations	1,307	849	778	756
Sample Countries	All	All	Adv	Balanced
Num of Countries	33	33	29	28
Sample Years	1970-2021	1995-2021	1995-2021	1995-2021
Num of Years	52	27	27	27
Country & Year FE	✓	✓	✓	✓

Notes: This table reports panel regressions based on equation (9) with country fixed effects and time fixed effects. The dependent variable is $\ln RER_{i,t} - \ln RER_{U,t}$ and the regressor is $\ln(A_{i,T,t}/A_{i,NT,t}) - \ln(A_{U,T,t}/A_{U,NT,t})$. Column (1) uses the full 1970–2021 sample. Column (2) restricts to 1995–2021 to match EU KLEMS 2023 coverage. Column (3) restricts to advanced economies. Column (4) uses a balanced panel for 1995–2021. Standard errors are panel-corrected to allow correlation across countries within a year using the Beck–Katz method with period correlation.

Table 3: Single-reference time-FE regressions with Penn effect control

	(1)	(2)	(3)	(4)
Log Rel. ALP	0.31		0.26	0.25
	(0.06)		(0.07)	(0.05)
Log Rel. GDP per Workers		0.22	0.11	-0.33
		(0.08)	(0.09)	(0.11)
Observations	849	849	849	1,307
Sample Countries	All	All	All	All
Num of Countries	33	33	33	33
Sample Years	1995-2021	1995-2021	1995-2021	1970-2021
Num of Years	27	27	27	52
Country & Year FE	✓	✓	✓	✓

Notes: This table reports panel regressions based on equation (9) augmented with $\ln(\text{GDPpw}_{i,t}) - \ln(\text{GDPpw}_{U,t})$ from the Penn World Table. All specifications include country fixed effects and time fixed effects. Columns report estimates for 1995–2021 and for 1970–2021, as indicated in the table body. The dependent variable is $\ln RER_{i,t} - \ln RER_{U,t}$. The BS regressor is $\ln(A_{i,T,t}/A_{i,NT,t}) - \ln(A_{U,T,t}/A_{U,NT,t})$. Standard errors are panel-corrected to allow correlation across countries within a year using the Beck–Katz method with period correlation.

Different data. We test portability to a broader setting. We switch to the GGDC 10-Sector Database which extends coverage to a larger set of economies and a longer period. We estimate the same time-FE specification with and without the Penn effect control on the full 1960–2013 span and on the post-1995 subset. Table 4 indicates that the elasticity remains positive across both samples. The magnitude is smaller in the late period but the sign and significance persist, showing that the time-FE design is robust to alternative sources and wider coverage.

Labor composition adjustment. We ask whether changing the labor input measure affects the time-FE result. We replace hours with the KLEMS composition-adjusted labor index that tracks changes in worker composition over time. We re-estimate the same single-reference time-FE specification on the standard sample set so that the only change is the labor input definition. Table 5 shows that the elasticity remains positive and similar in magnitude across columns. This indicates that the result is not driven by how labor input is measured.

Table 4: Single-reference time-FE regressions: GGDC 10-sector data

	(1)	(2)	(3)	(4)
Log Rel. ALP	0.84	0.45	0.65	0.38
	(0.07)	(0.06)	(0.07)	(0.06)
Log Rel. GDP per Workers			1.42	-0.36
			(0.13)	(0.06)
Observations	2,106	741	2,106	741
Sample Countries	All	All	All	All
Num of Countries	39	39	39	39
Sample Years	1960-2013	1995-2013	1960-2013	1995-2013
Num of Years	54	19	54	19
Country & Year FE	✓	✓	✓	✓

Notes: This table reports panel regressions based on equation (9) estimated on the GGDC 10-Sector Database. All specifications include country fixed effects and time fixed effects. Columns report results for 1960–2013 and for 1995–2013, each without and with the Penn effect control as indicated in the table body. The dependent variable is $\ln RER_{i,t} - \ln RER_{U,t}$. The BS regressor is $\ln(A_{i,T,t}/A_{i,NT,t}) - \ln(A_{U,T,t}/A_{U,NT,t})$. Standard errors are panel-corrected to allow correlation across countries within a year using the Beck–Katz method with period correlation.

Table 5: Single-reference time-FE regressions: labor composition adjusted

	(1)	(2)	(3)	(4)
Log Rel. ALP	0.74	0.76	0.73	0.91
	(0.05)	(0.05)	(0.05)	(0.04)
Observations	654	583	530	324
Sample Countries	All	Adv.	Europe	Balanced
Num of Countries	33	29	27	12
Sample Years	1995-2021	1995-2021	1995-2021	1995-2021
Num of Years	27	27	27	27
Country & Year FE	✓	✓	✓	✓

Notes: This table reports panel regressions based on equation (9) where labor productivity uses the KLEMS composition-adjusted labor index in place of hours. All specifications include country fixed effects and time fixed effects. Column (1) covers 1970–2021. Column (2) restricts to 1995–2021. Column (3) restricts to advanced economies. Column (4) is a balanced panel for 1995–2021. The dependent variable is $\ln RER_{i,t} - \ln RER_{U,t}$. The Balassa–Samuelson regressor is $\ln(A_{i,T,t}/A_{i,NT,t}) - \ln(A_{U,T,t}/A_{U,NT,t})$. Standard errors are panel-corrected to allow correlation across countries within a year using the Beck–Katz method with period correlation.

3 Model Setup

The empirical results point to a qualitative Balassa–Samuelson (BS) relationship. To quantify its contribution to price differences across countries, we build a quantitative Armington trade model that features trade frictions and sectoral productivity differences. [9]

3.1 Countries and Sectors

Countries are indexed by $i, j = 1, \dots, N$. Sectors are indexed by $s = 1, \dots, S$. In the quantitative exercises, we focus on $S = 2$, which we label G (goods) and S (services). All prices are expressed in U.S. dollars.

3.2 Preferences

Households in country j have Cobb–Douglas preferences over sectoral composites,

$$U_j = \prod_{s=1}^{S} \left(\frac{C_{j,s}}{\alpha_{j,s}} \right)^{\alpha_{j,s}}, \qquad \sum_{s=1}^{S} \alpha_{j,s} = 1,$$

where $C_{j,s}$ is consumption of sector s in country j, and $\alpha_{j,s}$ are (possibly country-specific) expenditure shares. Let C_j denote aggregate final demand in country j. Optimal allocation implies

$$P_{j,s} C_{j,s} = \alpha_{j,s} P_j C_j, \qquad P_j = \prod_{s=1}^{S} P_{j,s}^{\alpha_{j,s}}.$$

Within each sector s, final demand is a CES composite of varieties produced in different countries,

$$C_{j,s} = \left(\sum_{i=1}^{N} \mu_{i,j,s}^{1/\theta} C_{i,j,s}^{(\theta-1)/\theta} \right)^{\theta/(\theta-1)}, \qquad \theta > 1,$$

where θ is the elasticity of substitution and $\mu_{i,j,s} > 0$ are taste (quality) shifters. Non-tradables are obtained as a limiting case: for $s = \text{NT}$, demand is restricted to the domestic variety only by setting $\mu_{i,j,\text{NT}} = 1$ if $i = j$ and $\mu_{i,j,\text{NT}} = 0$ otherwise.

The corresponding sectoral price index is

$$P_{j,s} = \left(\sum_{i=1}^{N} \mu_{i,j,s} P_{i,j,s}^{1-\theta} \right)^{1/(1-\theta)},$$

[9] In Appendix D, we show a version with input–output linkages.

where $P_{i,j,s}$ is the delivered price in country j of the good produced in country i in sector s.

The implied expenditure share of country i in country j's spending on sector s is

$$\pi_{i,j,s} \equiv \frac{P_{i,j,s}C_{i,j,s}}{\sum_{\ell=1}^{N} P_{\ell,j,s}C_{\ell,j,s}} = \frac{\mu_{i,j,s}P_{i,j,s}^{1-\theta}}{\sum_{\ell=1}^{N} \mu_{\ell,j,s}P_{\ell,j,s}^{1-\theta}}.$$

3.3 Technology and Pricing

Labor is the only factor of production. Output in country i and sector s is

$$Y_{i,s} = A_{i,s}L_{i,s},$$

where $A_{i,s}$ is labor productivity and $L_{i,s}$ is labor used in that sector. The unit cost is

$$c_{i,s} = \frac{w_i}{A_{i,s}},$$

where w_i is the wage in country i.

3.4 Trade Costs

Iceberg trade costs $\tau_{i,j,s} \geq 1$ apply to shipments from country i to country j in sector s. Delivering one unit in j requires shipping $\tau_{i,j,s}$ units from i. The delivered (buyer-facing) price is

$$P_{i,j,s} = \tau_{i,j,s}c_{i,s} = \frac{w_i\tau_{i,j,s}}{A_{i,s}}.$$

Frictionless trade corresponds to $\tau_{i,j,s} \equiv 1$. Non-tradables are captured by shutting down imports in the non-tradable sector, i.e., by setting $\mu_{i,j,\mathrm{NT}} = 1$ if $i = j$ and $\mu_{i,j,\mathrm{NT}} = 0$ otherwise.

3.5 Labor and Income

Total labor supply in country i is $L_i = \sum_s L_{i,s}$ and is taken as exogenous. Labor is perfectly mobile across sectors within a country, so a single wage w_i clears all sectoral labor markets in country i.

We allow for exogenous trade imbalances. Let β_j denote country j's trade surplus as a fraction of world income, scaled by

$$Y \equiv \sum_{\ell=1}^{N} w_\ell L_\ell.$$

By construction, the surpluses sum to zero,

$$\sum_{\ell=1}^{N} \beta_\ell = 0,$$

so that world absorption equals world income. Nominal expenditure (absorption) in country j is

$$P_j C_j = w_j L_j - \beta_j Y.$$

A country with $\beta_j > 0$ runs a surplus and spends less than its income; a country with $\beta_j < 0$ runs a deficit and spends more than its income. When we normalize world income to one, $Y = 1$, this becomes

$$P_j C_j = w_j L_j - \beta_j.$$

3.6 Equilibrium

An equilibrium is a set $\{w_i, P_{j,s}, \pi_{i,j,s}\}_{i,j,s}$ such that goods markets clear, expenditure shares are consistent with CES demand, and the exogenous trade surpluses $\{\beta_i\}$ are respected.

Let

$$Y \equiv \sum_{\ell=1}^{N} w_\ell L_\ell$$

denote aggregate world income. Since β_i is defined as the trade surplus of country i as a share of Y, $\beta_i > 0$ means that country i exports more than it absorbs, and $\sum_i \beta_i = 0$.

Goods-market clearing (in revenue) for exporter i requires

$$\sum_{j=1}^{N} \sum_{s=1}^{S} \pi_{i,j,s} \alpha_{j,s} w_j L_j = w_i L_i + \beta_i Y. \tag{10}$$

The left-hand side is total revenue that country i earns from selling to all destinations and sectors. The right-hand side is factor income $w_i L_i$ plus the exogenous trade surplus $\beta_i Y$. Because the surpluses sum to zero, summing (10) over i yields $Y = Y$.

Bilateral expenditure shares are given by CES demand and delivered prices:

$$\pi_{i,j,s} = \frac{\mu_{i,j,s}\left(\frac{w_i \tau_{i,j,s}}{A_{i,s}}\right)^{1-\theta}}{\sum_{\ell=1}^{N} \mu_{\ell,j,s}\left(\frac{w_\ell \tau_{\ell,j,s}}{A_{\ell,s}}\right)^{1-\theta}}. \tag{11}$$

Given $\{\pi_{i,j,s}\}$ from (11), solving (10) for $\{w_i\}$ delivers sectoral price indices $\{P_{j,s}\}$ and, in turn, the consumer price index,

$$P_j = \prod_{s=1}^{S} P_{j,s}^{\alpha_{j,s}}.$$

The CPI-based bilateral real exchange rate between countries i and j is

$$RER_{i,j} = \frac{P_i}{P_j},$$

which is our target.

4 Quantitative Importance of the Balassa–Samuelson Force

4.1 Overview

This section implements the model in a sequence of experiments that increase in realism and data content. We first study a traditional 2×2 free-trade economy (goods sector as tradables T and service sector as non-tradables NT; two countries) to isolate the Balassa–Samuelson mechanism in its simplest form. We then introduce iceberg trade costs in the same 2×2 environment to quantify how costly trade attenuates the transmission from sectoral productivity to prices and the real exchange rate. Finally, we move to the full $N \times 2$ model disciplined by data for a large set of countries, and—when presenting results—we also report an $N \times 2$ specification with input–output linkages. The IO extension is used for quantitative exercises in the main text, while its structure and solution are detailed in Appendix D. In each case, we solve the static equilibrium year by year, recover sectoral price indices and the CPI, and construct $REER_{i,t}$.

We keep the baseline structure deliberately minimal—one factor, Cobb–Douglas across sectors, and Armington within sectors—so that the quantitative contribution of sectoral productivity $\{A_{i,s,t}\}$ and iceberg costs $\{\tau_{i,j,s,t}\}$ to real exchange rates is transparent and directly comparable to the reduced-form evidence. When we employ the IO extension in the $N \times 2$ environment, we maintain the same calibration strategy; implementation details are in Appendix D.

We calibrate the model to 1995 and conduct counterfactual simulations using exact hat algebra, with $\hat{x} \equiv x'/x$. From the data in 1995, we construct share parameters $\{\alpha_{j,s}\}$ and $\{\mu_{i,j,s}\}$. We set $\theta = 5$. The shocks are sectoral productivity $\{\widehat{A}_{i,s}\}$ and iceberg trade costs $\{\widehat{\tau}_{i,j,s}\}$. See Appendix C for details.

4.2 Shock Feeds

Productivity. Sectoral labor productivity $\{\widehat{A}_{i,s,t}\}$ is taken from the same sources as in the empirical section: EU KLEMS, Asia KLEMS (Korea, Taiwan, India), and the RIETI China Industrial Productivity database. We compute sectoral $A_{i,s,t}$ as real value added divided by labor input, aggregate industries into T (agriculture, mining, manufacturing) and NT (services), and apply the Hodrick–Prescott filter with $\lambda = 6.25$ to the log series to isolate low-frequency movements.

When the model includes input–output linkages, we use an *intermediate-adjusted productivity* that accounts for the role of intermediate inputs in production. Specifically, using cost shares from KLEMS, we back out effective productivity as

$$\ln A^{IO}_{i,s,t} = \ln \text{Gross Output}_{i,s,t} - \gamma_{i,s} \ln L_{i,s,t} - (1 - \gamma_{i,s}) \ln \text{Intermediate}_{i,s,t},$$

where $\gamma_{i,s}$ is the labor cost share. This adjustment ensures that productivity growth reflects efficiency gains net of changes in intermediate input prices, consistent with the IO model's unit-cost structure. We construct $\widehat{A}^{IO}_{i,s,t}$ as the ratio of the filtered series between t and 1995.

Trade Costs. Bilateral iceberg trade costs for tradables are inferred from the OECD Inter-Country Input–Output (ICIO) database (2025 Extended Edition). For the baseline model, we compute $\{\widehat{\tau}_{i,j,s,t}\}$ from total bilateral flows following Head and Ries (2001). When using the IO model, we separately back out trade costs for final and intermediate goods, $\widehat{\tau}^{F}_{i,j,s,t}$ and $\widehat{\tau}^{X}_{i,j,s,t}$, using bilateral final-use and intermediate-use trade flows from ICIO. Both sets of trade-cost series are HP-filtered, normalized to 1995, and used as shocks in the quantitative exercises.

Trade Costs. Bilateral iceberg trade costs $\{\widehat{\tau}_{i,j,s,t}\}$ for tradables are inferred from trade data using the OECD Inter-Country Input–Output (ICIO) database (2025 Extended Edition). The ICIO provides bilateral trade flows by sector. We classify agriculture, mining, and manufacturing as goods sectors, and all remaining sectors as services.

In the model with input–output linkages, we separately calibrate iceberg costs for *final* and *intermediate* goods, denoted $\tau^F_{i,j,s,t}$ and $\tau^X_{i,j,s,t}$, respectively. Both are recovered from bilateral trade flows using the standard gravity-based inversion method. In counterfactual simulations, we allow both $\widehat{\tau}^F_{i,j,s,t}$ and $\widehat{\tau}^X_{i,j,s,t}$ to evolve along their estimated low-frequency paths, so that trade frictions adjust consistently across final and intermediate markets.

Following Head and Ries (2001), we back out $\{\tau_{i,j,s,t}\}$ using

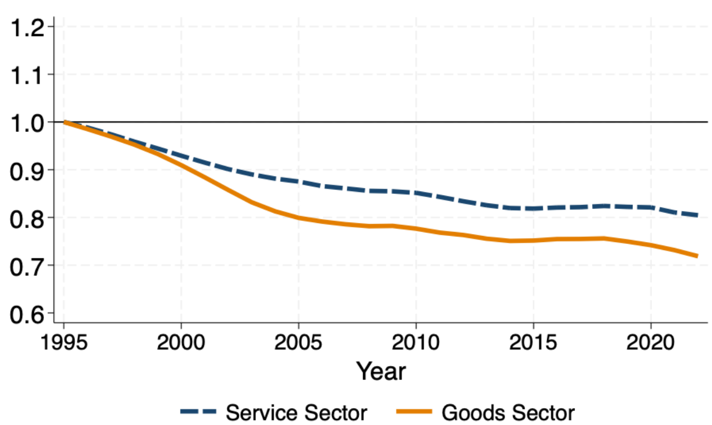
$$(\tau_{ijst})^{1-\theta} = \sqrt{\frac{X_{ijst}X_{jist}}{X_{iist}X_{jjst}}},$$

where X_{ijst} is a gross trade flow from i to j in sector s in year t.

As with productivity, we work with the low-frequency component of $\ln \tau_{i,j,T,t}$ obtained by applying the HP filter with $\lambda = 6.25$. We then normalize the level relative to 1995 and obtain $\widehat{\tau}_{i,j,s}$.

As an illustration, Figure 4 shows the time series of estimated trade costs for China. Since 1995, trade costs in the goods sector from China have declined by about 28% by 2021, while service sectors have experienced about an 18% decline. This indicates that, all else equal, the reduction in trade costs operates as if productivity in the goods sector increased more than in the service sector.

Figure 4: Estimated Trade Costs (China as an Exporter)



Notes: This figure shows the estimated paths of $\widehat{\tau}$ for China as an exporter over time. We take the weighted average within each sector using the share of bilateral flows as weights.

Normalization and Timing. We simulate the model year by year over 1995–2021. For countries with missing data, we log-linearly extrapolate the series to obtain a balanced panel. All nominal variables are expressed in U.S. dollars; the model determines wages $\{w_{i,t}\}$ and price indices $\{P_{i,s,t}\}$ up to a common scalar each year. We remove this indeterminacy by fixing world GDP to one. Country-specific expenditure weights $\{\alpha_{i,s}\}$ are set to average expenditure shares in the data. The trade elasticity is set to $\theta = 5.0$, which is standard.

4.3 Results

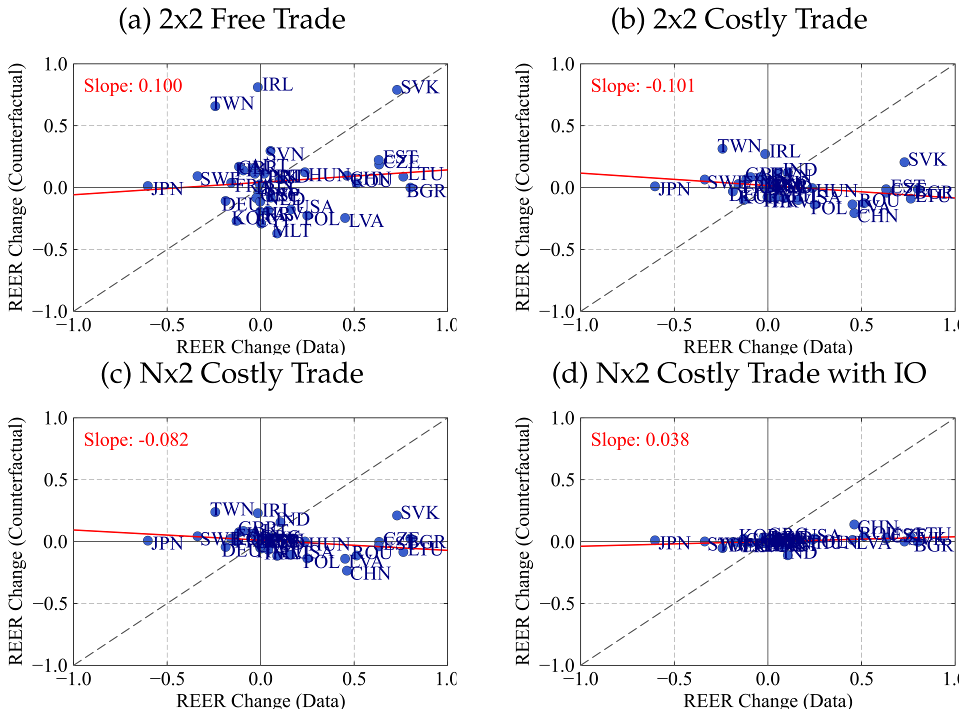
Figure 5 summarizes our first counterfactual. We feed only sectoral productivity shocks—that is, we allow $\{A_{i,s,t}\}$ to evolve as in the data, hold trade costs $\{\tau_{i,j,s,t}\}$ fixed at their 1995 values, and solve the model year by year. We then compare, for each country, the long difference in its real effective exchange rate (REER) between 1995 and 2021 in the data with the corresponding long difference implied by the model. Each panel of Figure 5 reports this comparison under a different version of the model: a 2×2 free-trade case, a 2×2 case with iceberg trade costs, an $N \times 2$ costly-trade case that allows for third-country reallocation, and an $N \times 2$ version that also incorporates input–output linkages.[10] Points on the 45-degree line would indicate that the model with productivity shocks alone reproduces the observed REER movement.

The scatter plots in Figure 5 show that productivity shocks alone cannot replicate observed REER changes. Large appreciations and depreciations in the data translate into much smaller movements in the model, even in the richer $N \times 2$ environments. In many cases, the sign is already off: some countries that appreciate in the data appear as mild depreciations in the model, and vice versa. In other cases, the sign is correct, but the slope is far too flat—the model's long-run REER movement is only a fraction of what is observed.

To examine the time dimension behind these long differences, Appendix Figure A1 plots full REER paths for six large economies—China, Germany, France, India, Japan, and the United States—under the productivity-shock experiment. In each panel, we show the HP-filtered REER from the data (normalized to one in 1995) and the model-implied REER paths generated by feeding only $\{A_{i,s,t}\}$. The qualitative failure is clearest in Japan: the data exhibit a depreciation of more than 0.7 log points between 1995 and 2021, while the model predicts an appreciation. China moves the wrong way as well: the REER in the data appreciates sharply, whereas the model delivers a slight depreciation. Some countries move in the right direction qualitatively—for example, Germany's observed appreciation is upward in the model—but even there, the amplitude is far too small. Only in a few cases, such as India, do the counterfactuals roughly trace the movements. Over-

[10] See Appendix D for model details.

Figure 5: REER Long-Difference Fit vs. Data, 1995–2021: Productivity Shock Only



Notes: Each dot represents one country. The horizontal axis is the HP-filtered change in log REER between 1995 and 2021 in the data (BIS). The vertical axis is the corresponding change implied by the model when only sectoral productivity $\{A_{i,s,t}\}$ is allowed to move, while iceberg trade costs $\{\tau_{i,j,s,t}\}$ are fixed at 1995 levels. Panels differ by model environment. The 45-degree line indicates a perfect quantitative match.

all, feeding observed sectoral productivity—the traditional Balassa–Samuelson channel—rarely gets the sign right and is far too weak quantitatively.

4.4 Robustness

Adding Trade-Cost Shocks. Figure A2 repeats the long-difference exercise after adding trade-cost shocks. Bilateral iceberg trade costs $\{\tau_{i,j,s,t}\}$ evolve along their estimated low-frequency paths. We solve the model year by year, compute sectoral price indices and the CPI, and compare each country's 1995–2021 change in the data with the model-implied change. We report three environments: 2×2 with costly trade, $N \times 2$ with costly trade, and $N \times 2$ with costly trade plus input–output linkages.

Results are similar to Figure 5. Allowing time variation in $\{\tau_{i,j,s,t}\}$ delivers only minor improvements, and the quantitative fit remains weak. Large appreciations and depreciations in the data translate into muted changes in the model-implied series. The overall conclusion is unchanged: adding trade-cost reductions does not account for the observed scale of real effective exchange rate movements.

Timing of Long Differences. Figure 5 focused on the 1995–2021 long difference. Figure A3 reports the results for alternative windows, including 1995–2019, 2000–2021, and 2000–2019, using the $N \times 2$ costly-trade environment.

Results are stable across windows. The model's quantitative shortfall is not driven by the particular long-difference period. Estimated slopes of model-implied versus observed changes remain close to zero, with wide dispersion around the 45-degree line, confirming that the weak fit is not a window-specific artifact.

4.5 Productivity and Sectoral Prices

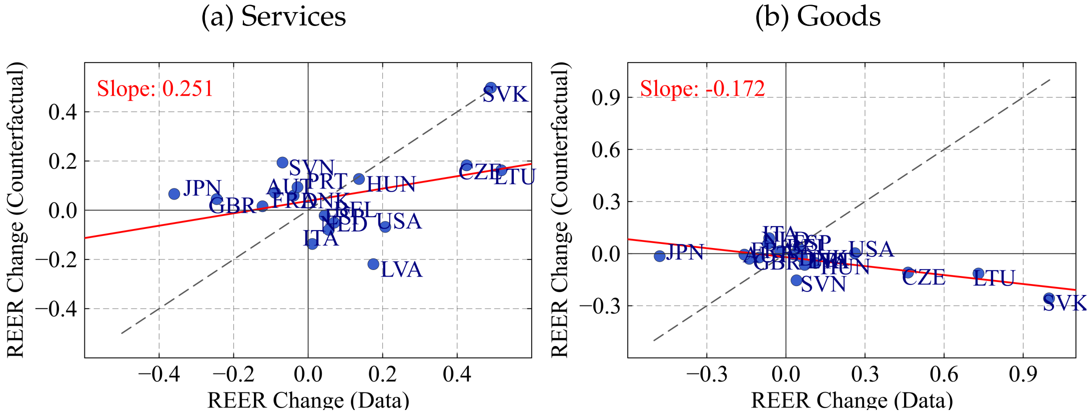
The previous subsection shows that the Balassa–Samuelson mechanism cannot replicate the magnitude—and in some cases even the sign—of international price movements for aggregate CPI. Prior work reports that the Balassa–Samuelson relationship is not evident for aggregate real exchange rates but appears qualitatively in sectoral prices, especially for services (Engel, 1999; Canzoneri et al., 1999), and in some cases for goods (Lee and Tang, 2007). We examine whether observed sectoral productivity changes can quantitatively match international differences in sectoral prices.

Additional Data Sources. Cross-country sectoral CPI coverage is limited. We use the OECD CPI by sector for 17 countries between 1997 and 2019, including the United States, Japan, the United Kingdom, and 14 EU economies. For services, we use services CPI excluding housing when available; otherwise, we use the broad services CPI. For goods, we recover a goods CPI using total CPI and expenditure-share weights from the same source. We then convert country-level sectoral CPIs into effective terms using the BIS double-weighting scheme so that sectoral price indices are comparable across countries. As with aggregate CPI, we work with HP-filtered log series to isolate low-frequency movements.

Results. We run the same quantitative exercises as for aggregate CPI, feeding sectoral productivity shocks with time-varying trade costs in the $N \times 2$ environment for the 17-country sample. Figure 6 reports the long-difference comparison for services and goods.

At the sectoral level, service prices move in the expected direction with sectoral productivity and align better than aggregate CPI in qualitative terms, but the magnitudes remain too small. Goods prices do not improve the quantitative fit. Allowing trade costs to vary jointly with sectoral productivity does not alter this conclusion.

Figure 6: Sectoral Effective Prices: Model Fit vs. Data, Long Differences 1997–2019



Notes: Each dot represents one country. The horizontal axis is the HP-filtered change in the sectoral effective price index between 1997 and 2019, constructed from OECD sectoral CPIs. For services, we use services CPI excluding housing when available; otherwise, services CPI. For goods, we recover a goods CPI from total CPI and expenditure-share weights from OECD. Country-level sectoral CPIs are converted to effective terms using BIS double weighting, following the multilateral index construction in Klau and Fung (2006). The vertical axis shows the corresponding model-implied change from the $N \times 2$ environment when sectoral productivities $\{A_{i,s,t}\}$ evolve along their estimated low-frequency paths, with time variation in bilateral iceberg trade costs $\{\tau_{i,j,s,t}\}$. A 45-degree line would indicate a perfect quantitative match. Services display the expected sign but muted magnitudes; goods show no quantitative improvement.

4.6 Discussion: Perfect Competition and TFP Measurement.

Our baseline maintains perfect competition. A natural concern is that the model's difficulty matching long-run RER paths could instead reflect imperfect competition in product or factor markets (markups or labor markdowns), suggesting that one might introduce wedges and replace sectoral productivities with "markup-adjusted TFP." The measurement problem is that EU KLEMS/INTANProd growth accounting constructs capital compensation using an *internal* rate of return that exhausts non-labor income, so revenue shares proxy output elasticities only under perfect competition; with markups or markdowns, revenue-based residuals mix technology with wedges and non-CRS forces, and a markup-consistent A is not identified from KLEMS alone (Bontadini et al., 2023).[11] Recent aggregation methods show how to incorporate market power using firm-level markups and imputed input shares for the United States (Baqaee and Farhi, 2020), but extending this approach country-by-country would require comparable firm-level coverage and harmonized capital-cost imputations that are not available across our panel. Therefore, we retain perfect competition as our maintained assumption and interpret sectoral A from KLEMS accordingly.

[11] See Takahashi and Takayama (2025a,b) for detailed discussion on this point.

5 Conclusion

This paper places the Balassa–Samuelson hypothesis on an empirically cleaner footing and examines its quantitative relevance. Pairwise panel estimates deliver the expected signs and symmetry, overturning the mixed, sample-dependent results from the single-benchmark panels commonly used in the literature. However, when we translate these elasticities into country-level paths, the quantitative impact is small. The shortfall is most striking for Japan: a 0.7-log depreciation in the data between 1995 and 2021 is essentially invisible to the standard sectoral-productivity Balassa–Samuelson channel.

A standard multi-country Armington model reinforces this message. Feeding observed productivity growth does not replicate the paths of real exchange rates. Matching both the level and slope of RERs likely requires additional ingredients—time-varying wedges (Devereux et al., 2025), demand shifts (Bergstrand, 1991), factor intensity (Bhagwati, 1984), sectoral markups, demographics, or terms-of-trade movements. These remain promising directions for future research.

References

Baqaee, D. R. and Farhi, E. (2020). Productivity and misallocation in general equilibrium. *The Quarterly Journal of Economics*, 135(1):105–163.

Bergstrand, J. H. (1991). Structural determinants of real exchange rates and national price levels: Some empirical evidence. *American Economic Review*, 81(1):325–334.

Berka, M., Devereux, M. B., and Engel, C. (2018). Real exchange rates and sectoral productivity in the euro area. *American Economic Review*, 108(6):1543–1581.

Berka, M. and Steenkamp, D. (2018). Deviations in real exchange rate levels in the oecd countries and their structural determinants. Technical report, Reserve Bank of New Zealand.

Bhagwati, J. N. (1984). Why are services cheaper in the poor countries? *The Economic Journal*, 94(374):279–286.

Bontadini, F., Corrado, C., Haskel, J., Iommi, M., and Jona-Lasinio, C. (2023). Euklems & intanprod: industry productivity accounts with intangibles. sources of growth and productivity trends: methods and main measurement challenges (deliverable d2.3.1). Technical report, Luiss Lab of European Economics. Methodological documentation; capital compensation uses an internal (endogenous) rate of return.

Canzoneri, M. B., Cumby, R. E., and Diba, B. T. (1999). Relative labor productivity and the real exchange rate in the long run: Evidence for a panel of OECD countries. *Journal of International Economics*, 47(2):245–266.

Cardi, O. and Restout, R. (2015). Imperfect mobility of labor across sectors: A reappraisal of the balassa–samuelson effect. *Journal of International Economics*, 97(2):249–265.

Chahrour, R., Cormun, V., De Leo, P., Guerrón-Quintana, P. A., and Valchev, R. (2024). Exchange rate disconnect revisited. Technical report, National Bureau of Economic Research.

De Gregorio, J., Giovannini, A., and Wolf, H. C. (1994). International evidence on tradables and nontradables inflation. *European Economic Review*, 38(6):1225–1244.

Devereux, M. B., Fujiwara, I., and Granados, C. (2025). Productivity and wedges: Economic convergence and the real exchange rate. Technical report, National Bureau of Economic Research.

Engel, C. (1999). Accounting for U.S. real exchange rate changes. *Journal of Political Economy*, 107(3):507–538.

Feenstra, R. C., Inklaar, R., and Timmer, M. P. (2015). The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–3182.

Froot, K. A. and Rogoff, K. (1995). Perspectives on PPP and long-run real exchange rates. In *Handbook of International Economics*, volume 3, pages 1647–1688. Elsevier.

Gornemann, N., Guerrón Quintana, P. A., and Saffie, F. (2025). Real exchange rates and endogenous productivity. *American Economic Journal: Macroeconomics*, 17(4):204–261.

Head, K. and Ries, J. (2001). Increasing returns versus national product differentiation as an explanation for the pattern of US–canada trade. *American Economic Review*, 91(4):858–876.

Hsieh, D. A. (1982). The determination of the real exchange rate: The productivity approach. *Journal of International Economics*, 12(3-4):355–362.

Irwin, D. A. and Obstfeld, M. (2024). Understanding korea's long-run real exchange rate behavior. Technical report, National Bureau of Economic Research. NBER Working Paper No. 32769 / PIIE Working Paper 24-17.

Itskhoki, O. (2021). The story of the real exchange rate. *Annual Review of Economics*, 13:423–455.

Klau, M. and Fung, S. S. (2006). The new BIS effective exchange rate indices. *BIS Quarterly Review*, pages 51–65.

Lee, J. and Tang, M.-K. (2007). Does productivity growth appreciate the real exchange rate? *Review of International Economics*, 15(1):164–187.

Lothian, J. R. and Taylor, M. P. (2008). Real exchange rates over the past two centuries: How important is the harrod–balassa–samuelson effect? *The Economic Journal*, 118(532):1742–1763.

Officer, L. H. (1976). The productivity bias in purchasing power parity: An econometric investigation. *IMF Staff Papers*, 23(3):545.

Ravn, M. O. and Uhlig, H. (2002). On adjusting the Hodrick–Prescott filter for the frequency of observations. *Review of Economics and Statistics*, 84(2):371–376.

Takahashi, Y. and Takayama, N. (2025a). Global technology stagnation. Working Paper.

Takahashi, Y. and Takayama, N. (2025b). Online appendix for "Global technology stagnation". Working Paper.

Tica, J. and Družić, I. (2006). The harrod–balassa–samuelson effect: A survey of empirical evidence. Technical Report 07, EFZG Working Paper Series.

A Additional Data Details

A.1 Coverage of the Data

Table A1: Coverage in Empirical Analyses

Country	Start year	End year	Data source
AUT	1970	2021	EU KLEMS
BEL	1970	2021	EU KLEMS
BGR	1995	2021	EU KLEMS
CHN	1981	2017	CIP
CYP	1995	2021	EU KLEMS
CZE	1995	2021	EU KLEMS
DEU	1970	2021	EU KLEMS
DNK	1970	2021	EU KLEMS
ESP	1970	2021	EU KLEMS
EST	1995	2021	EU KLEMS
FIN	1970	2021	EU KLEMS
FRA	1970	2021	EU KLEMS
GBR	1970	2021	EU KLEMS
GRC	1970	2021	EU KLEMS
HRV	1995	2021	EU KLEMS
HUN	1992	2021	EU KLEMS
IND	1981	2009	Asia KLEMS
IRL	1970	2021	EU KLEMS
ITA	1970	2021	EU KLEMS
JPN	1973	2021	EU KLEMS
KOR	1980	2012	Asia KLEMS
LTU	1995	2021	EU KLEMS
LUX	1970	2021	EU KLEMS
LVA	1995	2021	EU KLEMS
MLT	2000	2021	EU KLEMS
NLD	1970	2021	EU KLEMS
POL	1995	2021	EU KLEMS
PRT	1970	2021	EU KLEMS
ROU	1995	2021	EU KLEMS
SVK	1995	2021	EU KLEMS
SVN	1995	2021	EU KLEMS
SWE	1970	2021	EU KLEMS
TWN	1980	2009	Asia KLEMS

Table A2: Single-reference time-FE regressions: low-frequency variation

	(1)	(2)	(3)	(4)	(5)
Log Rel. ALP	0.11	0.22	0.35	0.36	0.31
	(0.13)	(0.16)	(0.13)	(0.18)	(0.21)
Observations	133	82	116	72	44
Sample Countries	All	All	Balanced	Balanced	Balanced
Num of Countries	33	33	28	28	28
Sample Years	10 Years	20 Years	10 Years	20 Years	40 Years
Num of Years	5	3	5	3	2
Country & Year FE	✓	✓	✓	✓	✓

Notes: This table reports panel regressions based on equation (9) with country fixed effects and time fixed effects. Variables are first HP-filtered to extract trends and then aggregated. Columns report 10-year averages, 20-year averages, and long differences where indicated. Balanced-panel variants repeat the same constructions holding country composition fixed. The dependent variable is $\ln RER_{i,t} - \ln RER_{U,t}$ and the regressor is $\ln(A_{i,T,t}/A_{i,NT,t}) - \ln(A_{U,T,t}/A_{U,NT,t})$. Standard errors are panel-corrected to allow correlation across countries within a year using the Beck–Katz method with period correlation.

B More Robustness

Low-frequency variation. We examine whether the result reflects very slow-moving comovement. We average the trend components into non-overlapping 10-year and 20-year intervals and also estimate long differences over the maximum span where applicable. We repeat each exercise on the full set and on a balanced set. Table A2 shows that the estimated elasticity remains positive and of similar magnitude even when attention is restricted to long-horizon movements, confirming that the finding is not an artifact of higher-frequency noise.

No time-series filtering. We check that the time-FE result is not an artifact of trend extraction. We re-estimate the specification on unfiltered annual series while keeping the same country and time fixed effects and the same samples. Table A3 shows that the coefficient remains positive and close to the baseline magnitudes, though standard errors widen as expected when short-run noise is present.

Table A3: Single-reference time-FE regressions: unfiltered annual series

	(1)	(2)	(3)	(4)
Log Rel. ALP	0.08	0.25	0.25	0.27
	(0.04)	(0.06)	(0.07)	(0.07)
Observations	1,307	849	778	756
Sample Countries	All	All	Adv	Balanced
Num of Countries	33	33	29	28
Sample Years	1970-2021	1995-2021	1995-2021	1995-2021
Num of Years	52	27	27	27
Country & Year FE	✓	✓	✓	✓

Notes: This table reports panel regressions based on equation (9) estimated on unfiltered annual data. All specifications include country fixed effects and time fixed effects. Column (1) covers 1970–2021. Column (2) restricts to 1995–2021. Column (3) restricts to advanced economies. Column (4) is a balanced panel for 1995–2021. The dependent variable is $\ln RER_{i,t} - \ln RER_{U,t}$. The Balassa–Samuelson regressor is $\ln(A_{i,T,t}/A_{i,NT,t}) - \ln(A_{U,T,t}/A_{U,NT,t})$. Standard errors are panel-corrected to allow correlation across countries within a year using the Beck–Katz method with period correlation.

C Hat Algebra

Hats denote proportional changes relative to the baseline, e.g. $\hat{x} = x'/x$.

We allow shocks to productivity, trade costs, and labor supply, so the objects $\widehat{A}_{i,s}$, $\widehat{\tau}_{i,j,s}$, and $\hat{L}_i$ can all differ from one. Baseline world income is normalized to one,

$$\sum_{\ell=1}^{N} w_\ell L_\ell = 1,$$

and we keep each country's trade surplus β_i fixed across counterfactuals, with

$$\sum_{\ell=1}^{N} \beta_\ell = 0,$$

where $\beta_i > 0$ means that country i runs a trade *surplus* in the baseline.

Trade balance and labor-market clearing. With a labor-supply shock, post-shock income of country i is $\hat{w}_i \hat{L}_i w_i L_i$. Exporter-side market clearing is therefore

$$\hat{w}_i \hat{L}_i w_i L_i + \beta_i = \sum_{j=1}^{N} \sum_{s=1}^{S} \hat{\pi}_{i,j,s}\, \pi_{i,j,s}\, \alpha_{j,s}\, \hat{w}_j \hat{L}_j w_j L_j, \qquad i = 1, \ldots, N. \tag{12}$$

The left-hand side is factor income of i after the wage and labor shocks, plus its fixed trade surplus. The right-hand side is the revenue it earns from all destinations and sectors, using baseline expenditure shares and their hats. Summing (12) over i and using $\sum_i \beta_i = 0$ implies

$$\sum_{i=1}^{N} \hat{w}_i \hat{L}_i w_i L_i = 1.$$

Thus, post-shock world income remains normalized to one.

Unit costs.

$$\hat{c}_{i,s} = \frac{\hat{w}_i}{\widehat{A}_{i,s}}.$$

Delivered prices.

$$\hat{P}_{i,j,s} = \frac{\hat{w}_i\, \widehat{\tau}_{i,j,s}}{\widehat{A}_{i,s}}.$$

Sectoral price index.

$$(\hat{P}_{j,s})^{1-\theta} = \sum_{i=1}^{N} \pi_{i,j,s} \, (\hat{P}_{i,j,s})^{1-\theta}.$$

Bilateral expenditure shares.

$$\hat{\pi}_{i,j,s} = \left(\frac{\hat{P}_{i,j,s}}{\hat{P}_{j,s}} \right)^{1-\theta} = \left(\frac{\hat{w}_i \, \widehat{\tau}_{i,j,s}}{\widehat{A}_{i,s} \, \hat{P}_{j,s}} \right)^{1-\theta}.$$

Aggregate CPI.

$$\hat{P}_j = \prod_{s=1}^{S} (\hat{P}_{j,s})^{\alpha_{j,s}}.$$

D Model with Input–Output Linkages

This appendix extends the baseline Armington model with Cobb-Douglas aggregation across sectors and CES aggregation within sectors to allow for intermediate-input use. The structure, notation, and timing follow the main text.

D.1 Countries and Sectors

There are N countries indexed by $i, j = 1, \ldots, N$ and S sectors indexed by $s, r = 1, \ldots, S$. A good produced in sector s of country i can be (i) absorbed as final demand in any country or (ii) used as an intermediate input by any sector in any country.

D.2 Preferences

Households in country j have Cobb–Douglas preferences over sectoral CES composites,

$$U_j = \prod_{s=1}^{S} \left(\frac{C_{j,s}}{\alpha_{j,s}} \right)^{\alpha_{j,s}}, \qquad \sum_{s=1}^{S} \alpha_{j,s} = 1,$$

where $C_{j,s}$ is final consumption of sector s in country j and $\alpha_{j,s}$ are expenditure shares. Let $P_{j,s}$ be the final-use price index for sector s in j and let P_j be the CPI. Optimal allocation implies

$$P_{j,s} C_{j,s} = \alpha_{j,s} P_j C_j, \qquad P_j = \prod_{s=1}^{S} P_{j,s}^{\alpha_{j,s}}.$$

World income is

$$Y \equiv \sum_{\ell=1}^{N} w_\ell L_\ell.$$

Trade imbalances are exogenous. Let β_j denote country j's trade surplus as a fraction of world income, with

$$\sum_{j=1}^{N} \beta_j = 0.$$

Final absorption in j is

$$P_j C_j = w_j L_j - \beta_j Y,$$

so $\beta_j > 0$ means that country j spends less than its income.

D.3 Technology and Costs

Production uses labor and sectoral intermediate-input composites. For country i, sector s, the unit cost is

$$c_{i,s} = \frac{v_{i,s}\, w_i^{\gamma_{i,s}} \prod_{r=1}^{S} \left(P_{i,r}^X\right)^{\eta_{i,r,s}(1-\gamma_{i,s})}}{A_{i,s}},$$

where:

- $A_{i,s}$ is productivity,
- $\gamma_{i,s} \in (0,1]$ is the labor cost share,
- $\eta_{i,r,s} \geq 0$ is the cost share of input r in sector s of country i, with $\sum_{r=1}^{S} \eta_{i,r,s} = 1$,
- $P_{i,r}^X$ is the price index of intermediate inputs from sector r used in country i,
- $v_{i,s}$ is a cost shifter that is fixed across counterfactuals.

Thus, a fraction $\gamma_{i,s}$ of costs is labor and a fraction $(1-\gamma_{i,s})$ is intermediates, allocated across input sectors according to $\eta_{i,r,s}$.

D.4 Trade Costs

Goods used for *final* demand face iceberg trade costs $\tau_{i,j,s}^F \geq 1$,

$$P_{i,j,s}^F = c_{i,s}\tau_{i,j,s}^F = \frac{v_{i,s}\, w_i^{\gamma_{i,s}} \prod_r (P_{i,r}^X)^{\eta_{i,r,s}(1-\gamma_{i,s})}}{A_{i,s}}\, \tau_{i,j,s}^F.$$

Goods used for *intermediate* demand face (possibly different) iceberg trade costs $\tau_{i,j,s}^X \geq 1$,

$$P_{i,j,s}^X = c_{i,s}\tau_{i,j,s}^X.$$

D.5 Final and Intermediate CES Aggregators

Within each sector s, final demand in country j is a CES composite of varieties produced in different countries,

$$C_{j,s} = \left(\sum_{i=1}^{N} \mu_{i,j,s}^{1/\theta}\, C_{i,j,s}^{(\theta-1)/\theta}\right)^{\theta/(\theta-1)}, \qquad \theta > 1,$$

with taste shifters $\mu_{i,j,s} > 0$. This implies the sectoral price index

$$P_{j,s} = \left(\sum_{i=1}^{N} \mu_{i,j,s} \left(P^F_{i,j,s} \right)^{1-\theta} \right)^{1/(1-\theta)}.$$

The corresponding final-expenditure share is

$$\pi^F_{i,j,s} \equiv \frac{P^F_{i,j,s} C_{i,j,s}}{\sum_{\ell=1}^{N} P^F_{\ell,j,s} C_{\ell,j,s}} = \frac{\mu_{i,j,s} \left(P^F_{i,j,s} \right)^{1-\theta}}{\sum_{\ell=1}^{N} \mu_{\ell,j,s} \left(P^F_{\ell,j,s} \right)^{1-\theta}}.$$

For intermediate use, we allow for a distinct set of taste weights $\mu^X_{i,j,s}$. The intermediate-input price index for sector s in country j is

$$P^X_{j,s} = \left(\sum_{i=1}^{N} \mu^X_{i,j,s} \left(P^X_{i,j,s} \right)^{1-\theta} \right)^{1/(1-\theta)},$$

and the corresponding intermediate-expenditure share is

$$\pi^X_{i,j,s} = \frac{\mu^X_{i,j,s} \left(P^X_{i,j,s} \right)^{1-\theta}}{\sum_{\ell=1}^{N} \mu^X_{\ell,j,s} \left(P^X_{\ell,j,s} \right)^{1-\theta}}.$$

D.6 Gross Output

Let $Y_{i,s}$ denote gross output (revenue) of country i, sector s. Demand for $Y_{i,s}$ has two components.

Final demand. Country j spends $\alpha_{j,s}(w_j L_j - \beta_j Y)$ on sector s; exporter (i,s) gets the share $\pi^F_{i,j,s}$. Final-demand revenue of (i,s) is

$$\sum_{j=1}^{N} \pi^F_{i,j,s}\, \alpha_{j,s}\, (w_j L_j - \beta_j Y).$$

Intermediate demand. Country j, user sector r, spends a fraction $(1 - \gamma_{j,r})$ of its gross output $Y_{j,r}$ on intermediates; of that, the share $\eta_{j,s,r}$ is spent on inputs from sector s; and of that, exporter (i,s) receives the share $\pi^X_{i,j,s}$. Intermediate-demand revenue of (i,s) is

$$\sum_{j=1}^{N}\sum_{r=1}^{S}\pi^X_{i,j,s}\,(1-\gamma_{j,r})\,\eta_{j,s,r}\,Y_{j,r}.$$

Combining the two parts, gross output in country i and sector s satisfies

$$Y_{i,s}=\sum_{j=1}^{N}\pi^F_{i,j,s}\,\alpha_{j,s}\,(w_jL_j-\beta_jY)+\sum_{j=1}^{N}\sum_{r=1}^{S}\pi^X_{i,j,s}\,(1-\gamma_{j,r})\,\eta_{j,s,r}\,Y_{j,r}. \tag{13}$$

D.7 Labor Market Clearing

Labor income in country i is the labor-cost share across its sectors,

$$w_iL_i=\sum_{s=1}^{S}\gamma_{i,s}Y_{i,s}. \tag{14}$$

Summing (13) over i and s, and using

$$\sum_{i=1}^{N}\pi^F_{i,j,s}=1,\qquad \sum_{i=1}^{N}\pi^X_{i,j,s}=1,\qquad \sum_{s=1}^{S}\alpha_{j,s}=1,\qquad \sum_{s=1}^{S}\eta_{j,s,r}=1,\qquad \sum_{j=1}^{N}\beta_j=0,$$

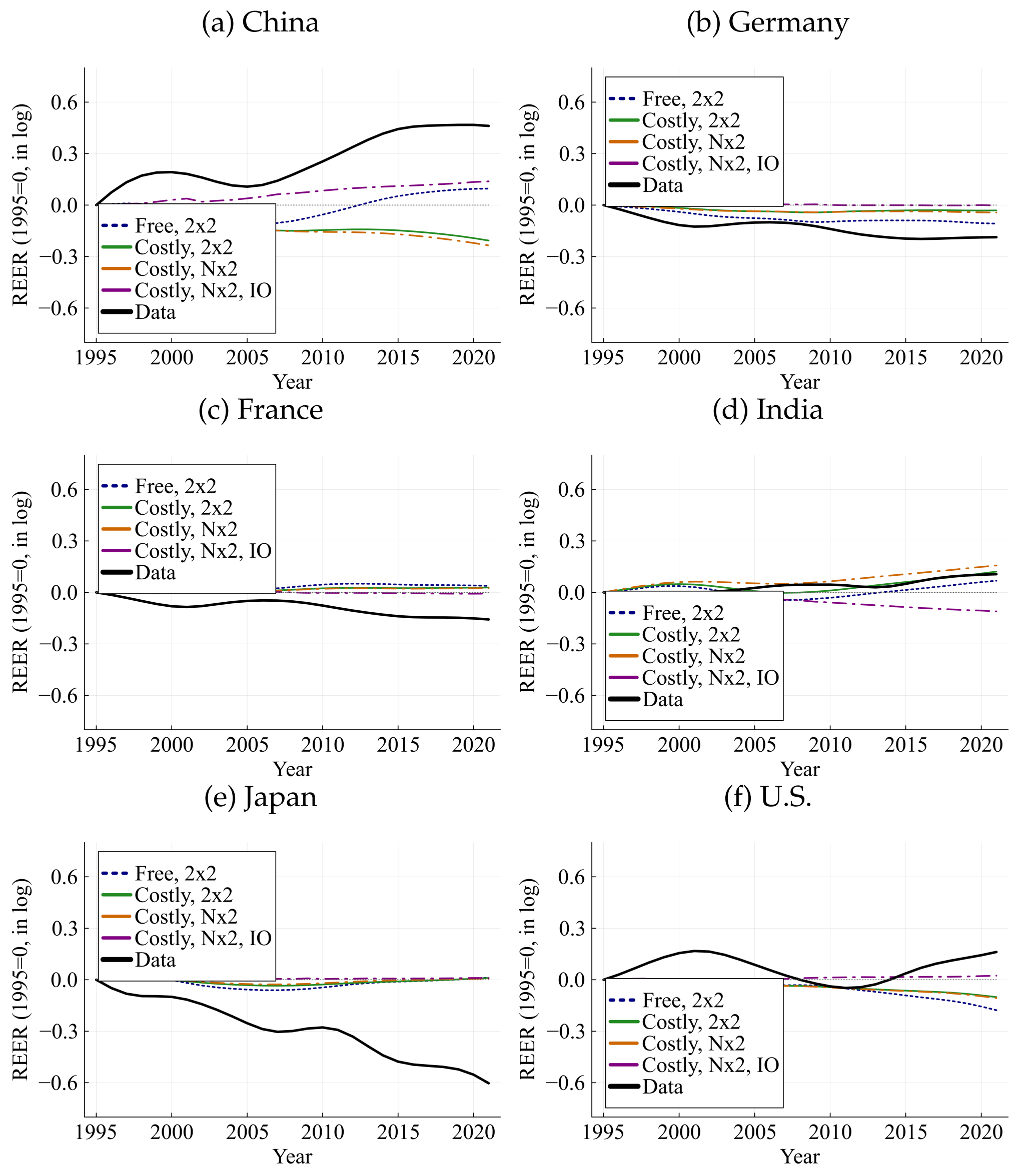
we obtain

$$\begin{aligned}\sum_{i=1}^{N}\sum_{s=1}^{S}Y_{i,s}&=\sum_{j=1}^{N}\sum_{s=1}^{S}\alpha_{j,s}(w_jL_j-\beta_jY)+\sum_{j=1}^{N}\sum_{r=1}^{S}(1-\gamma_{j,r})Y_{j,r}\\&=\sum_{j=1}^{N}(w_jL_j-\beta_jY)+\sum_{j=1}^{N}\sum_{r=1}^{S}(1-\gamma_{j,r})Y_{j,r}\\&=\sum_{j=1}^{N}w_jL_j+\sum_{j=1}^{N}\sum_{r=1}^{S}(1-\gamma_{j,r})Y_{j,r}.\end{aligned}$$

The first term is world value added; the second term is total intermediate use. Hence, with input–output linkages.

E Further Quantitative Results

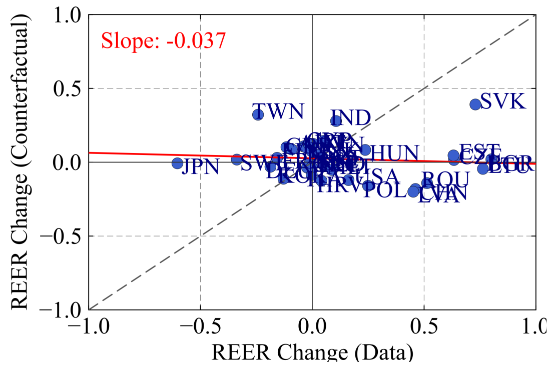
Figure A1: REER Fits: Productivity Shock



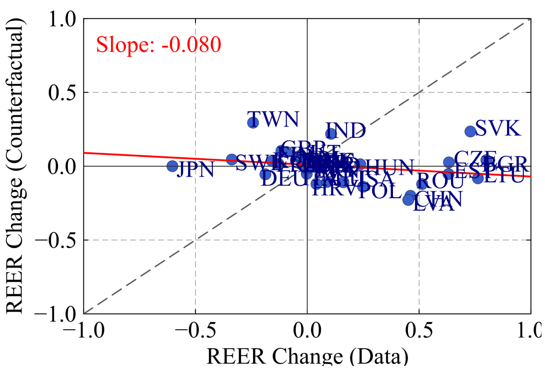
Notes: This figure compares the observed and model-generated REER paths across countries.

Figure A2: REER Long-Difference Fit vs. Data, 1995–2021: Adding Trade-Cost Shocks

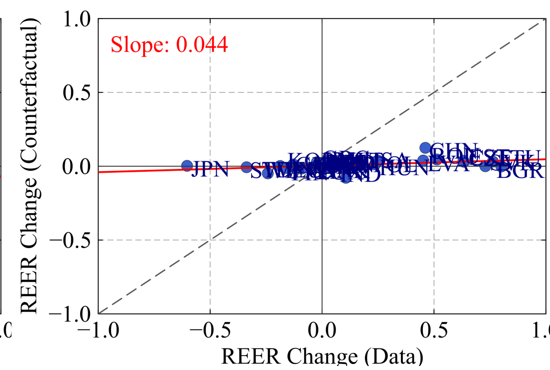
(a) 2×2 Costly Trade



(b) $N \times 2$ Costly Trade



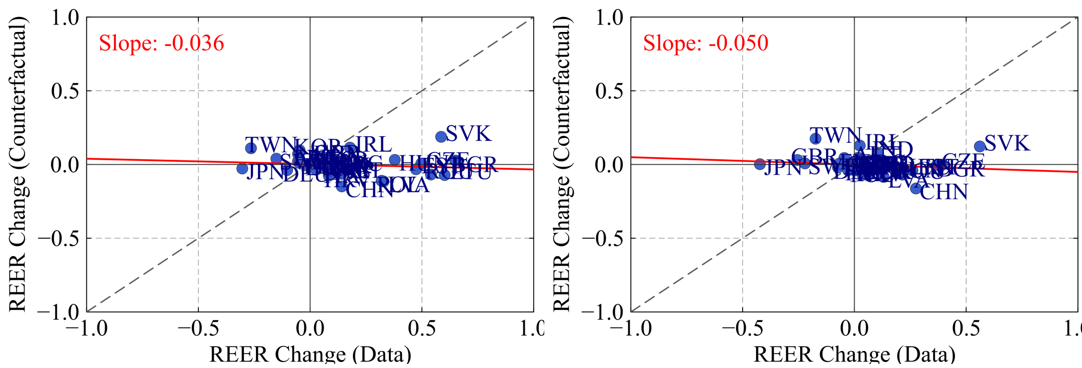
(c) $N \times 2$ Costly Trade with IO Linkages



Notes: Each dot represents one country. The horizontal axis shows the HP-filtered change in log real effective exchange rate (REER) between 1995 and 2021 from the BIS data. The vertical axis shows the corresponding model-implied change when both bilateral iceberg trade costs $\{\tau_{i,j,s,t}\}$ and sectoral productivities $\{A_{i,s,t}\}$ evolve along their estimated low-frequency paths. Panels differ by model environment: (a) 2×2 with costly trade, (b) $N \times 2$ with costly trade, and (c) $N \times 2$ with costly trade and input–output linkages. A 45-degree line would indicate a perfect quantitative fit. The fitted variation remains a small fraction of that in the data, confirming that allowing time-varying trade costs does not materially improve the model's explanatory power.

Figure A3: REER Long-Difference Fit vs. Data, Alternative Time Windows

(a) 1995–2007 (b) 2000–2019



Notes: Each dot represents one country. The horizontal axis shows the HP-filtered change in log REER from the BIS data over the indicated period. The vertical axis shows the model-implied change in the $N \times 2$ costly-trade environment when sectoral productivities $\{A_{i,s,t}\}$ evolve along their estimated low-frequency paths. Country-level REER is constructed from bilateral series using fixed 1995 trade weights following the Bank for International Settlements weighting scheme and the multilateral index construction in Klau and Fung (2006). Results are similar across windows, confirming that the weak quantitative fit does not depend on the choice of time horizon.