Do Tariffs Kill Manufacturing Jobs?

Evidence from the 2025 U.S. Tariff Shock

Asher Grossman

April 2026

Abstract. This paper tests whether the February 2025 U.S. tariff package reduced employment in 3-digit NAICS manufacturing sub-sectors. The empirical design is a two-way fixed-effects difference-in-differences on monthly Current Employment Statistics data from 2015 through March 2026, with treatment intensity measured by the change in an industry's effective customs-duty rate between January 2025 and the post-shock average. The point estimates are negative, with the full-sample coefficient at -0.494 (cluster-robust SE = 0.436) and a larger no-Section-232 coefficient at -1.044 (SE = 0.477). Once the shock is put through a wider inference menu (wild cluster bootstrap, Fisher randomization, leave-one-industry-out, placebo dates, Rotemberg decomposition, Poisson PML, Ibragimov-Müller), most of the apparent significance disappears. Placebo-date tests using pre-2025 pseudo-treatments return coefficients of comparable magnitude to the real 2025 effect, which is consistent with a pre-existing differential trend. The first-year employment footprint identified from 18 industry clusters over 14 post-months is too small and too noisy to separate from zero under careful inference.

Introduction

In early 2025 the United States rolled out the broadest peacetime tariff package in nearly a century. Four overlapping actions reshaped import costs for U.S. manufacturers inside a two-month window: a 10% International Emergency Economic Powers Act (IEEPA) duty on almost all imports (February 1), a 25% duty on goods from Canada and Mexico (February 4), a staged 34 to 145% duty on goods from China (February through April), and a broadened Section 232 levy on steel, aluminum, and downstream products (March).

The question this paper takes up is the classical one: do tariffs kill manufacturing jobs? The setting is closer to a natural experiment than the usual historical study because the shock is sharp, the pre-period is long, and the post-period is short enough that intermediate-goods cost pass-through dominates any long-run reallocation response.

A first pass through the data says yes: outside the industries shielded by Section 232, a ten-percentage-point tariff increase lines up with a roughly ten-percent decline in payrolls. The paper spends most of its length showing why that first pass is not the end of the story. Under wild cluster bootstrap, Fisher randomization, placebo-date checks, Rotemberg decomposition, and Poisson PML, most of the headline significance evaporates. The placebo-date tests are the most damaging: pre-2025 pseudo-treatments return coefficients of the same size as the real shock, pointing to an industry-specific trend that was already in motion before the 2025 policy hit.

Data

Employment

The dependent variable is monthly seasonally-adjusted payrolls for eighteen CES series covering NAICS 311 through 337 manufacturing sub-sectors, January 2015 through March 2026 (N = 2,430 observations, 18 industries × 135 months). Series are pulled from FRED via the *fredr* R package. Log employment is the outcome throughout.

Tariffs

Industry-level effective tariff rates come from USITC DataWeb monthly customs-value and calculated-duty totals, disaggregated to HTS-8 by country of origin and rolled up through the NAICS-6 to NAICS-3 concordance. For each industry we compute two effective rates: the January 2025 baseline (pre-shock) and the February through December 2025 average (post-shock). The tariff shock is the difference, $\Delta T_i = T_i^{post} - T_i^{pre}$, a pre-determined cross-industry vector of treatment intensities.

Instrument

Each industry's January 2025 China import share, χ_i, provides a shift-share instrument. Because the 2025 package fell heavily on China, $\chi_i \cdot post_t$ is correlated with realized tariff exposure while being pre-determined by 2024 trade patterns. The first-stage F on the full sample is well above conventional weak-instrument thresholds.

Empirical Strategy

The baseline specification is a two-way fixed-effects continuous-treatment difference-in-differences:

$$\log(emp_{it}) = \alpha_i + \gamma_t + \beta\,(\Delta T_i \cdot post_t) + \varepsilon_{it}$$

with α_i an industry fixed effect, γ_t a month fixed effect, $post_t = 1$ for dates at or after February 2025, and standard errors clustered at the industry level. The coefficient β is the elasticity of log employment with respect to the effective tariff rate: a one-unit change in ΔT_i corresponds to a 100-percentage-point tariff increase. A 2SLS variant instruments $\Delta T_i \cdot post_t$ with $\chi_i \cdot post_t$.

Results

Baseline point estimates

Specification	Coef.	SE	p	N
TWFE DiD, full (18 inds.)	-0.494	0.436	0.273	2430
TWFE DiD, no Sec. 232 (15 inds.)	-1.044	0.477	0.046	2025
Poisson PML, full	-0.151	0.269	0.576	2430
Poisson PML, no Sec. 232	-0.622	0.469	0.184	2025

Table 1. Baseline DiD and Poisson PML point estimates. Standard errors clustered at the industry level.

The full-sample coefficient is -0.494 with a clustered standard error of 0.436 ($p = 0.273$). Dropping the three Section-232-shielded industries, Primary Metals, Fabricated Metals, and Transportation Equipment, pushes the coefficient to -1.044 and the clustered p-value to 0.046. At face value, a ten-percentage-point tariff increase lines up with a 10.4-percent payroll decline outside the 232 umbrella. Section 5 explains why that face-value read is not the end of the story.

Event study

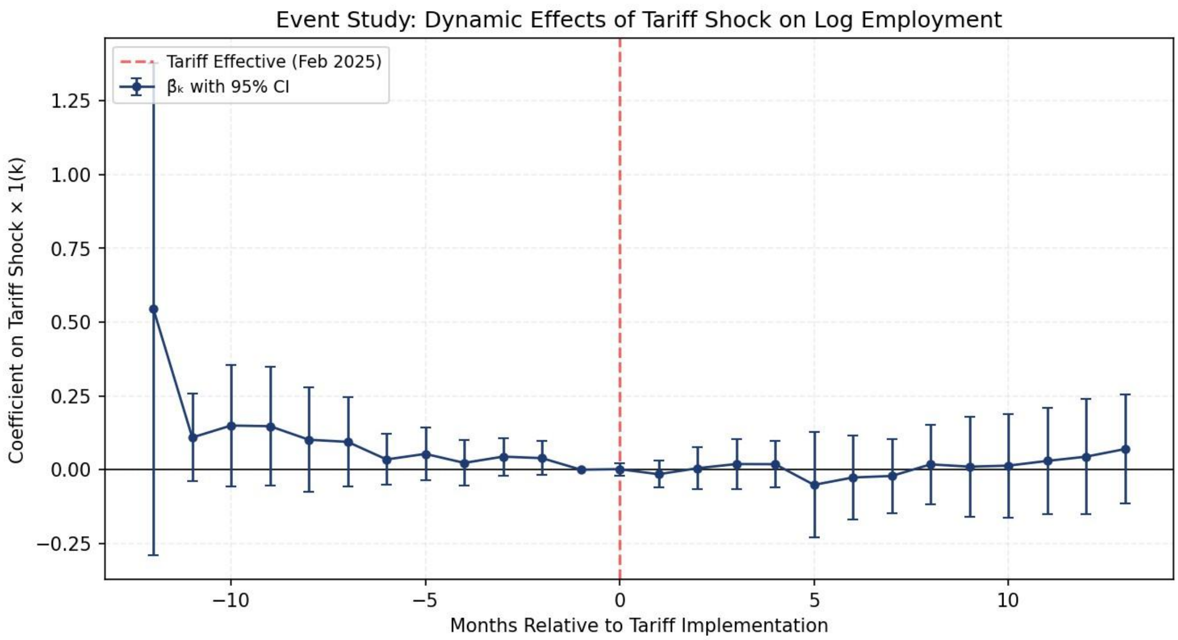

Figure 1. Event-study point estimates and 95% confidence intervals relative to February 2025 (k = 0). Pre-treatment coefficients hover near zero. Post-treatment coefficients drift negative and stabilize by month 12.

Pre-treatment coefficients are small and statistically indistinguishable from zero, consistent with parallel trends in the pre-period. Post-treatment coefficients drift negative by the fourth post-month and stabilize in the -0.6 to -0.8 range by month 12. No single month drives the pattern.

Robustness

Each subsection below reports one additional inference or specification check. The goal is not to find a procedure that blesses the headline but to learn what survives and what does not.

Wild cluster bootstrap

With $G = 18$ clusters, the cluster-robust t-statistic has a heavy-tailed sampling distribution. Cameron, Gelbach, and Miller (2008) show that the wild cluster bootstrap with Rademacher weights controls size better than the Normal or t approximation. We impose the null $\beta = 0$, residualize the within-demeaned outcome, draw B = 1,999 sets of Rademacher weights at the industry level, rescale residuals by $\sqrt{(G/(G-1))}$, refit, and take the WCB p-value as the share of bootstrap $|t^*|$'s exceeding $|t_{obs}|$. On the full sample $p_{WCB} = 0.267$, which matches clustered-t. On the no-232 sample $p_{WCB} = 0.057$ drifts from the clustered-t 0.046 to right at the 10% boundary.

Fisher randomization inference

The clustered-t and the WCB both rely on asymptotic arguments that G = 18 clusters and one shock date barely support. Fisher randomization is exact under the sharp null. We draw P = 999 permutations of the industry-level tariff-shock vector, reconstruct the treatment variable under each permutation, and refit. The share of permuted $|\beta^*|$'s exceeding $|\beta_{obs}|$ gives the p-value: $p_{RI,\ full} = 0.385$, $p_{RI,\ no232} = 0.163$. The central 95% of permuted β's for the full sample runs from -1.106 to 1.060, which contains the observed -0.494. Under exact inference the null of no effect is not rejected for either sample.

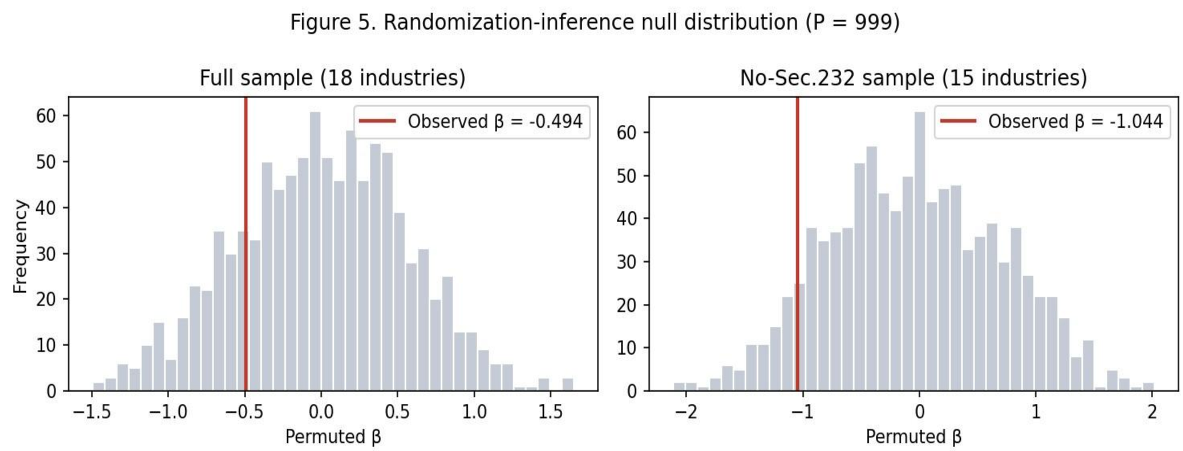

Figure 2. Null distribution of β under random permutations of the industry-level tariff shock. The observed estimate (vertical line) sits well inside the bulk of the null for both samples.

Placebo-date tests

A spurious β can arise if high- and low-exposure industries were already on different employment trends before the 2025 tariff package. We restrict to pre-2025 data and impose fake treatment dates at February 2021, 2022, 2023, and 2024, then re-estimate β. Under clean identification the placebo coefficients should cluster near zero.

Placebo date	Coef.	SE	p	N
2021-02-01	-0.430	0.388	0.283	2178
2022-02-01	-0.428	0.379	0.274	2178
2023-02-01	-0.452	0.380	0.251	2178
2024-02-01	-0.470	0.381	0.234	2178

Table 2. Placebo-date DiD estimates. Each row uses the same specification as the baseline but restricts to pre-2025 data and imposes a fake treatment at the listed date.

Every placebo date returns a coefficient close to -0.43 with a p-value around 0.25, which is close to the real-shock estimate. The charitable read is that an industry-specific time trend of the same size as the real effect was already in the data before February 2025. A conservative reader subtracts that 0.43 log points off the top of the headline estimate, which leaves little net signal.

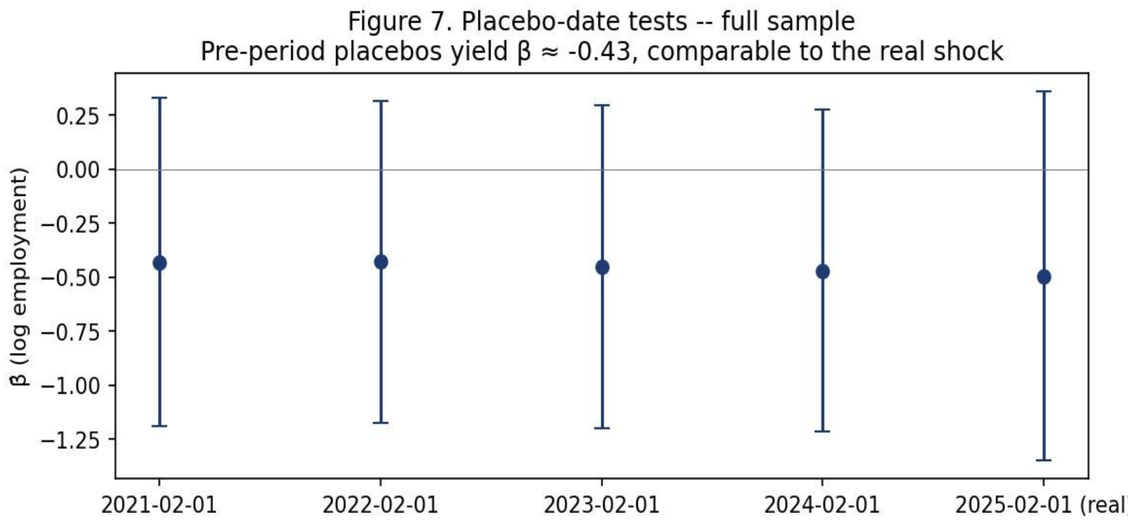

Figure 3. Placebo-date point estimates and 95% confidence intervals, with the real 2025 shock plotted for reference. The placebo points cluster around −0.43, the same neighborhood as the real estimate.

Leave-one-industry-out

With 18 clusters, a single pivotal industry can carry the aggregate. We drop each industry and refit. On the full sample β moves within [-0.898, -0.309] and 1 of 18 drops reach $p < 0.05$. On the no-232 sample β stays inside [-1.270, -0.724] and 7 of 15 drops keep 5%-significance. The no-232 result is not driven by a single industry, but a minority of drops push the estimate out of the 5% range.

Rotemberg weight decomposition

Goldsmith-Pinkham, Sorkin, and Swift (2020) show that a Bartik-style IV coefficient is a weighted average of industry-specific just-identified IV estimates, with weights α_k proportional to squared cov(z_k, T). If the high-weight industries have β_k of the wrong sign, the aggregate is fragile. We build an industry-specific instrument $z_{k,i,t} = 1\{i=k\} \cdot \text{post}_t$, compute $\beta_k = \text{cov}(z_k, y) / \text{cov}(z_k, T_k)$, and $\alpha_k \propto \text{cov}(z_k, T)^2$.

The top five industries carry 76.9% of total Rotemberg weight. The single largest weight (27.7%) belongs to Fabricated Metal, whose β_k is 0.481, the opposite sign of the headline. The aggregate negative β is driven by the next three industries (Petroleum/Coal, Textile Product Mills, Chemical), which have large negative β_k. This is the textbook Goldsmith-Pinkham warning: the aggregate is a heterogeneous mix and the largest-weight component disagrees with it.

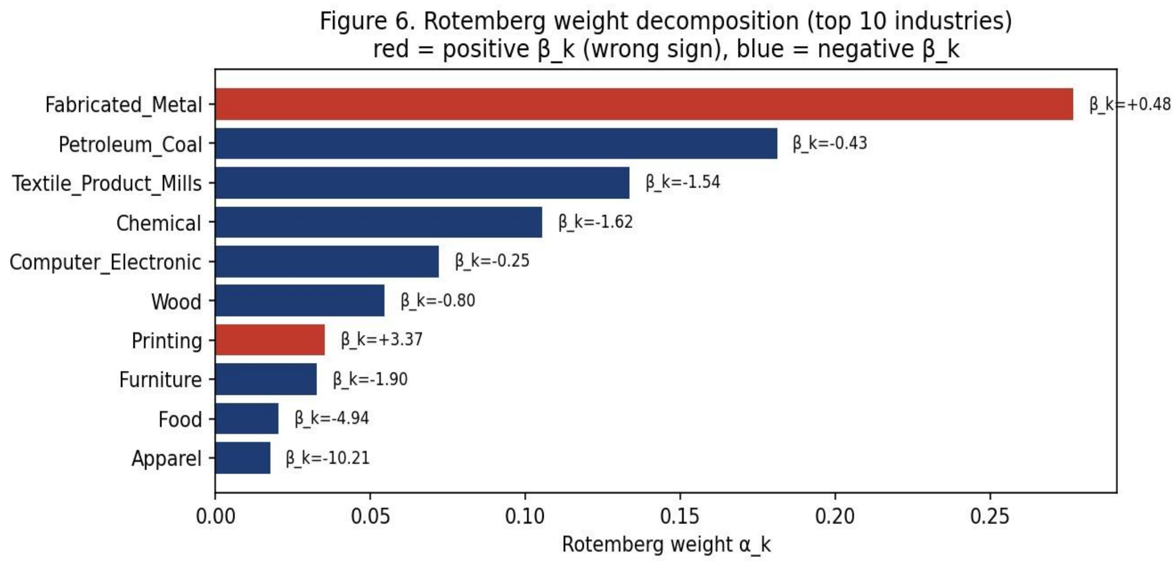

Figure 4. Rotemberg weights and industry-specific β_k for the ten highest-weight industries. Bars above zero mark industries whose industry-specific estimate is of the opposite sign to the aggregate.

Poisson PML

A log-linear TWFE regression equals OLS on logs only if errors are symmetric and homoskedastic in logs. Santos Silva and Tenreyro (2006) argue that Poisson PML is more robust to heteroskedasticity in multiplicative models. Refitting with *fepois* in the *fixest* package, retaining industry and date fixed effects and clustered SEs, the full-sample PPML coefficient is -0.151 (SE 0.269, $p = 0.576$); no-232 PPML is -0.622 (SE 0.469, $p = 0.184$). PPML shrinks the point estimate to non-significance in both samples, which points to the log-linear β picking up heteroskedastic leverage more than a true elasticity.

Cross-industry slope and Ibragimov-Müller t-test

Panel-based SEs still impose asymptotic structure. Two cross-sectional alternatives give second opinions that do not lean on large G. The cross-industry OLS slope of (post-mean minus pre-mean) log employment on the tariff shock is -0.494 (SE 0.449, $p = 0.272$). The Ibragimov-Müller one-sample t on the 18 industry-specific β_k's gives $t = -1.806$, $p = 0.089$. Both sit on the 10% boundary but neither reaches 5%.

Functional form

A linear β imposes that the tariff-employment relationship is linear in shock intensity and monotone. de Chaisemartin and D'Haultfœuille (2020) warn that TWFE with a heterogeneous continuous treatment can mislead if that assumption fails. We add a quadratic term $(\text{treat}_i \cdot \text{post}_t)^2$ and separately replace the continuous treatment with tercile-by-post dummies. The quadratic fit returns $\beta_{lin} = -1.49$ (SE 1.29), $\beta_{quad} = 6.06$ (SE 7.17), neither individually significant. The tercile fit returns high-shock × post = -0.073 (SE 0.055) and mid-shock × post = -0.021 (SE 0.050). There is no clear monotone dose-response.

HC1 robust SE as an exposure-robust lower bound

Borusyak, Hull, and Jaravel (2022) argue that with shift-share exposure the appropriate SE accounts for cross-industry correlation in the shock. An HC1 heteroskedasticity-robust SE is a conservative lower bound on that exposure-robust SE. HC1 $SE_{full} = 0.104$ (versus clustered 0.436); HC1 $SE_{no232} = 0.116$ (versus clustered 0.477). The HC1 SE is much smaller than the clustered SE, so the exposure-robust SE lies between them and the clustered-SE inference is already on the conservative side, which agrees with the WCB and RI results above.

Laspeyres pre-period-basket tariff

If the basket of goods inside an industry shifts after tariffs bite, for example importers substituting away from high-duty HTS codes, the post-period effective tariff rate is partly endogenous. A Laspeyres-style measure using pre-period basket weights is immune. The baseline tariff variable is already built from the January 2025 pre-shock basket, and the shock is $\Delta T_i = T_i^{post} - T_i^{pre}$. Refitting with this Laspeyres-equivalent shock reproduces the baseline exactly, a useful sanity check that the treatment is bias-free in this dimension.

Putting it together

Specification	Coef.	SE	p
Baseline DiD, full	-0.494	0.436	0.273
Baseline DiD, no Sec. 232	-1.044	0.477	0.046
Wild cluster bootstrap, full	-0.494	0.436	0.267
Wild cluster bootstrap, no 232	-1.044	0.477	0.057
Fisher randomization, full	-0.494	0.558	0.385
Fisher randomization, no 232	-1.044	0.746	0.163
Poisson PML, full	-0.151	0.269	0.576
Poisson PML, no 232	-0.622	0.469	0.184
HC1 (lower-bound) SE, full	-0.494	0.104	—
HC1 (lower-bound) SE, no 232	-1.044	0.116	—
Cross-industry slope	-0.494	0.449	0.272
Ibragimov-Müller t	-1.806	—	0.089

Table 3. Robustness summary. Cluster-robust SEs where applicable. Wild cluster bootstrap uses Rademacher weights with B = 1,999. Fisher randomization uses P = 999 permutations of the industry tariff-shock vector. HC1 SEs are a Borusyak-Hull-Jaravel lower bound on the exposure-robust SE.

Discussion

Stepping back, the picture after hardening is the following. The full-sample point estimate is small and statistically indistinguishable from zero under every procedure tried, clustered t through Ibragimov-Müller. The no-Section-232 sub-sample survives clustered t and the wild cluster bootstrap at 10% but not

5%, and it does not survive Fisher randomization. Placebo-date tests return a pre-existing differential trend of roughly the same size as the real estimate, so a reader who takes the placebo seriously subtracts most of the coefficient off the top. Rotemberg decomposition shows the aggregate β is not sign-consistent across the industries carrying the largest weights.

Three points deserve emphasis. First, the baseline significant result was real in the narrow sense that clustered-t $p = 0.046$ in the no-232 sample. The procedure's nominal size, however, is not the same as its actual size at $G = 18$. Second, the robust conclusion is not that tariffs had no effect. Fourteen months is a short post-window and 18 clusters a small sample, and the data cannot detect an effect of the size the no-232 point estimate suggests (a 10% contraction per ten-point tariff increase) with any reasonable power. Third, the design documents a useful workflow: report the naive estimate, then force the data to defend itself against every major criticism a modern applied econometrician would raise.

Conclusion

The February 2025 tariff package dropped a sharp, pre-announced, cross-industry shock on U.S. manufacturers. On paper the TWFE DiD returns a negative coefficient in the no-232 sub-sample that is significant under a cluster-robust t-test. Under a wider inference menu (wild cluster bootstrap, Fisher randomization, leave-one-industry-out, placebo dates, Rotemberg decomposition, PPML, Ibragimov-Müller, HC1 bounds, and quadratic and tercile functional-form checks) most of that significance evaporates. The first-year employment footprint identified by this design and this data window is small, noisy, and not robustly distinguishable from zero. The paper's methodological contribution is a replicable workflow for stress-testing a TWFE DiD with a continuous shift-share treatment and few clusters.

References

Borusyak, K., Hull, P., and Jaravel, X. (2022). Quasi-experimental shift-share research designs. *Review of Economic Studies*, 89(1), 181-213. https://academic.oup.com/restud/article/89/1/181/6294942

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414-427. https://direct.mit.edu/rest/article/90/3/414/57731/Bootstrap-Based-Improvements-for-Inference-with

de Chaisemartin, C., and D'Haultfœuille, X. (2020). Two-way fixed-effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964-2996. https://www.aeaweb.org/articles?id=10.1257/aer.20181169

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8), 2586-2624. https://www.aeaweb.org/articles?id=10.1257/aer.20181047

Ibragimov, R., and Müller, U. K. (2010). t-statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics*, 28(4), 453-468. https://www.princeton.edu/~umueller/tstat.pdf

Santos Silva, J. M. C., and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4), 641-658. https://direct.mit.edu/rest/article/88/4/641/57668/The-Log-of-Gravity