Commentary and Critique **Research Commentary**
Vol. 9 (2026) · ISSN 3006-6220 journal.pe.ac/Commentary-and-Critique

An Eightfold Aggregation Trap in Mirror Trade Evidence on Third-Country Rerouting of Controlled Goods

Farhod Rakhimov[a], Nargiza Qodirova[a]

[a]Academy of Public Policy and Economics, London E1 0SG, United Kingdom

Corresponding author: Farhod Rakhimov, farhod.rakhimov@pe.ac.

Published 20 March, 2026.

DOI: https://doi.org/10.62432/APPE.CC.2026.9.34

Article webpage: https://journal.pe.ac/Commentary-and-Critique/paper/view/810782

Abstract

Third-country rerouting of export-controlled goods is a documented procurement pattern, and the temptation is to evidence it from mirror trade statistics — the gap between what an exporter reports sending and what the importing partner reports receiving. This article tests whether that gap can separate rerouting of controlled goods from ordinary statistical attribution differences, using the United Nations Comtrade public preview interface, the joint Common High Priority Items List as the operable definition of "controlled", and a same-chapter off-list contrast. It reports two results. The first is a measurement result that any user of the interface needs: the endpoint's response for a single reporter–partner–commodity–year–flow cell is **not an additive cross-tabulation**. Across 160 non-empty cells the naive sum of a cell's rows divided by that cell's maximum is approximately **8.00 within the recorded precision in 133 cells (83.1%)**; the total-bearing row is the one with customsCode=C00 **and** motCode=0 **and** partner2Code=0 in **160 of 160 cells**, yet no field in the response marks it: the descriptor fields are null in all 9,790 returned records. The factor of eight is approximate at raw floating-point precision. It decomposes into a 2.00 partner-2 multiplier and a 4.00 mode-by-customs multiplier across the affected cells; the worked example shows the latter as two 2.00 components. The second result concerns the controlled-versus-off-list comparison. On the eight mirror relations this design can build (2023, Germany↔Kazakhstan, four listed and four off-list commodities), the median absolute log value gap is **0.620** — a median percentage gap of **+85.9%** — and the controlled group does **not** separate from the off-list group: exact two-sided Mann-Whitney U test p = **0.343**, permutation p = **0.314**, n = 4 versus 4. With four observations per group, this comparison detects no difference and has little power to distinguish plausible effects. The weight dimension, which carries no valuation basis, shows a median gap of **+93.5%**, but in six of the eight relations one leg's net weight is publisher-imputed, so that comparison is reported under a flag and is not used as the decisive argument. On the data this design can reach, mirror gaps do not isolate rerouting; testing the effect requires a larger controlled comparison and evidence of physical movement beyond mirror records.

Keywords: mirror trade statistics; export controls; third-country rerouting; UN Comtrade; aggregation bias; product classification

1. Introduction

Mirror trade statistics compare two records of the same shipment: the value the exporter reports sending and the value the importer reports receiving. The gap between them is large, persistent and increasing. In the dataset documentation closest to this question, the authors report that in 2010 approximately 50%

© 2026 Academy of Public Policy and Economics · Open Access (CC BY 4.0).

of all possible country pairs had **neither** partner reporting any bilateral trade, 9.3% were reported by the exporter only, 10% by the importer only, and roughly 31% by both; among the flows reported by both partners, **nearly two-thirds exhibit discrepancies greater than 25 percent**, and the fraction diverging by more than 25 percent **has grown** over time (Bustos et al., 2026). That same work is explicit that its reconciliation ”does not attribute discrepancies to specific drivers” (Bustos et al., 2026).

Two other literatures do attribute. One reads mirror gaps as evasion: in the United States–China corridor, the authors find that under-reporting at the Chinese border to avoid value-added tax accounts for approximately two-thirds of the discrepancy, with USD 6.5 billion of tax avoided over 2002–2008 and an estimated USD 2 billion of United States tariff revenue lost to evasion at the United States border — and they state that the discrepancy ”is not caused by re-exporting through Hong Kong” (Ferrantino et al., 2012). Because that study excludes the re-export channel on its own corridor, it is the single most important adverse result for the reading this article began with, and it is carried here as a contradicted expectation rather than as a settled fact about rerouting generally: it is an abstract-level result on one corridor and one period. The other literature reads mirror gaps as rerouting: after United States anti-dumping actions against China, the positive correlation between United States imports from a third country and Chinese exports to that same third country strengthens, more so for products actually subject to duty than for similar products that are not, and more so where origin certificates are easier to alter and rerouting costs are low (Liu & Shi, 2019). Case-based work on sanctions shows the same phenomenon reached by different evidence: car re-exports to Russia move through third countries, and that study’s identification rests on administrative and trade records rather than on mirror gaps — the words ”mirror” and ”discrepancy” each occur once in its full text (Partsvaniya & Pirveli, 2024).

What none of these establishes is whether the *controlled* status of a good makes its mirror gap behave differently from an uncontrolled good’s. That is a question about the separability of a signal, and it is the question this article asks. It is worth asking because the policy documents that define ”controlled” for practitioners assert the third-country channel directly while offering them no measurement instrument (Foreign, Commonwealth & Development Office, 2024).

This article’s contribution is a boundary result. It does not estimate a rerouting effect, and it does not claim that rerouting is absent. It shows that on the data a public-preview design can actually assemble, mirror gaps do not separate controlled goods from a same-chapter control group — and it documents a measurement trap in that interface that makes the attempt harder than it looks, since the first thing a user will do with a multi-row response is add it up.

The article proceeds as follows. Section 2 sets out what the neighbouring literature establishes and what it leaves open. Section 3 describes the data, the only operable definition of ”controlled”, and the reproducible aggregation and verification rules. Section 4 reports the aggregation trap. Section 5 reports the mirror-gap noise floor and the treatment contrast. Section 6 states the evidence states this article is entitled to claim. Section 7 discusses what would have to be true for the question to close, and Section 8 concludes.

2. What the neighbouring literature establishes, and what it leaves open

2.1. Rerouting is documented, and controlled goods are its object

Third-country procurement of controlled goods is not inferred here from theory. The list that this article uses as its definition of ”controlled” exists precisely because the phenomenon was observed: it is a joint instrument of the European Union, the United Kingdom, the United States and Japan, published by the United Kingdom, and it exists to name the goods found in weapons systems and to warn businesses that Russia ”is seeking to procure restricted goods such as those on this list, via non-sanctioning countries”

(Foreign, Commonwealth & Development Office, 2024). The operational definition therefore follows the jointly maintained list rather than inferring control status from trade codes alone.

2.2. Mirror gaps as evidence, and the non-attribution gap

Mirror gaps are differences between two reporting systems, so attribution requires a model or independent evidence. The present comparison treats the off-list distribution as an empirical noise floor rather than assigning every discrepancy to rerouting.

The single most useful methodological statement in this literature is the one that declines the attribution. Bustos et al. (Bustos et al., 2026) reconcile exporter- and importer-reported values after standard cost-insurance-freight to free-on-board adjustments, cap the adjustment at 20% — a cap that applies to fewer than 1% of observations — and then state that the procedure ”does not attribute discrepancies to specific drivers” (Bustos et al., 2026). That sentence is the opening this article works in: if the best current reconciliation declines to attribute, then the question of what a mirror gap can be read as is live, and it is an empirical question about the gap’s distribution rather than a definitional one.

2.3. Contribution

This article makes two bounded contributions. First, it identifies the total-bearing row in a multi-row UN Comtrade preview response and quantifies the inflation caused by summing that response as though it were additive. Second, it compares mirror-value gaps for four listed and four same-chapter off-list commodities in one Germany–Kazakhstan cross-section. The comparison is a small-sample diagnostic of separability, not an estimate of rerouting.

The design therefore tests whether controlled status shifts the observed mirror-gap distribution in this sample. It does not infer physical movement from discrepancies and makes no novelty claim beyond the literature actually used.

3. Data, the definition of ”controlled”, and verification posture

3.1. Sources

Trade records are the United Nations Comtrade public preview interface, in the HS annual reporter–partner form (United Nations Statistics Division, 2026). Two hundred and forty-four responses were retrieved and archived; 160 of them returned data and 84 returned none. The 160 non-empty responses contain 9,790 individual returned records. The responses were parsed field by field from the archived JSON, and every quantity printed in this article is re-derived from those responses by a single script held with the project record, not read back from a prior reduction.

The interface has three properties that shape the design. It returns rows that vary along several dimensions inside one cell; it does not return a summary for the world as a partner; and it throttles requests, with the extraction run here spaced at intervals of at least 2.9 seconds (an operational parameter of the extraction, recorded in the extraction log; it is not re-derivable from the archived responses and is reported as such). Country identities are not returned in the response payload; the reporter code is carried in the request and the M49 partner codes were taken from the publisher’s own reference table.

3.2. The only operable definition of ”controlled”

”Controlled” is defined throughout as **appearing on the Common High Priority Items List** — a list-based definition, not a legal one. The list names 50 six-digit HS codes in six tiers: 4 in tier 1, 5 in tier 2, 16 in tier 3A, 9 in tier 3B, 11 in tier 4A and 5 in tier 4B. The version parsed here was first published on 30 May 2023 and last updated on 8 August 2024, both within this article’s evidence cut-off, and it is published by the Foreign, Commonwealth & Development Office (Foreign, Commonwealth & Development Office, 2024).

The alternative granularity is the United States Commerce Control List, which controls by parameterised Export Control Classification Number rather than by HS code (United States, 2026). The two are not interchangeable and this article does not map one to the other. The Commerce Control List text was retrieved but its endpoint returned no last-modified header, so the currency of the retrieved text against the 2026-02-20 cut-off could not be established; consequently no item counts or keyword counts from that text are used anywhere in this article, and the reference is used only for the structural point that the two lists control at different granularities.

The four study commodities that are on the list are HS 854231 (tier 1), 854239 (tier 1), 853690 (tier 3A) and 848210 (tier 3B). The four controls are same-chapter codes absent from the list: 848180, 848340, 853229 and 854442. Membership was tested against the list text itself, and all four controls are absent from it.

3.3. Verification rules

All printed statistics were re-derived from the 244 archived JSON responses. A cell total is the row with customsCode=C00, motCode=0 and partner2Code=0; component rows are not added to that total. Bibliographic identities were checked against registry records, and author names follow those records.

The verified corpus contains 160 non-empty cells and 9,790 returned rows. Of the 133 cells with an approximately eightfold naive-sum ratio, 132 are within 1e-6 of 8 and all 133 are within 1e-5. The partner-2 multiplier is approximately 2 and the combined mode-by-customs multiplier approximately 4 across those affected cells; separate twofold mode and customs components are established only in the worked examples. Evidence-state bookkeeping and superseded intermediate counts are retained in the internal editorial record, not used as results.

4. The aggregation trap: one cell is not one number

4.1. The response is not an additive cross-tabulation

A user who asks the preview interface for one reporter, one partner, one six-digit commodity, one year and one flow does not receive one number. Across the 160 non-empty cells, a cell returns between 1 and 384 records, with a mean of 61.2. The natural thing to do with such a response — add the rows — is wrong by a large, stable factor.

Dividing a cell's naive row sum by that cell's maximum value gives a median of **8.00**. In 133 of the 160 cells (83.1%) the factor exceeds 1.00 by more than one per cent; in 132 of those it is within one part in a million of 8.00, and one cell reaches 8.000004 — a floating-point artefact of the same factor. The remaining 27 cells have a single row and a factor of exactly 1.00. Table 1 gives the distribution.

Table 1. Aggregation diagnostics, 160 non-empty preview cells (reporter × partner × HS6 × year × flow).

Diagnostic	Value
Cells parsed	160
Records returned across those cells	9,790
Records per cell	minimum 1, maximum 384, mean 61.2
Naive row sum ÷ cell maximum	minimum 1.00, median 8.00, maximum 8.000004
Cells overstated by more than 1%	133 (83.1%)
Of those, cells within 1e-6 of 8.00	132
Cells with a single row (factor exactly 1.00)	27
Cells whose total-bearing row is customsCode=C00 and motCode=0 and partner2Code=0	160 (100%)
Cells where that row is the cell maximum	160

(continued)

Table 1 (continued)

Diagnostic	Value
Cells where it is not the cell maximum	0
Cells carrying a non-zero partner2Code label	133 (83.1%)
— of which the only such label is 899 (”unspecified”)	81
— of which other labels appear as well	52
Total-bearing rows flagged isNetWgtEstimated=true	102 (63.8%)

Note. the two percentages in the last block are of the 160 cells; the 133 labelled cells are the same 133 cells that carry the eightfold factor. The factor is re-derived from the raw responses, not from any earlier reduction.

4.2. Where the factor of eight comes from

The factor is not a coincidence and it is not a fitting parameter. Across the 133 affected cells it is the product of a measured partner-2 multiplier of about 2.00 and a combined mode-by-customs multiplier of about 4.00. The worked examples below show how the fourfold component can split into two further twofold components; that separate split is not established cell by cell across the corpus.

Partner-2 duplication. In the 133 affected cells, the same values appear a second time under a non-zero partner2Code; the sum over rows with partner2Code=0 is exactly half the naive sum, giving a multiplier of 2.00. In 81 of those cells the only non-zero label is 899, ”unspecified”; in 52 cells other labels appear as well. Code 899 is therefore the dominant duplicate label, not the only one in the corpus.

Mode-of-transport decomposition. Within one cell, motCode=0 carries the total and the non-zero motCode rows decompose it, so summing them again doubles the figure. For the 2023 German export of HS 854231 to Kazakhstan, the mode rows sum to the total exactly: 1,101,306.936 plus 220,968.794 equals 1,322,275.730, which is the motCode=0 value, and the net weights agree as well, 970.0 plus 375.0 equalling 1,345.0 kilograms.

Customs-procedure duplication. Customs procedures are not a flat duplicate either: for the 2023 German export of HS 848210 to Kazakhstan, the motCode=0 value under C00 is 9,097,409.261, under C03 it is 9,073,908.100 and under C04 it is 23,501.160, so that C00 is the parent of C03 plus C04 — the identity closes to within 0.001, the printed precision of the returned values, rather than to zero. Summing the procedures therefore doubles the figure again.

Table 2. The measured multipliers, and the two worked examples.

Quantity	Re-derived value
Cells where the partner-2 multiplier is 2.00	133
Affected cells where the combined mode-by-customs multiplier is within 1e-5 of 4.00	133
Affected cells where the two measured components reproduce the observed factor within 1e-6	133
Worked example, HS 854231: mode 1000 + mode 3200	1,101,306.936 + 220,968.794 = 1,322,275.730
Worked example, HS 854231: motCode=0 value	1,322,275.730
Worked example, HS 854231: net weight, modes 1000 + 3200	970.0 + 375.0 = 1,345.0 kg
Worked example, HS 848210: C00 (mode 0)	9,097,409.261
Worked example, HS 848210: C03 (mode 0)	9,073,908.100
Worked example, HS 848210: C04 (mode 0)	23,501.160
Worked example, HS 848210: C00 − C03 − C04	0.001

Note. both worked examples are single archived responses, 276_398_2023_854231_X and 276_398_2023_848210_X, re-derived here. The second row of the table is the value a naive row sum would never produce and a naive reader would never find.

4.3. Nothing in the response marks the total

The trap is not that the interface returns extra rows; it is that it gives no usable signal about which row is the total. Across all 9,790 returned records, the six descriptor fields (reporterDesc, partnerDesc, motDesc, customsDesc, cmdDesc, flowDesc) are null — 58,740 null values and no non-null value at all. The two flags that are populated do not identify the row either: isAggregate is true on 7,033 of the 9,790 records and is never true on exactly one row of a cell (0 of 160 cells), so it cannot pick out the parent row; isReported is true on 2,757 records and is true on exactly one row in only 41 of the 160 cells.

The rule that does work — customsCode=C00 and motCode=0 and partner2Code=0 — is the rule this article used, and it was verified per cell rather than assumed: the selected row is the cell maximum in 160 of 160 cells, with no exceptions. That rule is a statement about this endpoint's current response shape, established by inspection, not a documented contract; a user who adopts it is relying on an observed regularity.

5. The mirror-gap noise floor, and a contrast that is not detected

5.1. What this design can assemble

A mirror relation requires both partners to report the flow, in opposite directions of the same flow: one partner's export record against the other's import record. Reporting coverage is the binding constraint. Germany's cells were retrieved 162 times and returned data 133 times; Kazakhstan's returned data 17 times out of 50 retrievals; Japan's 10 times out of 32. In the specific direction that matters, Germany-to-Kazakhstan cells returned data in 51 of 80 retrievals (63.75%), while Kazakhstan-to-Germany returned data in 15 of 32, and in 2022 every one of the 16 Kazakhstan-to-Germany cells was empty. Table 3 gives the coverage that produced the sample.

This is the same coverage asymmetry that Bustos et al. describe at country-pair level (Bustos et al., 2026), appearing here at commodity level for an intermediary economy: whether a mirror relation can be built at all depends on who happened to report the commodity that year. Any design of this kind must therefore show that its sample was not selected by reporting behaviour, and this article can only show the size of the problem, not solve it.

Table 3. Reporting coverage, 244 retrieved cells.

Direction	Cells retrieved	Cells returning data	Share
All, by reporter 276 (Germany)	162	133	—
All, by reporter 398 (Kazakhstan)	50	17	—
All, by reporter 392 (Japan)	32	10	—
276 → 398	80	51	63.75%
398 → 276	32	15	46.88%
398 → 276, year 2022 only	16	0	0%

Note. cells returning data, by year, are 19 in 2021, 75 in 2022 and 66 in 2023. The Kazakhstan-to-Germany direction contains 32 queried cells, of which 15 returned data.

5.2. The noise floor, and the sample it rests on

The value gap for a relation is the absolute natural logarithm of the ratio of the importer's reported value to the exporter's reported value — the same absolute-log-difference form that the Growth Lab documentation uses (Bustos et al., 2026) — and the weight gap is the same expression applied to net weight.

Under the primary reduction rule — one relation per unordered country pair, year and commodity — eight relations exist, all of them 2023 German–Kazakhstani, four on the list and four off it. The median absolute log value gap is 0.620, a median percentage gap of +85.9%. Seven of the eight exceed

25% (87.5%), five exceed 50% (62.5%) and four exceed 100% (50.0%); the ratio of importer-reported to exporter-reported value ranges from 0.392 to 5.023, a spread of 12.8 times.

5.3. Controlled versus off-list: not detected

The point estimates point in the direction the design was looking for. The listed group's median absolute log gap is 0.879 and every one of its four relations exceeds 25%; the off-list group's median is 0.330 and three of its four exceed 25%. But neither test separates them. The Mann-Whitney U is 12 of a possible 16. For n = 4 versus 4 with no ties, the exact two-sided Mann-Whitney U test — the appropriate test for this sample, with no continuity correction — gives p = **0.343** (unrounded 0.342857). A normal approximation without continuity correction gives z = 1.155 and p = 0.248; that value is reported only as a sensitivity check on the exact test, not as an alternative or equivalent statistic, and it is not the primary result. A separate permutation test on the difference in medians (20,000 iterations, seed 20260220) gives an observed difference of +0.549 and p = **0.314** (United Nations Statistics Division, 2026).

With four relations against four, this test has almost no power against any small or moderate effect. The finding is therefore **"not detected"**. It is **not** a demonstration that the two groups are indistinguishable, and it is **not** a finding of no effect. Any sentence stronger than "not detected" would be unsupported by these eight relations. Table 4 reports the relations individually, so that a reader can see how few observations the contrast rests on.

Table 4. The eight mirror relations under the primary reduction rule.

HS6	Group	Tier	Export value	Import value	Ratio	Log gap	Derived?	Weight imputed?
848180	off-list	—	37,049,486.161	25,080,293.96	0.677	0.390	yes	yes
848210	listed	3B	9,097,409.261	4,006,490.31	0.440	0.820	yes	yes
848340	off-list	—	15,572,040.064	11,896,409.66	0.764	0.269	yes	no
853229	off-list	—	10,780.344	54,145.73	5.023	1.614	yes	yes
853690	listed	3A	5,034,235.700	3,307,919.78	0.657	0.420	yes	yes
854231	listed	1	1,322,275.730	5,214,757.91	3.944	1.372	yes	no
854239	listed	1	874,969.178	342,554.07	0.392	0.938	yes	yes
854442	off-list	—	2,273,342.642	2,170,590.02	0.955	0.046	yes	yes

Note. "export leg derived" records that the exporter-side row is flagged isAggregate=true and isReported=false; the importer-side row is flagged isReported=true in all eight relations. "Export leg imputed weight" records isNetWgtEstimated=true on that row. All eight relations are year 2023, reporter 276, partner 398.

5.4. The weight comparison, and a flag that limits it

The median absolute log weight gap is 0.660, a median percentage gap of +93.5%, above the value gap's +85.9%, and six of the eight relations exceed 25%. Since net weight carries no valuation basis, a gap in weight cannot be produced by the cost-insurance-freight to free-on-board convention, by tariffs or by exchange rates, but publisher imputation prevents that comparison from establishing whether the gaps are valuation artefacts.

That argument cannot be carried at full strength, and the reason is a flag in the data rather than a weakness of the logic. In six of the eight relations — all but HS 848340 and HS 854231 — the exporter-side row carries isNetWgtEstimated=true, which marks a publisher-imputed net weight; and in all eight relations the exporter-side row is itself a derived aggregate rather than a reported record. The weight comparison therefore mixes a reported quantity with an imputed one in six relations out of eight, and on such rows the quantity is an imputation product rather than a trade observation. Restricting to the two relations with no imputed weight on either leg leaves gaps of 0.087 and 0.850, a median of 0.468 or +59.7% — two observations, which establish nothing.

This article therefore reports the weight result as an observation under a flag and does not use it as the decisive argument. The conclusion of Section 5.3 does not depend on it: the treatment contrast is not detected on the value dimension, where no imputation is involved, and that is enough to make the separability claim fail.

Table 5. Noise floor and treatment contrast, under two reductions of the same archive.

Statistic	Primary reduction	Complete archive
Mirror relations	8	10
Median absolute log value gap	0.620	0.620
Median percentage value gap	+85.9%	+85.9%
Relations exceeding 25%	7 of 8	9 of 10
Listed group, n	4	5
Listed group, median absolute log gap	0.879	0.938
Off-list group, n	4	5
Off-list group, median absolute log gap	0.330	0.275
Mann-Whitney U	12 of 16	20 of 25
Mann-Whitney p — exact two-sided test, no continuity correction, no ties (primary statistic)	0.343	0.151
Mann-Whitney p — normal-approximation sensitivity check, no continuity correction (not the primary statistic)	0.248	0.117
Permutation p (20,000 iterations, seed 20260220)	0.314	0.122
Median absolute log weight gap	0.660	—
Median percentage weight gap	+93.5%	—

Note. the second column is not an alternative analysis chosen after the fact; it is the set of mirror relations the archived responses actually support. The primary reduction keeps one relation per unordered country pair, year and commodity, which discards the reverse flow direction. Two further relations are recoverable: the 2023 Kazakhstani export of HS 853690 to Germany (absolute log gap 0.942) and of HS 854442 (0.275). Adding them leaves the median gap unchanged and shifts the exact two-sided contrast from p = 0.343 to p = 0.151 with five relations against five (U = 20 of 25, no tied values). The normal approximation without continuity correction gives p = 0.117 for that sample. Neither reduction detects a group difference.

5.5. The unit-value contrast, reported but not used

A unit-value contrast gives the median ratio of importer unit value to exporter unit value is 1.795 for listed goods and 1.207 for off-list goods, and the record read the same-direction result as showing that the gap appears in unit prices as well as in totals. That reading is not available here. Because the exporter-side quantity is publisher-imputed in six of the eight relations (Section 5.4), a unit value computed on those rows measures the imputation parameter rather than trade. The two medians are therefore reported for completeness and are given no evidential weight in this article.

Table 6. Two quantities reported under a flag and excluded from the argument.

Quantity	Re-derived value	Why it is not used
Median weight gap, all eight relations	+93.5%	Six of eight relations carry an imputed weight on one leg
Median weight gap, two unimputed relations	+59.7%	Two observations only
Median unit-value ratio, listed group	1.795	Quantity imputed on the exporter leg in six of eight
Median unit-value ratio, off-list group	1.207	Same

6. Scope of supported claims

The archived responses support the aggregation result within the 2021–2023, eight-code sample and the reported four-versus-four mirror-gap comparison. They do not establish the stability of the eightfold pattern outside that sample, the cause of any mirror discrepancy, or physical rerouting. The treatment contrast is underpowered, and the weight comparison is not decisive because one leg is publisher-imputed in six of eight relations.

7. Discussion: what would close the question

The result is a boundary, and boundaries are only useful if it is clear what lies beyond them. Three things would change the answer.

Power. Eight relations cannot detect a moderate effect. A design with enough mirror relations to give the contrast real power would need either a wider country set or a longer period, and it would meet the coverage constraint documented in Section 5.1 first: the constraint is not request volume but who reports.

A quantity that is observed rather than imputed. The weight dimension was the right instinct, because it is free of valuation conventions, but it fails here because the exporter-side weight is publisher-imputed in six of eight relations. A design that uses a quantity which is reported on both legs, and that discloses the estimation flags before computing anything from it, would test the valuation-artefact hypothesis properly.

Evidence independent of mirror statistics. This is the binding limitation. Nothing in a mirror gap can establish where goods went; the mirror gap is a discrepancy between two records, and rerouting is a fact about physical movement. Confirming movement requires enterprise-level customs records, transport or bill-of-lading data, enforcement cases or end-use certificates, all of which are non-public. Even a significant treatment contrast would license a statement about reporting behaviour, not about diversion. The gap between a discrepancy and a movement is a gap in kind, not in degree, and no amount of additional mirror data closes it.

The comparison provides a sample-specific result. It does not show that rerouting is absent, small or unchanged. It shows that a public mirror-statistics design, run carefully on the data it can reach, does not detect the difference it was built to detect — and that the interface it must use contains an eightfold aggregation trap that would silently inflate any total a less careful user computed from it.

8. Conclusion

Third-country rerouting of controlled goods is real, documented and policy-relevant, and mirror trade statistics look like the obvious public instrument for measuring it. This article tested that instrument on the data it can actually reach and reports two things.

First, a measurement result that stands independently of the rerouting question: one cell of the Comtrade public preview interface is not one number. Over 160 cells, the naive sum of a cell's rows exceeds that cell's maximum by a factor of approximately 8.00 within the recorded precision in 133 cells (83.1%); the total-bearing row is the one with customsCode=C00, motCode=0 and partner2Code=0 in all 160 cells; and nothing in the response marks it, because the descriptor fields are null throughout and the populated flags do not single the row out. Across the affected cells the measured decomposition is 2.00 times 4.00; the worked example shows the fourfold component as two further twofold components.

Second, on the eight mirror relations this design can build, the median absolute log value gap is 0.620, a median gap of +85.9%, and the listed group does not separate from a same-chapter off-list group: exact two-sided Mann-Whitney U test $p = 0.343$, permutation $p = 0.314$, four relations against four. With four observations per group, this comparison detects no difference and has little power to distinguish plausible effects. The same result holds in the alternative reduction that recovers two further relations

from the archive (exact two-sided p = 0.151, five against five; normal-approximation sensitivity check p = 0.117). The weight dimension, which would have been free of valuation conventions, is reported under an imputation flag and is not used as the decisive argument.

The result sets a measurement condition for studies of rerouting. Mirror gaps have a large noise floor — nearly two-thirds of jointly reported flows exceed 25% even before any question of control is asked (Bustos et al., 2026) — and their interpretation on at least one named corridor has been shown to be dominated by tax evasion that the authors explicitly distinguish from re-export (Ferrantino et al., 2012). In this sample, the *controlled* status of a good does not separate the mirror-gap distributions. Episode-specific claims therefore require evidence beyond that status comparison.

9. Declarations

Evidence cut-off. No cited item was created or published after 20 February 2026, with two exceptions that are live data/legal-text services rather than dated publications: the UN Comtrade public preview interface (United Nations Statistics Division, 2026) and the Commerce Control List text on the eCFR (United States, 2026) were both retrieved on 2026-09-21. That retrieval date is a property of when this article's data collection ran; it is not evidence that the underlying trade records or regulatory text postdate the cut-off, and it does not certify the historical vintage of either source. The 244 archived Comtrade responses report 2021–2023 annual flows, within the cut-off period; the Commerce Control List text's own currency against the cut-off could not be established from its response headers, which is why (Section 3.2) no item or keyword counts from it are used anywhere in this article.

Issue metadata. Issue date, volume, article number and DOI are withheld pending confirmation by the editor-in-chief. They appear nowhere in this article.

Data and scripts. The 244 archived interface responses, the primary list document, the reference table, the extraction scripts and a verification script that re-derives every number printed here from the archived responses are held with the project record. The verification script reads the raw responses, not any earlier reduction output.

Figures. This article contains no figures, and none is claimed. Tables are used throughout because no figure in this project could be checked by the author of this article, who has no image input.

Competing interests. None declared.

10. Tables

Table 7. Data and identification limits.

Item	Consequence
No public physical-movement evidence	Mirror discrepancies cannot establish rerouting
No verified HS-to-control-entry legal mapping	"Controlled" remains the list-based operational definition
Four observations per group	The differential comparison has little power
Publisher-imputed weight in six of eight relations	Weight and unit-value contrasts are not decisive

Statements

Funding. This work received institutional support from the Academy of Public Policy and Economics (APPE).

Competing interests. The authors are affiliated with APPE, which operates this journal. No external competing interests are declared.

References

Bustos, S., Jackson, E., Torun, D., Leonard, B., Tuzcu, N., Lukaszuk, P., White, A., Hausmann, R., & Yıldırım, M. A. (2026). Tackling discrepancies in trade data: The Harvard Growth Lab international trade datasets. *Scientific Data*, *13*(1). https://doi.org/10.1038/s41597-025-06488-2

Ferrantino, M. J., Liu, X., & Wang, Z. (2012). Evasion behaviors of exporters and importers: Evidence from the U.S.–China trade data discrepancy. *Journal of International Economics*, *86*(1), 141–157. https://doi.org/10.1016/j.jinteco.2011.08.006

Foreign, Commonwealth & Development Office. (2024). *Russia sanctions – Common high priority items list*.

Liu, X., & Shi, H. (2019). Anti-dumping duty circumvention through trade rerouting: Evidence from Chinese exporters. *The World Economy*, *42*(5), 1427–1466. https://doi.org/10.1111/twec.12747

Partsvaniya, V., & Pirveli, E. (2024). Western sanctions evasion through third countries: The case of sanctioned cars re-export to Russia. *NISPAcee Journal of Public Administration and Policy*, *17*(2), 80–108. https://doi.org/10.2478/nispa-2024-0014

United Nations Statistics Division. (2026). *UN Comtrade public preview interface (HS, annual, reporter–partner)*. https://comtradeapi.un.org/public/v1/preview/C/A/HS

United States. (2026). *Commerce Control List, 15 C.F.R. pt. 774*.