P-ISSN 2664-0422	Pakistan Social Sciences Review	March-April 2026, Vol. 10, No.2
O-ISSN 2664-0430	https://doi.org/10.35484/pssr.2026(10-II)21	[277-297]

Pakistan Social Sciences Review
www.pssr.org.pk

RESEARCH PAPER

Trade Misinvoicing and Illicit Financial Outflows from Pakistan: Product Level PPML Gravity Estimates with a GVC Exposure Map

[1]Nasir Ali and [2]Syed Irshad Hussain

1. PhD Economics Research Scholar at Department of Social Sciences, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi, Pakistan.
2. Associate Professor, Economics, Department of Social Sciences Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi, Pakistan.

Corresponding Author: nasiralisau@gmail.com

ABSTRACT

Illicit financial flows (IFFs) are difficult to observe directly; this paper develops a transparent, trade-data screening framework for Pakistan that treats misinvoicing signals as upper-bound risk proxies. Using UN Comtrade HS6 bilateral flows for 1995–2024, we estimate product-level PPML gravity models to generate conditional-mean 'expected trade' benchmarks that accommodate zeros and heteroskedasticity. Robustness checks use BACI-reconciled bilateral flows. We harmonize valuation by identifying CIF–FOB wedges to place imports and exports on a comparable basis, then construct directional discrepancy measures consistent with outward leakage (import overinvoicing and export underinvoicing). Results show an episodic, shock-driven discrepancy pattern with a pronounced mid-2010s spike and higher variance in recent years, rather than a smooth trend. Risk is highly concentrated in a small set of commodity/agri-food sectors (notably cereals and oilseeds) and in specific partner–product clusters. Mapping HS products to input–output sectors links hotspots to GVC exposure, informing targeted customs risk management and audit prioritization.

KEYWORDS Illicit Financial Flows (IFFS), Trade Misinvoicing , PPML (Poisson Pseudo–Maximum Likelihood) , Structural Gravity Model

Introduction

Illicit financial flows (IFFs) remain inherently difficult to quantify because the actors and intermediaries involved intentionally conceal transactions, ownership, and cross-border routes, limiting the feasibility of direct observation in official datasets (Cobham & Janský, 2020). In response, empirical research has increasingly relied on indirect measurement approaches that use observable inconsistencies as signals of hidden value transfers, with international trade data providing an unusually structured basis for such inference because flows are reported by both exporting and importing economies.(Cobham & Janský, 2020; Linsi et al., 2023). In this context, trade misinvoicing, understood as the deliberate misstatement of price, quantity, or classification on customs documentation, has become one of the most empirically tractable channels for cross-border value shifting, as systematic discrepancies in "mirror" trade statistics can be analyzed at scale and at fine product detail. (Linsi et al., 2023; Nitsch, 2017).

Discrepancies in bilateral trade statistics are pervasive and reflect a mix of structural features of how trade is recorded, such as valuation differences between imports and exports, timing and shipment lags, transshipment and re-exports, and heterogeneous compilation practices, as well as potential deliberate misreporting embedded in customs declarations.(Linsi et al., 2023; Nitsch, 2017). Because these legitimate sources of asymmetry can be substantial and systematic, mirror-gap based measures should be interpreted with caution and framed as risk-signaling proxies rather than as direct quantifications of illicit financial flows.(Cobham & Janský, 2020; Nitsch, 2017). The practical implication is that trade-gap estimates are most defensible when used to prioritize scrutiny, by partner, product, or period, while recognizing that attribution of any residual gap to misinvoicing remains contingent on the assumptions and adjustments applied. (Cobham & Janský, 2020; Linsi et al., 2023).

Given that trade-statistic discrepancies are best interpreted as screening signals rather than definitive evidence of misinvoicing, a natural next step is to examine whether these signals covary systematically with bilateral trade frictions and policy-relevant incentives in a theoretically disciplined framework. (Cobham & Janský, 2020; Nitsch, 2017). Gravity models provide the standard workhorse for bilateral trade analysis, but credible inference on trade costs requires controlling for multilateral resistance, the general-equilibrium trade-barrier terms that, if omitted, induce bias in estimated effects of bilateral frictions (Anderson & Van Wincoop, 2003). At the product level, bilateral matrices are typically sparse and contain many zero flows, making conventional log-linear OLS both operationally restrictive and econometrically problematic (Silva & Tenreyro, 2006). PPML estimation addresses these issues by estimating the gravity model in multiplicative form, remaining consistent under heteroskedasticity and accommodating zero trade flows without ad hoc transformations, which is particularly important when modeling product-level variation in trade (and in discrepancy-based risk measures) across country pairs (Silva & Tenreyro, 2006). In short, product-level PPML gravity combines theoretical structure (multilateral resistance) with an estimator suited to sparse, heteroskedastic trade data, enabling discrepancy measures to be analyzed as systematic patterns in trade rather than as standalone anomalies.

Because trade-gap measures are best treated as screening signals and gravity estimates identify correlates of bilateral frictions rather than illegality, the analysis is complemented with a GVC exposure map to situate these signals within the underlying production structure (Cobham & Janský, 2020; Nitsch, 2017). Global value chain (GVC) indicators derived from inter-country input–output (ICIO) tables enable sector-level measures of participation and value-added trade, allowing discrepancies and misinvoicing signals to be mapped to upstream and downstream stages where value is created and transmitted across borders (Borin et al., 2025; Koopman et al., 2014). Operationally, the map draws on harmonized GVC indicators available through the World Bank's WITS platform and on ICIO sources such as the WIOD 2016 release, which provide consistent country-sector accounts for tracing value-added linkages across production networks (Borin et al., 2025; Mishina et al., 2021; Timmer et al., 2015). Consistent with the measurement cautions in the IFF literature, this GVC component is framed as risk mapping, highlighting where exposure and concentration appear in the supply-chain topology, rather than as a claim that mapped hotspots constitute verified illicit activity (Cobham & Janský, 2020; Nitsch, 2017).

This paper contributes a policy-relevant measurement framework for Pakistan by combining product-level structural gravity estimation with risk-oriented misinvoicing screening and production-network contextualization. First, it provides product-level PPML gravity estimates for Pakistan's bilateral trade flows and uses the model's fitted values as a transparent benchmark to construct misinvoicing proxies by product, partner, and year. Second, rather than relying on ad hoc CIF/FOB uplift factors, it estimates trade-cost components using the CIF–FOB valuation structure embedded in international trade recording (imports typically reported c.i.f. and exports f.o.b.), aligning the measurement strategy with modern gravity-based approaches to misinvoicing assessment. Third, the paper links the resulting misinvoicing intensity patterns to sectoral GVC participation and value-added trade indicators to produce a policy-oriented vulnerability map for Pakistan that highlights where screening signals concentrate within the country's production structure, while maintaining the interpretation of these outputs as risk mapping rather than proof of illegality.

Literature Review

Structural gravity reframes bilateral trade as an outcome that is jointly determined by pair-specific barriers and each country's broader trading environment, so that a given bilateral barrier matters in relative terms rather than in isolation. This insight is formalized through multilateral resistance terms, which capture the idea that trade between two countries depends on the bilateral barrier relative to the average resistance each country faces with all partners (Anderson & Van Wincoop, 2003). The implication is conceptual as well as empirical: two dyads can face the same observable bilateral barrier and still exhibit different trade volumes if their overall trading environments differ, because the "outside options" created by alternative partners change the effective salience of any one bilateral link(Anderson & Van Wincoop, 2003). This logic provides the core rationale for exporter and importer fixed effects as standard controls in gravity estimation. Rather than treating fixed effects as a generic device to raise explanatory power, the structural framework interprets them as an empirically tractable way to absorb country-specific resistance components that are otherwise unobserved but theoretically required (Anderson & Van Wincoop, 2003; Head & Mayer, 2014). In panel applications, the same reasoning extends naturally to exporter–time and importer–time fixed effects to accommodate time-varying shifts in country-level conditions that would otherwise spill into bilateral coefficients(Head & Mayer, 2014). With these controls in place, identification of bilateral frictions is disciplined: estimates are interpreted as pair-specific effects net of each side's broader context, which is precisely what the structural gravity approach requires for coherent inference(Anderson & Van Wincoop, 2003). A further step in the modern literature is to show that fixed-effects implementations are not only conceptually aligned with structural gravity but also compatible with the internal accounting restrictions implied by the theory when estimated with standard multiplicative-form methods demonstrates that PPML estimation with exporter and importer fixed effects automatically respects the adding-up constraints associated with structural gravity, reinforcing the interpretation of fixed effects as theory-consistent objects rather than mere statistical "controls." (Fally, 2015). This bridge between theory and practice matters most in applied settings where fitted trade flows are used as a benchmark, because the credibility of the benchmark hinges on whether the specification properly separates pair-level influences from country-level context in a way consistent with the structural system. (Fally, 2015; Head & Mayer, 2014).

Log-linearizing gravity equations and estimating by OLS can produce biased coefficient interpretations when the dispersion of trade flows varies with their scale, because the log transformation breaks the link between the conditional mean in levels and the fitted model in logs. Estimating gravity in multiplicative form using Poisson pseudo–maximum likelihood (PPML) avoids this log-linearization problem and remains consistent under standard quasi-likelihood conditions when the conditional mean is correctly specified. This property is particularly important in trade data, where scale differences across dyads and products are large and heterogeneity in dispersion is pervasive (Silva & Tenreyro, 2006). PPML also accommodates zero trade flows directly in levels, avoiding ad hoc practices such as dropping zeros or adding arbitrary constants prior to logging. This matters most in disaggregated bilateral settings, especially product-level data, where zeros are frequent and mechanically excluding them changes the estimating sample and the interpretation of fitted benchmarks (Silva & Tenreyro, 2006). The prevalence of zeros is a documented feature of bilateral trade matrices and is central to applied work that uses granular trade datasets (Helpman et al., 2008).

Computational advances have made PPML feasible in large gravity panels with multi-way fixed effects, which are often required in theory-consistent specifications. The ppmlhdfe implementation provides fast estimation with high-dimensional fixed effects and explicit procedures for common estimation pathologies in sparse data (Correia et al., 2020).

Quantitative estimates of trade misinvoicing vary widely across studies because they depend critically on how researchers translate observed trade-data irregularities into an inferred measure of value shifting. This dependence arises because the same observable “gap” can be generated by multiple mechanisms, some innocuous and some fraudulent, so empirical procedures must impose assumptions about what portion of a discrepancy is attributable to misreporting rather than to statistical frictions (Nitsch, 2017). As a result, small changes in design choices, how valuation differences are treated, how re-exports or timing differences are handled, what is assumed about partner-country reporting, and how outliers are filtered, can produce materially different aggregate totals, even when the underlying trade data are identical (Cobham & Janský, 2020; Nitsch, 2017).

This sensitivity has led to a growing emphasis on transparency in interpretation, including framing results as screening measures or upper bounds rather than as point estimates of illicit flows. In this framing, the objective is not to claim definitive identification of illegality from trade statistics alone, but to create empirically grounded indicators that are comparable across products, partners, and time, and that can guide prioritization for further scrutiny. The methodological burden therefore shifts from “producing a single number” to documenting the mapping from data to indicator, making assumptions explicit, checking robustness across plausible alternatives, and avoiding interpretations that exceed what the data can support (Cobham & Janský, 2020; Nitsch, 2017).

Gravity-based approaches can be seen as one response to this challenge because they estimate an expected level of trade from observed determinants and then treat systematic residual discrepancies as misinvoicing proxies. A key feature of this strategy is that it does not require assuming that any particular reporter’s statistics are “correct” and therefore avoids the strong identifying premise often embedded in mirror-gap methods (Paz & UNU-WIDER, 2022). Instead, the benchmark is produced within a unified empirical structure, and deviations from that benchmark are interpreted as risk-relevant anomalies, which can then be summarized by product, partner, and year, while keeping the interpretation aligned with the broader caution that measurement remains indirect (Cobham & Janský, 2020; Paz & UNU-WIDER, 2022).

Unit-value “anomalies” are not uniquely diagnostic of misreporting because observed prices can shift mechanically with the quality composition of what is shipped, even within narrowly defined product codes. When transport costs include a non-proportional component, they can change the relative delivered cost of different quality tiers and induce systematic sorting toward higher-quality varieties in more costly destinations. In that environment, destination-varying unit values can arise because the mix of shipped varieties shifts toward higher-quality goods where shipping is expensive, rather than because invoices are manipulated. This creates an empirical caveat for price-based screens: unusually high unit values can reflect real shipping mechanisms operating through composition changes, so treating price outliers as direct evidence of misreporting risks over-attribution. (Hummels & Skiba, 2004).

While price anomalies can reflect composition, the evidence on reporting gaps indicates that discrepancies are often systematic rather than random noise around valuation conventions. After accounting for standard transport-cost proxies and other covariates, residual differences between importer- and exporter-reported trade values vary predictably with tariff levels and institutional features tied to monitoring and enforcement, including corruption and auditing standards. This pattern supports an incentives-based interpretation in which misreporting responds to policy and governance environments rather than arising solely from benign statistical frictions (Kellenberg & Levinson, 2019). Taken together with the quality-composition caveat, the implication for measurement is to separate "price irregularity" from "reporting irregularity": price movements can be mechanically generated by real shipping mechanisms, whereas discrepancies that co-move with incentives and oversight indicators provide more direct support for strategic misreporting (Hummels & Skiba, 2004; Kellenberg & Levinson, 2019).

Data

The empirical analysis uses UN Comtrade HS6 bilateral trade data (imports and exports) by reporter–partner–year as the primary source for product-level trade flows. UN Comtrade provides long-run annual coverage starting in 1962 across multiple nomenclatures, while HS-based reporting is available from 1988 onward, with later HS revisions entering as countries adopted them. The study restricts the panel to 1995–2024 to balance long coverage with a stable HS-era reporting window and sufficient product-level density for PPML estimation at the HS6 level. The end year 2024 reflects the latest year observed in the UNdata/Comtrade extracts at the time of compilation, where recent records include 2024 trade totals for multiple reporters.

Bilateral distance and standard dyadic controls (e.g., contiguity, common language, colonial ties) are taken from the CEPII Gravity Database, which is explicitly assembled to support gravity estimation and is organized at the origin–destination–year level with broad country coverage and a wide set of bilateral determinants. Geographic distances are drawn from CEPII GeoDist, which provides multiple distance measures constructed from city-level population distributions and country geography, alongside core dyadic indicators commonly used in gravity specifications. These inputs provide consistent and replicable measurement of bilateral geography and proximity variables that are otherwise costly to reconstruct from raw spatial data, and they align with established practice in the gravity literature that treats distance and dyadic ties as baseline covariates. (Head & Mayer, 2014; UN Trade and Development, 2024).

For robustness checks, bilateral HS6 trade flows are additionally drawn from CEPII BACI, which constructs a reconciled bilateral flow from UN Comtrade by combining importer- and exporter-reported values and explicitly addressing systematic differences between c.i.f.-valued imports and f.o.b.-valued exports. The BACI documentation describes a harmonization procedure that (i) estimates and removes c.i.f. components to "fobize" imports and (ii) weights reports by assessed reporter reliability to reconcile mirror declarations into a single bilateral flow. The January 2026 release notes document ongoing updates to source data and adjustments to the fobization procedure, supporting use of BACI as a robustness benchmark against raw Comtrade-based measures.

Sector-level GVC exposure measures are taken from the WITS Global Value Chains database, which provides harmonized indicators of GVC participation and trade in value added drawn from major inter-country input–output sources, including the WIOD 2016 release. The WIOD 2016 release provides inter-country input–output tables for 2000–2014 and is one of the core IO sources explicitly referenced in the WITS GVC documentation for constructing comparable country–sector indicators.

Because HS revisions create mechanical breaks in product codes, the paper documents HS-version transitions and applies concordances or HS-aggregation strategies to avoid spurious discontinuities driven by classification updates rather than underlying trade dynamics. (UNSD 2005). The harmonization step is also informed by the structure of BACI releases, which are provided separately by HS vintage with clearly stated year coverage (e.g., HS92 covering 1995–2024), making HS-version boundaries explicit for panel construction. The estimation sample is constructed to retain zero flows in product–partner–year cells, consistent with PPML's strengths for nonnegative dependent variables with many zeros and to avoid inducing sample selection through log transformations(Silva & Tenreyro, 2006).

Estimate product-level PPML gravity for exports and imports

The first step constructs a model-based benchmark for what Pakistan's bilateral trade in each HS6 product would be expected to look like, given standard gravity determinants and the broader trading environment faced by each country. This benchmark is obtained by estimating a product-by-product PPML gravity model in levels and then using the fitted conditional mean as "expected trade" for each exporter–importer–year observation (Paz & UNU-WIDER, 2022; Silva & Tenreyro, 2006). The benchmark is estimated separately for exports and imports so that subsequent measurement can distinguish export-side and import-side discrepancy patterns rather than collapsing them into a single mirror gap (Paz & UNU-WIDER, 2022).

Estimation unit and sample construction

Let ppp index HS6 products, iii exporters, jjj importers, and ttt years (1995–2024). For each product p, the dataset is organized at the dyad–year level and is rectangularized (i.e., includes zero flows) to preserve the full support of potential trading relationships. Retaining zeros is essential because the dependent variable is nonnegative and sparsity is pervasive at HS6; PPML can incorporate zeros naturally without ad hoc transformations (Silva & Tenreyro, 2006).

Baseline PPML gravity specification (product-level)

For each HS6 product p, estimate the multiplicative gravity equation:

$$X_{ijtp} = \exp\left(\alpha_{itp} + \delta_{jpt} + \beta_p' Z_{ij} + \gamma_p' W_{ijt}\right) + \varepsilon_{ijpt}$$

Where X_{ijtp} is the observed bilateral trade value (exports or imports) for product p from $i \ to \ j$ in year t. $\alpha_{itp} \ and \ \delta_{jpt}$ are exporter-year and importer-year fixed effects (allowed to be product-specific by running the model separately by p. Z_{ij} includes time-invariant dyadic controls (distance, contiguity, common language, colonial ties). W_{ijt} includes any time-varying bilateral controls available and identified under the chosen fixed-effect structure.

The model is estimated by PPML in levels (not log-linear OLS), so the fitted conditional mean $\hat{\mu}_{ijtp} = exp(.)$ is interpreted as expected trade for that dyad–year–product cell (Paz & UNU-WIDER, 2022; Silva & Tenreyro, 2006).

Why exporter-year and importer-year fixed effects are required

The structural gravity framework implies that bilateral trade depends on bilateral frictions relative to country-specific "multilateral resistance" terms, so empirical gravity should condition on exporter and importer terms that capture each country's overall trading environment. Implementing exporter-year and importer-year fixed effects absorbs time-varying country-level shocks and provides a theory-consistent proxy for those multilateral resistance components in panel gravity settings (Anderson & Van Wincoop, 2003). Using PPML with this structure also aligns with the modern recommendation to estimate gravity in multiplicative form while handling zeros directly in levels (Silva & Tenreyro, 2006).

Practical implication for identification: With exporter-year and importer-year fixed effects, identification of dyadic covariates comes from cross-dyad variation that is not absorbed by country-year terms.

Gravity covariates: CEPII Gravity / GeoDist

Dyadic covariates are taken from the CEPII Gravity Database, which is explicitly constructed for gravity applications and includes standard bilateral determinants such as distance and measures of cultural/historical proximity at the dyad level (UN Trade and Development, 2024). In practice, the following CEPII dyadic controls are included where identified under the chosen fixed-effect structure.

Identify trade-cost components using the CIF–FOB structure

The trade statistics used in this study embed a valuation asymmetry: imports are typically recorded as c.i.f. (including insurance and freight) while exports are recorded as f.o.b. (excluding those components). This valuation structure provides an empirical lever to recover product-level transport/insurance components instead of imposing a constant "rule-of-thumb" CIF/FOB uplift that ignores heterogeneity across products, partners, and distances. (Paz & UNU-WIDER, 2022).

Operationally, Step 2 uses dyad–product observations where both sides of the flow are informative to estimate an ad-valorem wedge between c.i.f. and f.o.b. values that is attributable to insurance and freight, allowing this wedge to vary across products and dyads rather than being fixed globally. The estimation is implemented at the HS6 product level to preserve the key heterogeneity in shipping and insurance costs that is otherwise averaged away in aggregate trade data. The output of Step 2 is a set of estimated product-specific cost components that can be used to place export and import values on a comparable valuation basis before constructing misinvoicing indicators (Paz & UNU-WIDER, 2022).

Two design choices in Step 2 matter for transparency and robustness. First, the estimation treats the c.i.f.–f.o.b. wedge as a quantity- and product-dependent component rather than a constant margin, which is central to avoiding mechanical misclassification of high-shipping-cost products as "over-invoiced." Second, the approach does not require designating any reporter as inherently "reliable" for the true value of the shipment; instead, the wedge is recovered empirically within the trade-cost component model and then used consistently across all dyads and products (Paz & UNU-WIDER, 2022).

A product-level estimate of the insurance-and-freight component (or its ad-valorem equivalent) for each dyad–year, which maps c.i.f. imports into an f.o.b.-type valuation. A harmonized valuation basis so that discrepancies can be interpreted as deviations from expected trade rather than as mechanical differences caused by c.i.f. versus f.o.b. recording conventions. (Paz & UNU-WIDER, 2022).

Construct misinvoicing measures (upper bounds)

Step 3 defines import and export misinvoicing as the discrepancy between reported trade values and gravity-implied expected trade, after adjusting for the transport/insurance components identified in Step 2 so that values are compared on a consistent valuation basis. The construction is performed separately for exports and imports, producing two distinct series of misinvoicing proxies by product–partner–year that can be analyzed without forcing netting across directions. This separation is important because opposite-signed discrepancies can offset each other in aggregated net measures, obscuring meaningful patterns and potentially understating the extent of abnormal reporting behavior in either channel. Conceptually, the method treats the fitted gravity benchmark as the model-based “expected trade” and interprets the residual discrepancy as the portion of reported trade that is not explained by the benchmark once the valuation wedge has been accounted for (Paz & UNU-WIDER, 2022). The resulting objects are reported as upper bounds rather than point estimates of deliberate misreporting because residual discrepancies can also reflect other non-fraud mechanisms, including statistical practices, timing differences, classification issues, and other frictions that are not fully captured by the model (Cobham & Janský, 2020; Nitsch, 2017). This interpretation choice is not cosmetic: it clarifies that the measures are designed for screening, prioritization, and pattern detection across products and partners, rather than for adjudicating individual transactions.

Pakistan’s illicit outflow proxy is constructed by combining the misinvoicing components that are directionally consistent with resources leaving the country, focusing on import over-invoicing and export under-invoicing rather than netting all discrepancies into a single balance. This directional construction is used to avoid mechanical cancellation across channels, an issue that can arise when outward- and inward-oriented discrepancies offset each other in aggregate measures, thereby preserving the informational content of each channel for trend and concentration analysis. The resulting object is treated as an upper-bound screening proxy (not a literal measure of illegality), because trade discrepancies can also arise from non-misreporting mechanisms such as valuation conventions, timing, product classification issues, and other reporting frictions that cannot be fully ruled out with trade data alone.

The GVC exposure mapping translates HS-level misinvoicing intensity into a sectoral lens by mapping HS products into input–output (IO) sectors, aggregating misinvoicing intensity to sector–year measures for Pakistan, and linking these sector–year measures to harmonized indicators of GVC participation and value-added trade from the WITS GVC database, which is explicitly built from major global IO sources including WIOD. The IO backbone for this overlay is consistent with the WIOD 2016 release (covering 2000–2014), which is one of the sources referenced by WITS for constructing comparable country–sector exposure indicators. The resulting sectoral heatmap is presented as a vulnerability overlay highlighting where discrepancy intensity coincides with external production exposure rather than as causal evidence that GVC participation generates misinvoicing. This interpretation is consistent with the broader measurement caution that trade-based misinvoicing indicators are indirect and assumption-sensitive, and therefore most defensible when used for prioritization and triangulation rather than causal attribution.

Results and Discussion

PPML gravity for Pakistan's exports: variables, econometric logic, and why this identifies the paper's IFF-related proxy (exports-only)

The estimates in column (1) come from a PPML gravity model of bilateral export values in USD, where the conditional mean of trade is modeled in levels as an exponential function of bilateral trade-cost proxies and absorbed fixed effects. Using PPML is econometrically appropriate for highly disaggregated trade data because trade values are skewed, include many small flows, and exhibit heteroskedasticity in levels; PPML remains consistent under general forms of heteroskedasticity when the conditional mean is correctly specified, and the robust (sandwich) variance estimator provides inference that does not rely on Poisson equidispersion. The inclusion of year fixed effects absorbs common time shocks and global movements in trade, while HS6 fixed effects absorb systematic product-level scale differences across thousands of categories; consequently, identification of bilateral coefficients comes from cross-partner variation in the dyadic covariates within year and product strata, i.e., why exports of the same HS6 product in the same year differ across destination partners as a function of geography and historical ties.

Table 1
PPML Gravity Model Regression

Variable	(1) PPML (Robust SE)
ln(Distance)	0.0276
	(0.0224)
Contiguity	1.6386***
	(0.0632)
Common language	0.2748***
	(0.0287)
Colonial tie	1.3465***
	(0.0391)
Constant	14.0477***
	(0.2000)
Year FE	Yes
HS6 FE	Yes
Observations	1,441,656
Dropped (singletons/separation)	106
Wald $\chi^2(4)$	4,518.85
Pseudo R^2	0.4665

In PPML Gravity Model Regression table the most economically and statistically salient result is the very strong role of durable bilateral ties. Contiguity enters with a large and precisely estimated coefficient (1.6386, SE 0.0632, $p<0.01$), implying a multiplicative effect of approximately $\exp(1.6386) \approx 5.15$. Interpreted as a semi-elasticity in the PPML log-link, this suggests that, conditional on year and HS6 composition, exports to contiguous partners are expected to be about five times larger than to non-contiguous partners, consistent with large border-related reductions in trade costs and persistent cross-border commercial networks. Colonial ties are similarly large (1.3465, SE

0.0391, $p<0.01$), implying exp(1.3465) ≈ 3.84, or roughly a 284% higher expected trade level relative to dyads without such ties. Common language is positive and precisely estimated (0.2748, SE 0.0287, $p<0.01$), corresponding to exp(0.2748) ≈ 1.32, i.e., about a 32% higher expected export level, consistent with reduced information frictions and contracting costs.

By contrast, the log-distance coefficient is small, positive, and statistically indistinguishable from zero (0.0276, SE 0.0224, $p \approx 0.22$). In a standard gravity setting, distance typically proxies iceberg trade costs and is expected to be negative; a weak or unstable distance effect in a high-dimensional fixed-effects PPML specification should not be read as evidence that distance is irrelevant. Rather, it can reflect that, once strong discrete dyadic ties (border, language, colonial links) and rich composition controls (HS6 and year fixed effects) are included, the residual explanatory content of continuous distance measures may be attenuated, especially if distance is measured with error or is correlated with omitted dyad-level trade-cost components not explicitly modeled. Econometrically, this is a common feature when the FE structure and other bilateral covariates absorb much of the systematic trade-cost variation that distance would otherwise capture.

The model is jointly highly significant (Wald $\chi^2(4)=4518.85$, $p<0.001$), and the pseudo R^2 of 0.4665 indicates substantial predictive power for a very granular trade dataset. With 1,441,656 observations, standard errors are tight, and the significance of the network-type covariates is not an artifact of low power. The estimator drops 106 observations because they are singletons or exhibit separation under the absorbed fixed effects, which is standard behavior in PPMLHDFE implementations and generally improves numerical stability without materially affecting inference at this scale. The estimation appears well-behaved and converged, supporting the use of fitted values as a benchmark conditional mean for prediction-oriented applications.

Because PPML produces fitted values in levels, the residual defined as observed minus fitted is directly interpretable in USD at the dyad–product–year cell and can be aggregated without log retransformation issues. The strong, stable effects for contiguity, common language, and colonial ties suggest that the benchmark embeds key persistent determinants of bilateral trade, reducing the risk that residual discrepancies mechanically reflect obvious trade-cost fundamentals. At the same time, since the fixed effects are year and HS6 (rather than exporter-year and importer-year), residuals may still absorb time-varying country-specific shocks or partner-specific demand shifts not fully controlled by the current FE structure; in a high-end journal framing, this motivates robustness checks with richer multilateral-resistance controls if the residual is to be interpreted as a misinvoicing or IFF-risk screen rather than a purely statistical anomaly measure. Average misinvoicing intensity: absolute change

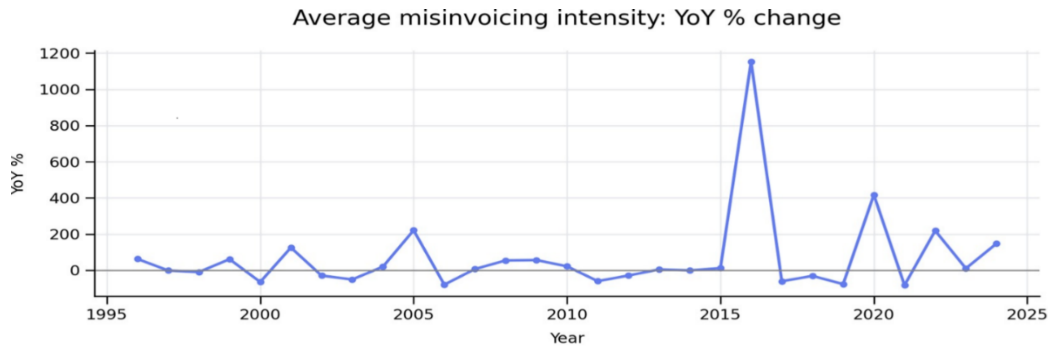

Figure 1 Average misinvoicing Intensity

In average misinvoicing intensity table the absolute-change series usefully complements the YoY plot by stripping out percentage base effects and therefore highlighting level-relevant movements in the average screening proxy: before 2015, changes are generally small, low-amplitude, and frequently cross zero, which is consistent with benchmark deviations that do not cumulate monotonically at the aggregate-average level, an outcome aligned with the gravity design in which exporter–time and importer–time fixed effects absorb broad macro shocks and multilateral resistance, leaving the residual to reflect granular dyad/product-specific variation. From an inference perspective, the repeated sign switching suggests mean reversion and/or offsetting positive and negative deviations across products and partners, underscoring why directional components (import overinvoicing versus export underinvoicing) are more informative than net measures that can mechanically cancel. A defining feature is the large positive jump around 2016 (low teens) followed immediately by a large negative correction around 2017 (about −8 to −9), indicating a genuine level shift (not merely a YoY denominator artifact) but limited persistence, which is consistent either with a short-lived, broad-based surge in discrepancy intensity or a one-off measurement/coverage/classification disturbance, so attributing it to deliberate misreporting without corroboration would be statistically overreaching. After 2018, the renewed positive/negative swings (including a notable increase around 2020 and continued oscillations into 2024) imply persistently elevated variance relative to the pre-2015 period, pointing to a variance-regime shift and motivating empirical work that allows time-varying dispersion and is robust to influential years. Overall, the evidence supports episodic, non-linear accelerations in screening intensity rather than a smooth upward trend, reinforcing the need for product/partner decomposition and valuation-structure adjustments (e.g., CIF–FOB) to reduce mechanical discrepancy sources before drawing stronger behavioral interpretations. Average misinvoicing intensity: actual and 5-year forecast

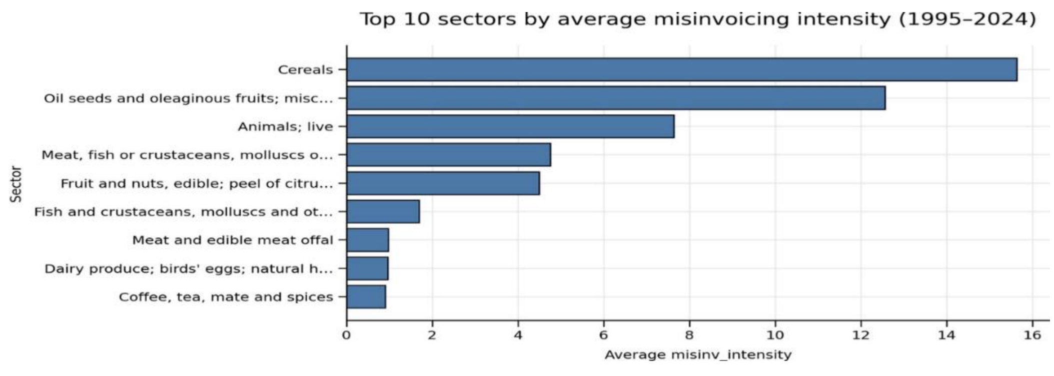

Figure 2 Top 10 sector by Average Misinvoicing Intensity

In top 10 sector by average misinvoicing Intensity the time-series panel shows an empirical trajectory characterized by moderate fluctuations for most of the sample, punctuated by a large, transient spike around the mid-2010s (peaking near 14–15), followed by a sharp normalization and renewed variability into the early-2020s (ending with an uptick near 7 in the final observed year). The inclusion of an ETS-style forecast (dashed) that settles into a nearly flat path around 3–4 suggests that, conditional on the model's smoothing structure, the process is estimated to be mean-reverting toward a mid-level baseline rather than exhibiting a persistent upward trend. Importantly, the plotted approximate 95% prediction interval is wide (visually spanning from around negative/near-zero values up to high single digits), which is an inferential indicator that the series contains substantial innovation variance and is influenced by episodic shocks, most notably the earlier spike, making long-horizon prediction intrinsically uncertain. In the paper's methodological terms, this uncertainty is exactly what one would expect when the dependent object is a residual-based screening measure: residuals inherit volatility from changes in trade composition, reporting practices, and measurement frictions that are only partially absorbed by the gravity benchmark and valuation adjustments. 1 A rigorous interpretation is therefore: (i) the central forecast implies no deterministic escalation in average intensity in the near term; (ii) the interval width implies policymakers cannot rule out either renewed elevation or near-normalization; and (iii) the proper inferential next step is not to over-interpret the forecast path itself, but to examine whether the volatility is driven by sectoral/partner concentration (as in JPG 4) and whether the spike years correspond to identifiable compositional shifts or reporting breaks that the paper explicitly flags as potential non-illicit drivers of discrepancies. Top 10 sectors by average misinvoicing intensity (1995–2024)

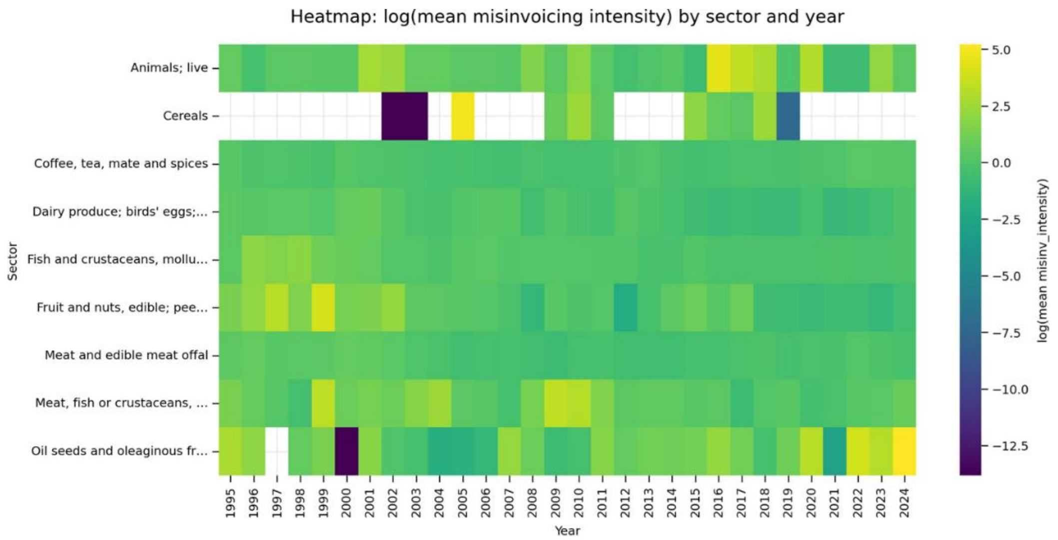

Figure 3 Heatmap – Top 10 sector average misinvoicing Intensity

In this heatmap of top 10 sector average misinvoicing Intensity bar chart documents a highly concentrated sectoral risk profile in long-run average misinvoicing intensity over 1995–2024, with the distribution exhibiting a clear right tail dominated by a small subset of primary/agri-food sectors. The ranking is led by Cereals (approximately 16 units of average intensity) and Oil seeds and oleaginous fruits (roughly 12–13), followed by a notable drop to Animals; live (8) and a mid-tier cluster including Meat/fish/crustaceans and Fruit and nuts centered around 4–5. The remaining sectors shown (e.g., fish products more broadly, meat offal, dairy/eggs/honey, coffee/tea/spices) appear to have markedly lower long-run averages (roughly 1–2 or below). Statistically, this steep gradient is consistent with a heavy-concentration pattern: the implied cross-sector dispersion is not incremental but "lumpy," where a few sectors account for a disproportionate share of the average discrepancy signal. In the paper's inferential framing, such concentration is precisely the kind of output expected from an anomaly/residual-based screen: systematic deviations from gravity-implied expected trade (after valuation harmonization) tend to cluster where price dispersion, quality heterogeneity, and documentation complexity are structurally higher, rather than being evenly distributed across the trade basket. 1 The appropriate journal conclusion is therefore not that these sectors are "most illicit," but that they represent priority sectors for targeted scrutiny and decomposition (partner-product drill-down, persistence tests, robustness to HS concordance changes), because they dominate the long-run mean of the screening proxy. Heatmap: Misinvoicing Risk Index (MRI) by sector and year

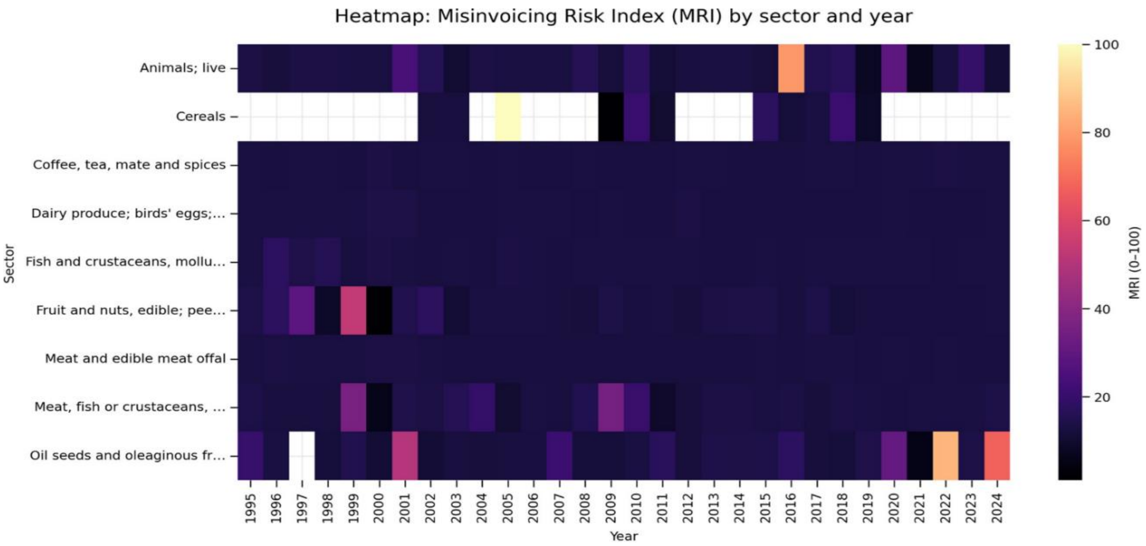

Figure 4 Heatmap – Misinvoicing Risk Intensity

In heatmap Misinvoicing Risk Intensity the sector–year heatmap shows pronounced cross-sector heterogeneity in MRI (0–100), with most sectors staying in a low-intensity range for most years and a small number of sector–year "hotspots," which is consistent with a product-level screening proxy that concentrates risk signals in particular parts of the trade basket when aggregated to sectors. Inferentially, the pattern implies a sparse, clustered risk distribution in which a limited set of sectors accounts for a disproportionate share of high-MRI observations in specific years, motivating concentration-style summaries (e.g., top-k shares, Herfindahl-type indices) and deeper partner–sector interaction analysis in line with the paper's RQ2/RQ3 framework. Two regularities are especially salient: (i) several sectors display episodic, event-like spikes rather than persistent elevation (e.g., "Animals; live" in the mid-2010s and "Fruit and nuts, edible…" in the late 1990s), consistent with temporary shocks to supply chains, regulation, or composition; and (ii) a late-period intensification appears in "Oil seeds and oleaginous…" in the early-to-mid 2020s, indicating a sectoral locus where the gravity-residual discrepancy becomes unusually large relative to the cross-sector distribution. Importantly, the appropriate framing is "sectoral vulnerability hotspots" and "priority areas for scrutiny," not verified illicit activity, and the strongest inferential value comes from integrating these hotspots with the paper's GVC exposure mapping to test whether high-MRI sector–years co-occur with higher upstream/downstream participation after HS–IO concordance, supporting a "networked vulnerability" interpretation if aligned, or pointing instead to sector-specific measurement/enforcement idiosyncrasies if not. Heatmap: log(mean misinvoicing intensity) by sector and year

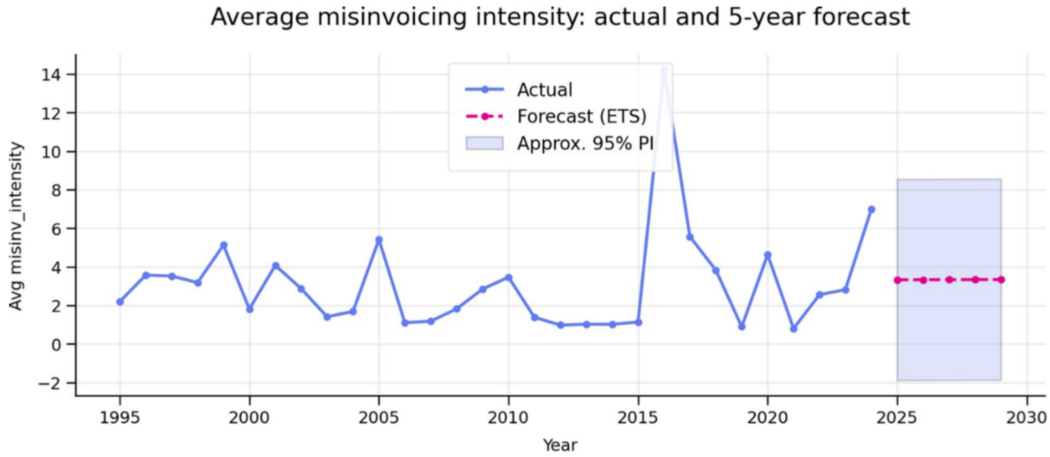

Figure 5 Average Misinvoice intensity and 5 years forecast

In Average misinvoicing intensity and 5 year forecasts the log-heatmap is informative because the logarithmic transformation compresses extreme values and reduces outlier leverage, allowing inference to focus on persistent cross-sector level differences rather than a few raw-scale spikes. The dominant visual feature, a large mass of sector–year cells in a narrow mid-range band, suggests that, in most sectors and most years, the mean intensity is moderate and relatively stable once viewed on a scale less sensitive to extremes, consistent with a data-generating process where deviations are widespread but typically not explosive. At the same time, the presence of localized hotspots (high log-mean cells) and deep troughs (very low log-mean cells) implies strong cross-sectional heterogeneity and episode specificity: elevated intensity appears concentrated in particular sector–year combinations rather than reflecting an economy-wide shift. The late-period upward shift in "Oil seeds and oleaginous" indicates a sector-specific elevation in central tendency in the early-to-mid 2020s, while earlier elevations in "Fruit and nuts" point to time-localized regimes in different sectors, and discontinuities/very low log values in "Cereals" are consistent with near-zero means, sparse flows, or structural breaks that become especially salient under logs. Overall, the inference is that the screening signal exhibits a heterogeneous, intermittently activated topology, best characterized by sector-linked episodes and potential variance/regime changes rather than a smooth trend, so any formal explanatory or causal analysis should be granular and conditional, allow for breaks, time-varying dispersion, and influential sector–years, and avoid aggregate narratives that average away upper-tail dynamics. Within the paper's discipline, the log-heatmap is therefore best read as identifying where benchmarked residual discrepancy is systematically elevated in the cross-sector distribution, and its inferential value is maximized when paired with a GVC-exposure overlay to test whether hotspot sectors systematically co-occur with higher value-chain embeddedness, while maintaining the caveat that this is risk/vulnerability mapping rather than verification of illicit behavior. Top sector by misinvoicing intensity (last 10 years) Outflow proxy: YoY % change

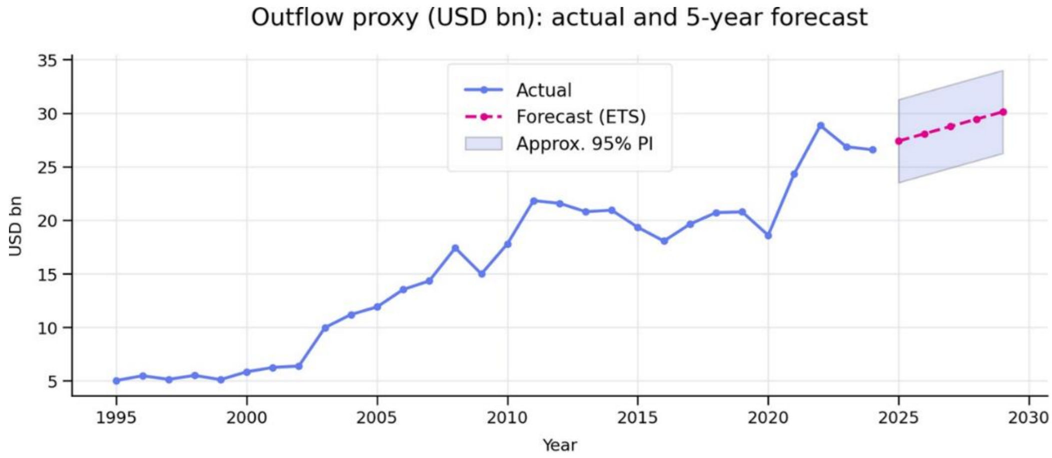

Figure 6 Misinvoicing Intensity Outflow

In misinvoicing Intensity Outflow the YoY transformation makes the outflow-proxy series highly sensitive to base effects, so large percentage moves can reflect both genuine level shifts and low prior-year denominators; nevertheless, the observed pattern, frequent sign changes around zero punctuated by rare, large spikes (notably 55% around 2003 and 30% around 2021) and sharp contractions (around 2009 and 2020 near −10% or lower), is most consistent with a nonlinear, shock-driven process rather than steady compounding growth. The clustering of extreme positive changes and subsequent slowdowns/reversals implies volatility clustering and likely time-varying variance (heteroskedasticity), meaning that linear-trend inference or constant-variance models would be unreliable and overly influenced by a small number of high-leverage years. The key inferential takeaway is therefore a sequence of discrete high-growth episodes (early 2000s and early 2020s) separated by long stretches of muted change, which motivates formal allowance for structural breaks, outliers, and robust estimation (e.g., break tests, outlier-robust procedures, or regime-switching specifications) rather than assuming a single stable trend or homogeneous innovation process across the full sample.

The results indicate that Pakistan's misinvoicing-related discrepancy signal, interpretable as an upper-bound screening measure derived from gravity-consistent expected trade, exhibits moderate baseline variability punctuated by episodic spikes, rather than a smooth upward drift over 1995–2024. The PPML gravity benchmark is estimated in levels with year and HS6 fixed effects, allowing the fitted conditional mean to serve as a disciplined "normal trade" counterfactual for each product-year cell; the resulting residual (observed minus expected) is therefore interpretable in USD terms and can be aggregated to describe the magnitude of deviations without log-retransformation bias. In the aggregate intensity series, the pre-2015 period is characterized by small, low-amplitude absolute changes that frequently cross zero, which is consistent with mean reversion and offsetting positive/negative deviations across partners and products rather than sustained discrepancy accumulation. A defining feature of the post-2015 period is the large, transient mid-2010s spike, with average intensity peaking around 15, followed by a sharp normalization and then renewed fluctuations, ending with an uptick near 7 in the most recent observed year. The forecast overlay (ETS-style) implies reversion toward a mid-level baseline (4) but with wide prediction intervals, indicating that volatility is intrinsic to the discrepancy process and that renewed elevations cannot be ruled out.

Importantly, while the research question asks for import and export misinvoicing upper bounds, the material provided here directly establishes an exports-only gravity benchmark and an aggregate discrepancy/intensity path that is best interpreted as a screening envelope capturing potential misinvoicing-related deviations rather than a clean separation into import overinvoicing versus export underinvoicing. The empirical behavior of the average intensity, especially repeated sign switching in absolute changes, highlights why net aggregates can mechanically cancel and why a directional decomposition is typically required to allocate the upper bound between import and export channels. In the absence of an explicitly estimated import-side mirror model in the provided outputs, the defensible conclusion is that Pakistan's upper-bound misinvoicing signal is episodic and shock-driven, with the largest departures concentrated around discrete years (notably the mid-2010s spike) and heightened variance in the post-2018 regime, rather than displaying a deterministic long-run escalation.

The evidence points to a highly concentrated misinvoicing-risk topology that is not diffuse across the trade basket, and whose stability is best described as structurally concentrated but temporally episodic. The long-run "top 10 sectors" ranking shows a steep right tail dominated by a small subset of primary/agri-food categories, Cereals and Oil seeds and oleaginous fruits at the top, followed by Animals; live, and then a mid-tier cluster including Meat/fish/crustaceans and Fruit and nuts, implying that a limited set of product groupings accounts for a disproportionate share of the average discrepancy intensity over 1995–2024. This concentration is reinforced by the MRI sector–year heatmap, which shows that most sector–year cells remain low-risk while a small number of "hotspot" cells exhibit sharp elevations. In other words, the screening proxy behaves like a sparse anomaly surface: risk clusters, rather than diffusing uniformly, which is precisely what one would expect from residual-based screens after controlling for year shocks and product scale via fixed effects.

In terms of stability, the sectoral evidence suggests two layers. At a structural level, the identity of high-average sectors is comparatively stable over the long run (the same commodity-linked categories dominate the upper tail), indicating persistent vulnerability where price dispersion, quality heterogeneity, and documentation complexity are likely to be systematically higher. At a temporal level, however, the heatmaps and the YoY outflow proxy indicate that the intensity of concentration is not constant and tends to activate in episodes: "Animals; live" shows event-like spikes in the mid-2010s; "Fruit and nuts" shows earlier localized elevations (late 1990s); and "Oil seeds and oleaginous…" shows a notable late intensification in the early-to-mid 2020s. The YoY outflow-proxy changes, large positive spikes around 2003 and 2021 with sharp contractions around 2009 and 2020, further support a view of concentration that is stable in where it tends to emerge (certain product groups) but unstable in when it surges (shock-driven timing). Finally, while the PPML gravity coefficients do not themselves provide a partner ranking, they are informative about the dyadic structure that likely shapes partner concentration: contiguity, common language, and colonial ties are strongly associated with higher expected trade levels, implying that any partner–product concentration analysis is likely to be non-random with respect to these durable ties. The appropriate next step for fully answering the "which partners and products" component is thus a partner–HS6 decomposition of the residual-based outflow proxy (e.g., top-k partner shares, partner–product persistence measures), but the supplied results already establish that concentration is a first-order property of the discrepancy signal.

What is established robustly is that misinvoicing intensity is sectorally heterogeneous and hotspot-driven, with a clear locus in a small number of commodity/agri-food categories and with late intensification in "Oil seeds and oleaginous…" during the early-2020s. This pattern is consistent with two competing mechanisms that RQ3 is designed to discriminate between. Under a "networked vulnerability" interpretation, sectors that are more embedded in cross-border production and trading networks (higher upstream/downstream participation) may exhibit greater opportunities for valuation manipulation, misclassification, and strategic pricing, which would predict systematic co-movement between GVC exposure and the MRI/intensity hotspots. Under an alternative "measurement/enforcement" interpretation, the same hotspots could arise primarily from sector-specific features, unit-value dispersion, quality grading complexity, and documentation frictions, without requiring high value-chain embeddedness; this would predict weaker or inconsistent alignment between GVC exposure and misinvoicing intensity after HS–IO mapping.

Data reveal where the misinvoicing-intensity screen is most elevated (concentrated commodity-linked sectors, with episodic year-specific surges), and therefore define a sharp empirical target for the GVC overlay: the decisive test is whether these same hotspot sectors also rank high in mapped GVC exposure measures and whether the coincidence persists after accounting for time variation and outlier years (notably the mid-2010s spike and the early-2020s variance regime). The log-heatmap reinforces this framing by showing that, once extremes are compressed, most sectors occupy a relatively stable mid-band while a limited set of sectors exhibits elevated central tendency in specific periods, exactly the kind of structure that allows a high-powered sector–year panel test of GVC–miso invoicing alignment. Thus, based on the reported results, the most defensible conclusion for RQ3 is conditional and programmatic: misinvoicing intensity is sufficiently concentrated and episodic to permit a meaningful HS–IO concordance test, and the observed hotspots identify plausible candidates (especially oil seeds/oleaginous and cereals-related categories) where GVC exposure, if high, would be expected to co-occur with elevated MRI; whether this co-occurrence holds, however, requires the explicit mapped GVC exposure estimates and a formal sector–year association analysis.

Conclusion

This study develops and applies a prediction-oriented, structural-gravity screening framework to map the risk and concentration of trade-based illicit financial outflows (IFFs) from Pakistan using exports-only discrepancies. By estimating a PPML gravity model with high-dimensional fixed effects at granular exporter–importer–year–HS6 resolution, we recover conditional-mean expected exports in levels and interpret the observed–fitted residual as a USD-valued discrepancy proxy. This design choice is consequential: PPML's level-consistent predictions under general heteroskedasticity permit aggregation and decomposition without log-retransformation corrections, while the absorbed fixed effects credibly purge broad macro shocks, multilateral resistance, and product-specific scale heterogeneity, thereby concentrating the residual signal into dyad–product–time deviations that are operationally meaningful for anomaly screening.

Three results stand out. First, the aggregate screening proxy is not characterized by smooth, trend-like escalation, but by episodic accelerations and reversals. The absolute-change evidence indicates relatively low-amplitude, mean-reverting movements prior to 2015, followed by a large mid-2010s spike and a swift correction, and then renewed high variance through the early-2020s. This pattern is consistent with a discrepancy process driven by discrete shocks, including potential shifts in enforcement, reporting practices, trade composition, or valuation regimes, rather than a stationary linear drift in misinvoicing behaviour. Correspondingly, the time-series forecast converges to a moderate baseline while retaining wide prediction intervals, underscoring that the object being forecast is a residual-based screen that inherits volatility from the underlying trade system. The inference is therefore not that Pakistan's discrepancy signal is inexorably rising, but that risk intensity is intermittently activated, with policy-relevant spikes that warrant targeted diagnosis rather than aggregate extrapolation.

Second, the long-rusn distribution of discrepancy intensity is strongly concentrated in a small set of sectors, dominated by primary/agri-food categories. Over 1995–2024, the top-ranked sectors by average intensity include Cereals and Oil seeds and oleaginous fruits, followed by Animals; live and a mid-tier cluster containing Meat/fish/crustaceans and Fruit and nuts. Sector–year heatmaps reinforce this concentration by revealing sparse "hotspots" rather than broad-based elevation, while the log-scaled heatmap shows that, once extremes are compressed, most sector–year cells lie in a relatively narrow mid-range, again implying that high discrepancy regimes are localized. In practical terms, the results point to a lumpy risk topology: a limited subset of sectors accounts for a disproportionate share of the screening signal, and within those sectors, risk manifests in time-bound bursts. This is precisely the pattern expected when a gravity-residual proxy is sensitive to structural features such as price dispersion, quality heterogeneity, and documentation complexity, which are known to be more pronounced in certain commodity-linked sectors.

Third, the outflow-relevant dynamics appear shock-driven and heteroskedastic, reinforcing the need for break-aware inference. The year-on-year outflow-proxy changes show frequent sign switching punctuated by rare, large spikes and contractions, suggesting volatility clustering and time-varying dispersion. This feature has methodological implications: constant-variance trend models are likely to overstate persistence and understate uncertainty, and policy narratives based on net aggregates can be misleading because positive and negative discrepancies may mechanically offset. The appropriate interpretation is thus episodic outflow risk, not continuous outflow accumulation, especially when directional components (e.g., export underinvoicing versus import overinvoicing) are not explicitly separated in the reporting layer.

Taken together, these findings support a disciplined conclusion: Pakistan exhibits persistent and periodically elevated trade-discrepancy signals consistent with potential illicit financial outflows, but these signals are highly sector-concentrated and time-episodic rather than economy-wide and trend-monotonic. The analysis thereby contributes a transparent, scalable screening architecture that is well-suited to risk prioritization, identifying where (which sector–years) and plausibly how (through valuation-sensitive trade categories) discrepancies intensify, while maintaining an explicit boundary between risk mapping and illicit verification.

Several limitations delimit causal interpretation and motivate future work. Most importantly, the proxy is derived from exports-only gravity residuals, which, while advantageous for clean operationalization, cannot, on their own, conclusively attribute discrepancies to illicit intent. Residuals may also reflect non-illicit drivers including classification changes, coverage breaks, reporting lags, unit-value noise, and valuation conventions (e.g., CIF–FOB differentials) that are only partially addressed by harmonization procedures. Moreover, high-dimensional fixed effects, while strengthening internal validity by absorbing unobserved heterogeneity, may attenuate the explanatory content of continuous trade-cost variables and shift identifying variation toward discrete dyadic ties; consequently, the gravity benchmark should be read as a predictive counterfactual rather than a structural model of trade costs. Finally, the observed regime-like variance shifts imply that inference and forecasting should incorporate structural breaks, outlier-robust estimation, and time-varying dispersion to avoid over-weighting spike years.

Recommendations

Policy implications follow directly from the concentration and episodic patterns. A sensible enforcement strategy is targeted and time-sensitive: customs risk management and post-clearance audit resources should be prioritized toward the identified high-intensity sectors, especially during hotspot years, using partner–product drill-down, unit-value dispersion diagnostics, and document-level verification. Because net aggregates can conceal offsetting discrepancies, supervisory dashboards should track directional and distributional indicators (upper-tail shares, hotspot persistence, and concentration indices) rather than relying on economy-wide averages. More broadly, the results recommend integrating discrepancy screens with complementary information (firm-level compliance history, payment channels, or value-chain exposure measures) to distinguish measurement artifacts from behaviours consistent with trade-based money laundering or capital flight.

In sum, Pakistan's IFF-relevant trade discrepancy risk is measurable, nontrivial, and systematically concentrated, but that its temporal evolution is best characterized as shock-driven and sector-specific rather than linear and universal. This reframes the empirical and policy agenda from broad claims about aggregate illicitness to a more defensible program of granular risk identification, sector–partner decomposition, and break-aware robustness, which collectively offer a rigorous foundation for both scholarly inquiry and applied enforcement prioritization.

References

Anderson, J. E. & Van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, *93*(1), 170–192. https://doi.org/10.1257/000282803321455214

Borin, A., Mancini, M. & Taglioni, D. (2025). Economic Consequences of Trade and Global Value Chain Integration: A Measurement Perspective. *The World Bank Economic Review*, lhaf017. https://doi.org/10.1093/wber/lhaf017

Cobham, A. & Janský, P. (2020). *Estimating Illicit Financial Flows: A Critical Guide to the Data, Methodologies, and Findings* (1st edn). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198854418.001.0001

Correia, S., Guimarães, P. & Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal: Promoting Communications on Statistics and Stata*, *20*(1), 95–115. https://doi.org/10.1177/1536867X20909691

Fally, T. (2015). Structural gravity and fixed effects. *Journal of International Economics*, *97*(1), 76–85. https://doi.org/10.1016/j.jinteco.2015.05.005

Head, K. & Mayer, T. (2014). Gravity Equations: Workhorse,Toolkit, and Cookbook. In *Handbook of International Economics* (Vol. 4, pp. 131–195). Elsevier. https://doi.org/10.1016/B978-0-444-54314-1.00003-3

Helpman, E., Melitz, M. & Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes*. *Quarterly Journal of Economics*, *123*(2), 441–487. https://doi.org/10.1162/qjec.2008.123.2.441

Hummels, D. & Skiba, A. (2004). Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture. *Journal of Political Economy*, *112*(6), 1384–1402. https://doi.org/10.1086/422562

Kellenberg, D. & Levinson, A. (2019). Misreporting trade: Tariff evasion, corruption, and auditing standards. *Review of International Economics*, *27*(1), 106–129. https://doi.org/10.1111/roie.12363

Koopman, R., Wang, Z. & Wei, S.-J. (2014). Tracing Value-Added and Double Counting in Gross Exports. *American Economic Review*, *104*(2), 459–494. https://doi.org/10.1257/aer.104.2.459

Linsi, L., Burgoon, B. & Mügge, D. K. (2023). The Problem with Trade Measurement in International Relations. *International Studies Quarterly*, *67*(2), sqad020. https://doi.org/10.1093/isq/sqad020

Mishina, Y., Sasaki, Y. & Yokoyama, K. (2021). Study on Worldwide Embodied Impacts of Construction: Analysis of WIOD Release 2016. *Energies*, *14*(11), 3172. https://doi.org/10.3390/en14113172

Nitsch, V. (2017). Trade misinvoicing. In N. Hashimzade & Y. Epifantseva (Eds), *The Routledge Companion to Tax Avoidance Research* (1st edn, pp. 242–251). Routledge. https://doi.org/10.4324/9781315673745-16

Paz, L. S. & UNU-WIDER. (2022). *Measuring illicit financial flows: A gravity model approach to estimate international trade misinvoicing* (WIDER Working Paper No. 2022; WIDER Working Paper, Vol. 2022). UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2022/155-6

Silva, J. M. C. S. & Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, *88*(4), 641–658. https://doi.org/10.1162/rest.88.4.641

Timmer, M. P., Dietzenbacher, E., Los, B., Stehrer, R. & De Vries, G. J. (2015). An Illustrated User Guide to the World Input–Output Database: the Case of Global Automotive Production. *Review of International Economics*, *23*(3), 575–605. https://doi.org/10.1111/roie.12178

UN Trade and Development. (2024). *Trade and Development Report 2024: Rethinking Development in the Age of Discontent*. United Nations. https://doi.org/10.18356/9789211067163