CGPA WORKING PAPER
2026-11

Instrumental-Variable Poisson PML with High-Dimensional Fixed Effects

Ohyun Kwon
Mario Larch
Jangsu Yoon
Yoto V. Yotov

DREXEL UNIVERSITY
Center for
Global Policy Analysis
LeBow College of Business

Instrumental-Variable Poisson PML with High-Dimensional Fixed Effects*

Ohyun Kwon Mario Larch Jangsu Yoon Yoto V. Yotov[†]

This version: April 24, 2026

Abstract

We implement an instrumental-variable Poisson pseudo-maximum likelihood estimator with high-dimensional fixed effects (IV-PPML-HDFE). To correct for incidental parameter bias, we use a split-panel jackknife (SPJ) routine with bootstrapped standard errors. Monte Carlo simulations across the three most common fixed-effect structures confirm that SPJ reduces the mean absolute bias by 42% and raises mean bootstrap confidence-interval coverage from 69% to 92%. We provide a robust and user-friendly `ivppmlhdfe` package, and deploy it in three empirical applications to establish the validity and usefulness of our methods.

JEL Codes: C13, C23, C26, F14.

Keywords: Poisson pseudo-maximum likelihood, instrumental variables, high-dimensional fixed effects, incidental parameter problem, gravity model, split-panel jackknife.

*Acknowledgments will be added later. During the preparation of this paper, the authors used Claude Code to assist with replication code development, data verification, and manuscript editing. After using this tool, the authors reviewed and edited the content as needed. The authors take full responsibility for the content of the article and for any errors.

[†]Contact information: Ohyun Kwon - Contact author. E-mail: theekwonomist@gmail.com; Mario Larch - Email: mario.larch@uni-bayreuth.de; Jangsu Yoon - Email: jangsu.yoon@uky.edu; Yoto V. Yotov - Email: yvy23@drexel.edu.

1 Introduction

The Poisson pseudo-maximum likelihood (PPML) estimator has become one of the most widely used methods in applied economics (Santos Silva and Tenreyro, 2022). Three of the key properties that contributed to the remarkable success of the PPML estimator include (i) its ability to address the heteroskedasticity-related issues by avoiding to log-linearize the model, (ii) the fact that its consistency requires only correct specification of the conditional mean, making it applicable to any non-negative outcome, and (iii) its ability to account for zero outcomes (Santos Silva and Tenreyro, 2006). As a result, the PPML estimator has gained popularity in various disciplines, e.g., economics (Santos Silva and Tenreyro, 2022; Chen and Roth, 2024), finance (Karolyi and Taboada, 2015; Cohn et al., 2022), and political science (Kinne, 2013; Morin et al., 2025).

Perhaps most prominently, the PPML estimator established itself as the leading estimator for 'gravity' equations in international trade (Santos Silva and Tenreyro, 2006; Larch et al., 2025). However, it has also been used for many gravity-type applications beyond trade, e.g., for foreign direct investment (Bergstrand and Egger, 2007; Larch and Yotov, 2025), migration (Poot et al., 2016; Beverelli, 2021), commuting (Persyn and Torfs, 2015; Dingel and Tintelnot, 2026), cross-border patent transfer (LaBelle et al., 2023), financial-asset flows (Mercado, 2022), greenfield investment (Carril-Caccia et al., 2023), mergers and acquisitions (Bradley et al., 2023), etc. The common practice for estimating such bilateral economic relationships in panel settings is that, in addition to using source-time and destination-time fixed effects, the econometric specifications often include bilateral/pair fixed effects as well (Larch et al., 2025). However, the use of such high-dimensional fixed effects structures poses two broad challenges related to (i) the incidental parameter problem (IPP) and (ii) computation and convergence issues.

To address the computation and convergence challenges, the original `ppml` command of Santos Silva and Tenreyro (2006) was extended and improved in terms of speed and convergence robustness by the `ppml_panel_sg` command of Larch et al. (2018) and, more recently, by the `ppmlhdfe` command of Correia et al. (2020). The latter quickly established itself as the best package for PPML estimations. On the IPP front, Fernández-Val and Weidner (2016) demonstrate that the PPML estimator with two-dimensional fixed effects (e.g., source-time and destination-time fixed effects) is immune to the incidental parameter problem. Extending the analysis to a panel setting with three-way fixed effects (e.g., source-time, destination-time, and pair fixed effects), Weidner and Zylkin (2021) demonstrate that the PPML estimates may suffer from IPP bias, and they propose and implement a bias correction procedure – `ppml_fe_bias`.

In parallel with these advancements, significant progress was made to develop instrumental-variable PPML (IV-PPML) procedures that address endogeneity in PPML regressions. For example, the approach of Windmeijer and Santos Silva (1997) to replace the endogenous regressors in the PPML score with excluded instruments while maintaining the exponential conditional mean has been successfully applied in a series of applications, e.g., Tenreyro (2007); Egger et al. (2011); Karolyi and Taboada (2015).[1] More recently, Jochmans and Verardi (2022) develop procedures to address endogeneity and IPP issues in a cross-section PPML setting with two-way (source and destination) fixed effects.[2] However, the existing literature remains largely limited to cross-sectional settings and does not provide a practical framework for handling IPP and estimation in panel models with three-way fixed effects, such as source-time, destination-time, and pair effects. In particular, although the IV-PPML mo-

[1]An alternative IV approach for count data with multiplicative unobserved heterogeneity is the transformation estimator of Mullahy (1997). The two estimators are not interchangeable: Mullahy's transformation is designed for multiplicative heterogeneity, while the additive moment of Windmeijer and Santos Silva (1997) is appropriate when the conditional mean is correctly specified after absorbing fixed effects, as in the panel and bilateral settings that we consider.

[2]The IPP arises in IV-PPML settings because replacing the endogenous regressors with instruments breaks the Poisson 'score cancellation' that eliminates leading-order IPP bias in standard/exogenous PPML settings, making bias correction necessary for the IV case even when it is unnecessary for the exogenous PPML case. Jochmans and Verardi (2022) address this problem by constructing moment conditions that difference out the fixed effects entirely (thus avoiding the IPP altogether). Lin and Wooldridge (2019) propose an alternative control function (CF) approach, which includes first-stage residuals as additional regressors in the Poisson FE regression rather than replacing the moment conditions. While CF avoids the IPP by not altering the score structure, it requires correct specification of the first stage and additive separability of the first-stage residual (control function) inside the exponential mean, whereas IV-PPML only requires instrument relevance and exogeneity.

ment conditions can in principle be carried over to such environments, estimation becomes substantially more difficult once high-dimensional fixed effects are introduced, both computationally and in terms of numerical convergence. Our IV-PPML-HDFE procedure is designed to address precisely these challenges.

Against this backdrop, we make two related contributions. First, we implement an instrumental-variable Poisson pseudo-maximum likelihood estimator with high-dimensional fixed effects (IV-PPML-HDFE) following the approach by Windmeijer and Santos Silva (1997). To mitigate the incidental parameter bias, we utilize the split-panel jackknife (SPJ) routine. Second, we provide robust and user-friendly `ivppmlhdfe` commands for practical implementation in Stata and Julia, and we deploy them in three empirical applications to establish the validity and usefulness of the proposed methods.

The methodological analysis focuses on the three most common panel fixed effects structures (i.e., individual + time, source-time + destination-time, and source-time + destination-time + pair), and we develop the IV-PPML-HDFE estimator in three steps. First, we demonstrate that, unlike the standard/exogenous PPML case, the IV-PPML-HDFE estimator suffers from IPP bias across all fixed effects structures that we consider. The reason is that the Poisson 'score cancellation' of the IPP bias in the standard PPML case does not extend to the IV-PPML-HDFE case. Second, to correct for the IPP bias in IV-PPML-HDFE, we capitalize on and extend the methods of Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021) and propose a split-panel jackknife routine that corrects for the leading-order bias by estimating on sub-panels that each eliminate one source of bias. Third, since the SPJ does not produce standard errors, we implement a bootstrap method that delivers standard errors that are consistent with the SPJ point estimates. This analysis adopts a large-N, large-T asymptotic framework developed in Section 2, under which the SPJ eliminates the leading-order incidental-parameter bias.

We demonstrate through Monte Carlo simulations that, when endogeneity is unaddressed, PPML is severely biased and its confidence intervals (CIs) are unreliable. `ivppmlhdfe` removes this endogeneity bias across the three most common panel structures, but remains subject to a residual IPP bias, and conventional confidence intervals do not deliver 95% coverage. The SPJ bias correction paired with pair bootstrap standard errors reduces the mean absolute bias by 42% and raises mean bootstrap CI coverage from 69% to 92%.

To facilitate the practical application of the IV-PPML-HDFE estimator, we develop `ivppmlhdfe` – a Stata command with a companion Julia package that can be called directly or through a Stata bridge. To this end, we rely on the latest practical applications of the IV and PPML estimators. Specifically, `ivppmlhdfe` follows the iteratively reweighted least squares (IRLS) approach of Correia et al. (2020) to estimate IV-PPML with high-dimensional fixed effects. As a result, `ivppmlhdfe` matches the computational and functional capabilities of `ppmlhdfe`. To make `ivppmlhdfe` familiar and user-friendly, its syntax combines the intuitive conventions and syntax structures of `ppmlhdfe` and `ivreg2`.[3] To ensure reliability across separation, collinearity, convergence edge cases, etc., we performed extensive stress tests of `ivppmlhdfe` against the `ppmlhdfe` and `ivreg2` benchmarks.

We apply `ivppmlhdfe` to three published studies with different high-dimensional fixed-effect specifications. In Chen et al. (2025), where the authors use a control-function approach, the `ivppmlhdfe` estimate is very close to the published estimate, and the SPJ version is almost identical. In Chang et al. (2022), `ivppmlhdfe` exactly reproduces the published just-identified estimates and substantially reduces computation time from 18 minutes to less than 1 second. In Curzi and Huysmans (2022), the two-step plug-in procedure used in the original paper attenuates coefficients by a factor of two to four relative to the joint `ivppmlhdfe` estimator, but the authors' main conclusion remains unchanged.

The rest of the paper is organized as follows. Section 2 introduces the IV-PPML-HDFE estimator, documents the incidental parameter problem under IV-PPML, develops the split-panel jackknife with bootstrap bias correction, and introduces our `ivppmlhdfe` package.

[3] The `ivppmlhdfe` package is distributed through the `ekwonomist/ivppmlhdfe` public GitHub repository, and it can be installed directly from Stata as follows: `net install ivppmlhdfe, from("https://raw.githubusercontent.com/ekwonomist/ivppmlhdfe/main/")`.

Section 3 reports findings from a Monte Carlo analysis. Section 4 presents results and comparisons from three empirical applications. Section 5 concludes with a summary of our contributions and a discussion of possible caveats and future work.

2 The IV-PPML-HDFE Estimator

This section develops the IV-PPML-HDFE estimator. Section 2.1 begins with the standard PPML moment conditions, shows how they extend to the IV case, and describes the algorithm used to concentrate out high-dimensional fixed effects. Section 2.2 demonstrates that IV-PPML suffers incidental parameter biases and characterizes them across the three fixed-effect cases. Section 2.3 describes the proposed split-panel jackknife with bootstrap procedures to correct for the biases. Finally, Section 2.4 introduces our `ivppmlhdfe` package.

2.1 From PPML-HDFE to IV-PPML-HDFE

The observations are indexed by $g = 1, \dots, n$, where g is a generic index that encompasses both standard panel structures ($g = (i, t)$, $n = NT$) and bilateral (gravity-type) panel structures ($g = (i, j, t)$, $n = N_c(N_c - 1)T$). Each observation consists of a non-negative outcome $y_g \geq 0$, a vector of exogenous covariates $x_g \in \mathbb{R}^K$, a vector of endogenous regressors $d_g \in \mathbb{R}^\ell$, and a vector of excluded instruments $z_g \in \mathbb{R}^\ell$. In the theory, we focus on the just-identified case, which is the most common setting in applied work. Since the just-identified case with vector-valued d_g and z_g does not introduce the additional weighting-matrix issues that arise under overidentification, we may, without loss of generality and to simplify notation, restrict attention in what follows to the scalar case $\ell = 1$.[4]

[4] The theoretical analysis focuses on the just-identified case. Extensions to the overidentified case require accounting for the additional bias induced by estimating the GMM weighting matrix; see Fernández-Val and Lee (2013). The `ivppmlhdfe` command includes an implementation for overidentified specifications based on iteratively reweighted two-stage least squares, analogous to `ivreg2`. However, the theoretical analysis in this paper is restricted to the just-identified case.

Let $w_g = \left(x_g', d_g\right)'$ collect all regressors with conformable slope coefficients β, and let ψ_g denote the sum of fixed effects relevant to observation g. The three fixed-effect structures that we consider are those most common in the empirical literature and they vary by model class as follows:

$$\begin{aligned}
\text{Class A:} \quad & \psi_g = \alpha_i + \gamma_t, \\
\text{Class B:} \quad & \psi_g = \alpha_{it} + \gamma_{jt}, \\
\text{Class C:} \quad & \psi_g = \alpha_{it} + \gamma_{jt} + \eta_{ij}.
\end{aligned}$$

Class A corresponds to the common panel-data setting with individual and time fixed effects, whereas Classes B and C apply to bilateral panel data (e.g., trade, FDI, migration, etc.), where each observation is indexed by a directed pair of units and time. We maintain the exponential conditional mean assumption:

$$\mathbb{E}\left[y_g \mid w_g, \psi_g\right] = \mu_g \equiv \exp\left(w_g'\beta + \psi_g\right), \tag{1}$$

and define the residual $u_g \equiv y_g - \mu_g$.

The PPML estimator of Santos Silva and Tenreyro (2006) maximizes the Poisson pseudo-log-likelihood $\sum_g \left[y_g \left(w_g'\beta + \psi_g\right) - \exp\left(w_g'\beta + \psi_g\right)\right]$, and the first-order conditions yield moment conditions of the form

$$\frac{1}{n}\sum_g w_g u_g = 0 \quad \text{and} \quad \sum_{g \in r} u_g = 0 \ \ \forall r, \tag{2}$$

where the first set determines β and the second pins down the fixed effects ψ_r within each group r. This is a just-identified GMM system in which the regressors w_g serve as their own instruments.

When d_g is endogenous, $\mathbb{E}\left[d_g u_g\right] \neq 0$ and these moment conditions are violated. Following Windmeijer and Santos Silva (1997), the IV-PPML approach that we adopt replaces the endogenous regressor in the slope moments with an excluded instrument while leaving the fixed-effect equations unchanged. Defining $q_g = \left(x_g', z_g\right)'$, the IV-PPML estimator solves

$$\frac{1}{n}\sum_{g} q_g u_g = 0 \quad \text{and} \quad \sum_{g\in r} u_g = 0 \ \ \forall r. \tag{3}$$

The identifying assumption is $\mathbb{E}\left[q_g u_g\right] = 0$: the instrument z_g is uncorrelated with the structural residual, and the exogenous covariates and fixed-effect group indicators remain valid moment conditions as in PPML. Thus, the key observation is that the transition from PPML to IV-PPML changes only the slope moments. The fixed-effect equations, the exponential mean specification, and the concentration-out procedure all remain identical.

To solve system (3), we apply the IRLS approach of Correia et al. (2020) to the IV setting. The algorithm proceeds as follows. At iteration k, given current estimates $\left(\beta^{(k)}, \psi^{(k)}\right)$ and fitted values $\hat{\mu}_g^{(k)} = \exp\left(w_g'\beta^{(k)} + \psi_g^{(k)}\right)$, we linearize the exponential mean around the current estimates, forming a pseudo-dependent variable

$$\tilde{y}_g^{(k)} = w_g'\beta^{(k)} + \psi_g^{(k)} + \frac{y_g - \hat{\mu}_g^{(k)}}{\hat{\mu}_g^{(k)}} \tag{4}$$

with observation weights $\omega_g^{(k)} = \hat{\mu}_g^{(k)}$. We then run weighted two-stage least squares of $\tilde{y}_g$ on w_g using instruments q_g, absorbing fixed effects via the iterative demeaning procedure of Correia et al. (2020). The resulting coefficient estimates provide $\beta^{(k+1)}$, and the fixed effects $\psi^{(k+1)}$ are recovered from the demeaning step. The procedure iterates until convergence, $\|\beta^{(k+1)} - \beta^{(k)}\| < \varepsilon$. For any given β, the fixed-effect equations $\sum_{g\in r} u_g = 0$ have a unique solution provided each group r contains at least one observation with $y_g > 0$. This follows from the strict concavity of the Poisson pseudo-log-likelihood in ψ for fixed β. Groups in which all outcomes are zero are dropped prior to estimation, following standard practice.

Table 1 summarizes the relationship between PPML and IV-PPML. The two estimators share the same exponential mean, the same fixed-effect equations, and the same computational approach to absorbing high-dimensional fixed effects (HDFE). They differ only in the

Table 1: PPML vs. IV-PPML

	PPML	IV-PPML
Slope moment condition	$\mathbb{E}\left[w_g u_g\right] = 0$	$\mathbb{E}\left[q_g u_g\right] = 0$
Fixed-effect condition	$\sum_{g \in r} u_g = 0$	$\sum_{g \in r} u_g = 0$
Instrument for β	w_g (regressors)	$q_g = \left(x_g', z_g'\right)'$

slope moments: PPML uses the regressors w_g as instruments for themselves, while IV-PPML substitutes q_g. This substitution, however, has an important consequence for finite-sample bias. As we show in Section 2.2, the Poisson score cancellation that eliminates leading-order incidental parameter bias in PPML does not carry over when w_g is replaced by q_g.

2.2 IV-PPML(-HDFE) and the Incidental Parameter Problem

In the standard PPML estimator, the score for β is $\sum_g w_g \left(y_g - \mu_g\right)$, where the same regressors w_g appear in both the score weights and the conditional mean $\mu_g = \exp\left(w_g'\beta + \psi_g\right)$. This alignment produces a 'score cancellation', which implies that the leading-order incidental parameter bias in PPML is effectively zero (Fernández-Val and Weidner, 2016; Weidner and Zylkin, 2021). Applied to our analysis, this means that, for Classes A and B, PPML is immune to first-order IPP bias. For Class C, however, Weidner and Zylkin (2021) show that the Poisson 'score cancellation' eliminates the $O\left(1/T\right)$ pair-effect contribution, leaving a residual bias of order $O\left(1/N_c\right)$ from the exporter-time and importer-time fixed effects.

The Poisson 'score cancellation' does not apply to the IV-PPML case. The reason is that, in IV-PPML, the moment condition replaces w_g with instruments $q_g \neq w_g$. Thus, because the moment weights q_g no longer coincide with the regressors inside μ_g, the cancellation fails and the IPP bias remains in each IV-PPML model class, even when such correction is not needed for PPML. We characterize the bias order for each of the three fixed-effect structures in Table 2, and we summarize them below. See Appendix A for the derivation of IPP bias orders across models.

Table 2: IPP bias orders and remedies by model class

	Class A	**Class B**	**Class C**
FE structure	$\alpha_i + \gamma_t$	$\alpha_{it} + \gamma_{jt}$	$\alpha_{it} + \gamma_{jt} + \eta_{ij}$
IPP bias (PPML)	0	0	$O(1/N_c)$
IPP bias (IV-PPML)	$O(1/T + 1/N)$	$O(1/N_c)$	$O(1/N_c + 1/T)$
SE bias	$O(1/T + 1/N)$	$O(1/N_c)$	$O(1/N_c + 1/T)$

Class A: Individual + time FE. With fixed effects $\psi_g = \alpha_i + \gamma_t$, estimation noise in both $\hat{\alpha}_i$ and $\hat{\gamma}_t$ contributes to the bias. The leading-order bias is $O(1/T + 1/N)$, vanishing as both dimensions grow (Fernández-Val and Weidner, 2016).

Class B: Two-way 'gravity' FE. With directional fixed effects $\psi_g = \alpha_{it} + \gamma_{jt}$, the bias arises from noise in $\hat{\alpha}_{it}$ and $\hat{\gamma}_{jt}$. The leading-order bias is $O(1/N_c)$, where N_c is the number of countries (Weidner and Zylkin, 2021).

Class C: Three-way 'gravity' FE. Adding pair fixed effects η_{ij} to the Class B specification introduces a second bias channel from the $O(1/T)$ noise in $\hat{\eta}_{ij}$. The total leading-order bias is $O(1/N_c + 1/T)$, and coverage can deteriorate sharply when both N_c and T are moderate.

In sum, the analysis in this section demonstrates that the incidental parameter problem is present in all IV-PPML settings. Because the bias expansions for Classes A and C contain a $1/T$ term, the IV-PPML-HDFE estimator is inconsistent under fixed-T asymptotics with $N \to \infty$ in those cases, and even under the large-N, large-T framework adopted here the leading-order bias remains non-negligible relative to the usual normalization. Bias corrections are therefore necessary for IV-PPML even when they are not needed for PPML. We propose such corrections next.

2.3 Split-Panel Jackknife with Bootstrap Bias Correction

The SPJ removes leading-order IPP bias by estimating on sub-panels that each eliminate one source of bias, then combining via inclusion-exclusion. The idea, developed by Fernández-Val and Weidner (2016) for two-way panel models and adapted by Weidner and Zylkin (2021) for three-way gravity, exploits the additive structure of the bias. If the bias takes the form $B/T + D/N$, then estimating on a sub-panel that halves T doubles the B/T component while leaving D/N unchanged, and vice versa. Appropriate linear combinations of the full-sample and subsample estimators cancel both bias terms simultaneously. Capitalizing on and extending the procedures from Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021), we propose the following SPJ formulas to account for the biases in each of the three fixed-effect structures, as summarized in Table 2:[5]

Class A. The bias has two components, $O(1/T)$ from individual fixed effects and $O(1/N)$ from time fixed effects. Following Fernández-Val and Weidner (2016), we split along both dimensions:

$$\hat{\beta}_{\text{SPJ}} = 3\hat{\beta} - \bar{\beta}_{T/2} - \bar{\beta}_{N/2},$$

where $\bar{\beta}_{T/2}$ averages the estimator over two temporal halves and $\bar{\beta}_{N/2}$ averages over two cross-sectional halves.

Class B. The bias is $O(1/N_c)$ from noise in the directional fixed effects. Following Weidner and Zylkin (2021), we randomly partition countries into two groups a and b and form four directed sub-panels $(a \to a, a \to b, b \to a, b \to b)$:

$$\hat{\beta}_{\text{SPJ}} = 2\hat{\beta} - \tfrac{1}{4}\left(\hat{\beta}_{aa} + \hat{\beta}_{ab} + \hat{\beta}_{ba} + \hat{\beta}_{bb}\right).$$

[5] The detailed derivations are included in Appendix B.

Class C. The bias has two components, $O(1/N_c)$ from directional effects and $O(1/T)$ from pair effects. We propose an 8-panel extension that crosses the country split with a time split:

$$\hat{\beta}_{\text{SPJ}} = 4\hat{\beta} - 2\bar{\beta}_{\text{country}} - 2\bar{\beta}_{\text{time}} + \bar{\beta}_{\text{8cell}},$$

where $\bar{\beta}_{\text{country}}$ averages over the four country sub-panels, $\bar{\beta}_{\text{time}}$ averages over the two temporal halves, and $\bar{\beta}_{\text{8cell}}$ averages over all $4 \times 2 = 8$ cells. This removes both bias terms to leading order.

While the SPJ reduces the leading-order bias in the point estimate, it does not produce standard errors. To obtain standard errors that are consistent with the SPJ estimator, we use a bootstrap method. Specifically, on each bootstrap draw, we re-estimate the full SPJ procedure, including all sub-panel estimations, and construct confidence intervals from the bootstrap distribution. The resampling unit varies by model class. For Class A, we resample individuals with replacement, retaining all T periods for each sampled individual. For Classes B and C, we resample directed country pairs (i, j) with replacement, retaining all T periods for each sampled pair. In all cases, the bootstrap preserves the within-unit dependence structure while generating variation in the cross-sectional composition.

2.4 Practical Implementation: The `ivppmlhdfe` Package

Consistent with our objectives and the methodological development of the IV-PPML-HDFE estimator, we rely on the latest practical applications of the IV and PPML estimators to build our `ivppmlhdfe` command in Stata, with a companion native Julia implementation that can be called directly or through a Stata bridge. Specifically, `ivppmlhdfe` capitalizes on the excellent computational and functional capabilities of `ppmlhdfe`, and its syntax combines the intuitive conventions and syntax structures of `ppmlhdfe` and `ivreg2` as follows:

```
ivppmlhdfe depvar [exogvars] (endogvar = instruments) [if] [in] [weight],
absorb(fe_list) [vce(type) options]
```

The `ivppmlhdfe` package is distributed through the `ekwonomist/ivppmlhdfe` public GitHub repository, and can be installed directly from Stata via

```
net install ivppmlhdfe, from("https://raw.githubusercontent.com/ekwonomist/ivppmlhdfe/main/")
```

To ensure reliability across separation, collinearity, convergence edge cases, etc., we performed extensive stress tests of `ivppmlhdfe` against the `ppmlhdfe` and `ivreg2` benchmarks. Further details on the implementation and results from these robustness experiments are summarized in Appendix D. Finally, we refer the reader to the `ekwonomist/ivppmlhdfe` GitHub repository for details, dependencies on existing packages, and example do-files.

3 Monte Carlo Simulations

Our Monte Carlo exercise has three objectives. First, we establish that `ivppmlhdfe` removes the endogeneity bias when a valid instrument is available. Second, we document that `ivppmlhdfe`, just as `ppmlhdfe`, remains subject to the IPP. Third, we show that the split-panel jackknife combined with a pair bootstrap attenuates the residual point bias and delivers confidence-interval coverage close to the 95% nominal level.

3.1 The Data Generating Process

We generate data from the exponential mean model $y_g = \exp\left(w_g'\beta + \psi_g\right) v_g$, where ψ_g collects class-specific fixed effects and v_g is a multiplicative error. Endogeneity arises because the first-stage error e_g in $x_{1g} = \pi_z z_g + e_g$ also enters $\ln v_g$, so $\mathrm{Cov}\left(x_1, v\right) \neq 0$. The instrument $z_g \sim$ i.i.d. $\mathcal{N}\left(0, 1\right)$ is independent of all errors and satisfies both relevance ($\pi_z = 0.8$) and exogeneity. We set $\beta_1 = 0.5$ and $\rho_{ev} = 0.5$ (moderate endogeneity), and run $R = 1{,}000$ replications per cell with $B = 1{,}000$ pair-bootstrap draws per replication, across the three model classes, varying N and T for Class A and N_c and T for Classes B and C. Full details of the DGP, parameter values, and simulation grids are in Appendix C.

3.2 IVPPMLHDFE Results

We compare three scenarios: (i) `ivppmlhdfe` applied to a DGP with endogeneity ($\rho_{ev} = 0.5$); (ii) `ppmlhdfe` applied to the endogenous DGP without IV correction; and (iii) `ppmlhdfe` applied to the same DGP with no endogeneity ($\rho_{ev} = 0$). Table 3 reports results for $T = 20$. Full results for $T = 10$ and $T = 30$ appear in Appendix C.

Table 3: Monte Carlo: `ivppmlhdfe` vs. `ppmlhdfe` ($T = 20$, $\beta_1 = 0.5$)

				`ivppmlhdfe`			`ppmlhdfe` (w/ endog.)			`ppmlhdfe` (no endog.)		
Class	FE structure	N	T	Bias	SE/SD	Cov	Bias	SE/SD	Cov	Bias	SE/SD	Cov
A	$\alpha_i + \gamma_t$	50	20	−.035	0.70	82.7	.446	0.58	0.0	−.005	0.71	84.6
		100	20	−.025	0.70	82.5	.448	0.59	0.0	−.005	0.75	86.0
		200	20	−.025	0.77	86.0	.445	0.61	0.0	−.006	0.78	88.1
		500	20	−.016	0.79	87.5	.453	0.65	0.0	−.000	0.81	89.4
B	$\alpha_{it} + \gamma_{jt}$	20	20	−.030	0.63	70.3	.451	0.57	0.0	−.002	0.66	80.7
		30	20	−.024	0.67	72.9	.448	0.55	0.0	−.003	0.69	81.1
		40	20	−.021	0.69	72.8	.449	0.60	0.0	−.002	0.76	86.4
		50	20	−.017	0.73	74.4	.450	0.60	0.0	−.001	0.79	87.9
C	$\alpha_{it} + \gamma_{jt} + \eta_{ij}$	20	20	−.036	0.61	60.1	.446	0.54	0.0	−.007	0.63	76.1
		30	20	−.034	0.60	49.6	.445	0.54	0.0	−.006	0.63	74.4
		40	20	−.031	0.64	43.1	.443	0.55	0.0	−.006	0.65	77.0
		50	20	−.030	0.66	36.6	.444	0.55	0.0	−.005	0.69	77.4

Notes. Monte Carlo results with $R = 1{,}000$ replications per cell and $B = 1{,}000$ pair-bootstrap draws per replication. For Classes B and C, N denotes the number of countries N_c. All standard errors are heteroskedasticity-robust (no clustering). SD is the Monte Carlo standard deviation of point estimates across replications. SE is the average analytical standard error reported by the estimator. Cov is the empirical coverage of the 95% confidence interval constructed from the analytical SE. 'w/ endog.' denotes `ppmlhdfe` applied to the endogenous DGP without IV correction. 'no endog.' denotes an exogenous DGP.

Table 3 shows four main findings.[6] First, `ivppmlhdfe` has far lower point bias than `ppmlhdfe` applied to the endogenous DGP without IV correction: `ppmlhdfe` without IV has bias of approximately 0.45, about 90% of the true coefficient, with zero coverage, whereas the `ivppmlhdfe` bias is at most −0.036 across all classes. Second, for Class A, `ivppmlhdfe` exhibits finite-sample bias that shrinks with N and T, ranging at $T = 20$ from −0.035 at $N = 50$ to −0.016 at $N = 500$, confirming consistency. Class B biases are of comparable magnitude (−0.017 to −0.030), and the rate of decay is consistent with the $O(1/N_c)$ order

[6] Appendix Tables C1, C2, C3, and C4 report the analogues of Tables 3 and 4 for $T = 10$ and $T = 30$.

predicted by theory. Class C biases are larger (-0.030 to -0.036) and shrink more slowly with N_c, reflecting the additional pair fixed effects. Third, `ivppmlhdfe` understates sampling variability. SE/SD ratios range from 0.60 to 0.79, and empirical coverage is 70–88% for Classes A and B and 37–60% for Class C. Fourth, `ppmlhdfe` applied to the exogenous DGP is consistent in its point estimate, but SE/SD ratios of 0.63–0.81 leave coverage at 74–89%, reflecting the downward bias in standard errors documented by Weidner and Zylkin (2021).

3.3 SPJ with Bootstrap Results

Table 4 compares `ivppmlhdfe` with the SPJ bias correction paired with bootstrap standard errors at $T = 20$.[7] For Class A, SPJ corrects 52–82% of the IV bias across the grid, and bootstrap CI coverage rises from 82–88% to 92–94%. For Class B, bias reduction ranges from 17 to 55%, with coverage rising from 70–74% to 89–98%. For Class C, bias reduction is more modest at small N_c (the 8-panel jackknife slightly worsens bias at $N_c = 20$) but reaches 36% at $N_c = 50$, and coverage improves from 37–60% to 83–97%. The appendix results for $T = 10$ and $T = 30$ follow the same qualitative pattern, with bias magnitudes shrinking in T. The computational cost is that SPJ requires 3 to 8 estimations per dataset, each repeated across B bootstrap draws. Pooled across the full grid (3 classes $\times$ 4 cross-section sizes $\times$ 3 time dimensions), SPJ reduces mean absolute bias by 42% and raises mean bootstrap CI coverage from 69% to 92%.

4 Empirical Applications

We apply `ivppmlhdfe` to three published studies that estimate IV-PPML with fixed effects. Each study uses a different strategy for the high-dimensional fixed effects (control function, dummy-variable GMM, and two-step plug-in), which we compare against `ivppmlhdfe` in turn.

[7] Results for $T = 10$ and $T = 30$ appear in Appendix Tables C3 and C4.

Table 4: IV-PPML vs. SPJ with bootstrap SE ($T = 20$)

Class	N	T	Estimator	Bias	Bias red.	SD	SE	SE/SD	Cov (%)
A	50	20	IV-PPML	−.035		.140	.097	0.70	82.7
			SPJ	−.014	60%	.183	.136	**0.74**	**93.6**
	100	20	IV-PPML	−.025		.109	.077	0.70	82.5
			SPJ	−.007	72%	.135	.104	**0.77**	**92.1**
	200	20	IV-PPML	−.025		.078	.060	0.77	86.0
			SPJ	−.012	52%	.095	.078	**0.82**	**92.0**
	500	20	IV-PPML	−.016		.054	.043	0.79	87.5
			SPJ	−.003	82%	.064	.055	**0.85**	**92.8**
B	20	20	IV-PPML	−.030		.050	.032	0.63	70.3
			SPJ	−.025	17%	.070	.071	**1.01**	**97.8**
	30	20	IV-PPML	−.024		.036	.024	0.67	72.9
			SPJ	−.015	38%	.046	.042	**0.91**	**91.4**
	40	20	IV-PPML	−.021		.029	.020	0.69	72.8
			SPJ	−.012	40%	.036	.032	**0.87**	**89.1**
	50	20	IV-PPML	−.017		.024	.017	0.73	74.4
			SPJ	−.008	55%	.029	.026	**0.88**	**89.8**
C	20	20	IV-PPML	−.036		.042	.025	0.61	60.1
			SPJ	−.037	−3%	.067	.075	**1.13**	**97.4**
	30	20	IV-PPML	−.034		.030	.018	0.60	49.6
			SPJ	−.028	18%	.043	.043	**1.00**	**89.4**
	40	20	IV-PPML	−.031		.023	.015	0.64	43.1
			SPJ	−.022	31%	.032	.031	**0.97**	**85.7**
	50	20	IV-PPML	−.030		.019	.012	0.66	36.6
			SPJ	−.019	36%	.026	.025	**0.95**	**83.2**

Notes. Monte Carlo results with $R = 1{,}000$ replications per cell and $B = 1{,}000$ bootstrap draws per replication. For Classes B and C, N denotes the number of countries N_c. 'IV-PPML' rows use the analytical heteroskedasticity-robust standard error reported by the estimator (matching Table 3). 'SPJ' rows use a bootstrap standard error, with individual resampling for Class A and pair resampling for Classes B and C. The reported SPJ standard error is a CI-implied SE: within each Monte Carlo replication, we take the width of the 95% percentile CI from the B bootstrap draws and divide by 2×1.96, the factor that relates CI width to SE under a normal approximation. The table reports the median of this CI-implied SE across the R replications. This construction is robust to a small fraction of divergent bootstrap draws, which inflate the bootstrap standard deviation but leave the percentiles unchanged. Class A SPJ: $\hat{\beta}_{\text{SPJ}} = 3\hat{\beta} - \bar{\beta}_{T/2} - \bar{\beta}_{N/2}$ (Fernández-Val and Weidner, 2016). Class B SPJ: 4-subpanel country-split jackknife (Weidner and Zylkin, 2021). Class C SPJ: country-split × time-split, yielding 8 subpanels, with $\hat{\beta}_{\text{SPJ}} = 4\hat{\beta}_{\text{full}} - 2\bar{\beta}_{\text{country}} - 2\bar{\beta}_{\text{time}} + \bar{\beta}_{\text{8cell}}$. Bold entries mark the SE/SD and coverage values for the SPJ rows.

4.1 Application 1: Flexible Pay and Labor Supply

Chen et al. (2025) exploit a randomized rollout of Uber's 'Instant Pay' feature, an earned-wage-access product that allows drivers to withdraw accumulated earnings on demand rather than waiting for the weekly pay cycle. Uber randomly assigned drivers in 12 U.S. cities to treatment or control between March and April 2016, and the outcome of interest is daily minutes worked. Because not all treated drivers take up the product, the authors estimate the effect using an IV strategy, where the randomly assigned treatment indicator (interacted with a post-enrollment indicator) instruments for the actual take-up.

Original Specifications and Results. Chen et al. (2025) estimate the effect via a control function (CF) approach. A linear first-stage regression of take-up on the instrument is estimated via `reghdfe`, absorbing city × cohort × day fixed effects (approximately 47,000 groups). The first-stage residual is then included as an additional regressor in a PPML regression of daily minutes worked on take-up, estimated via `ppmlhdfe`, with standard errors clustered at the driver level. Standard errors are computed via the delta method as suggested by Lin and Wooldridge (2019). Unlike IV estimation, which only exploits the moment condition $\mathbb{E}[q_g u_g] = 0$, the CF approach conditions on the first-stage residual and therefore requires that the first stage is correctly specified and that the structural error enters the exponential mean additively. Table 5, column (1), reports the main CF estimate from Table 1 of Chen et al. (2025): 0.1005 (SE 0.0453).

IVPPMLHDFE Results. Columns (2) and (3) of Table 5 report our `ivppmlhdfe` results. Applied directly to the same data, `ivppmlhdfe` yields 0.0970 (SE 0.0419), very close to the original result. Although the agreement is already tight, we apply the SPJ correction as a precaution against the downward IPP bias documented in our Monte Carlo simulations. The specification contains a single high-dimensional interacted fixed effect (city × cohort × day), with no separate driver-level or time-level fixed effect to correct for. We therefore

Table 5: Replication of Chen et al. (2025): Effect of Instant Pay on daily work minutes

	Original (CF)	`ivppmlhdfe`	`ivppmlhdfe` + SPJ
Effect on daily work minutes	0.1005**	0.0970**	0.1003**
	(0.0453)	(0.0419)	(0.0430)

Notes. N = 13,242,030 driver-day observations. Fixed effects: city × cohort × day (approximately 47,000 groups). Standard errors are clustered at the driver level. CF: linear first stage via `reghdfe`, then `ppmlhdfe` with residual, SE via delta method following Lin and Wooldridge (2019). SPJ: $\hat{\beta}_{\text{SPJ}} = 2\hat{\beta}_{\text{full}} - \bar{\beta}_{N/2}$ (N-only split on drivers). Bootstrap SE from $B = 200$ pair-bootstrap replications. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

apply only the cross-sectional (driver-level) split, yielding $\hat{\beta}_{\text{SPJ}} = 2\hat{\beta}_{\text{full}} - \bar{\beta}_{N/2}$. The SPJ estimate is 0.1003 with bootstrap SE 0.0430 ($B = 200$), within 0.2% of the CF estimate.

Takeaways. We draw two conclusions based on this analysis. First, `ivppmlhdfe` reproduces the CF regression results in a single command and does not require the linear first-stage or additive-residual assumptions that CF relies on.[8] Second, the close agreement between the two methods suggests that, in this case, the CF functional-form assumption is innocuous for estimation and the IPP is limited. Neither property is guaranteed in general.

4.2 Application 2: Country Image and Trade

Chang et al. (2022) study whether country image, measured by the share of positive public sentiment, affects bilateral trade flows. The specification includes exporter×year and importer×year fixed effects (Class B in our taxonomy), with two shift-share instruments constructed from language complementarity and genetic distance. The sample covers 15 exporters, 25 importers, and 10 years (N = 2,365), with standard errors clustered at the bilateral-pair level.

[8]While the original CF result is very close to our SPJ result, there is no guarantee that the CF approach is immune to the IPP by design.

Original Specifications and Results. Chang et al. (2022) implement IV-PPML with `ivpoisson gmm`, carrying 341 explicit dummy variables for the fixed effects (328 FE dummies plus controls) because `ivpoisson gmm` does not absorb high-dimensional fixed effects natively. Table 6, column (1), reports the original results from the three IV-PPML specifications in Table 5 of Chang et al. (2022): the language-based instrument alone (1.538, SE 0.574), the genetic-distance-based instrument alone (1.367, SE 0.581), and both instruments jointly via iterative GMM (1.453, SE 0.391).

Table 6: Replication of Chang et al. (2022): Effect of country image on trade

	Original (`ivpoisson gmm`)	`ivppmlhdfe`
(A) IV_1 only (just-id)	1.538***	1.538***
	(0.574)	(0.575)
(B) IV_2 only (just-id)	1.367**	1.367**
	(0.581)	(0.582)
(C) Both IVs (overid)	1.453***	1.426***
	(0.391)	(0.413)

Notes. N = 2,365 bilateral-year observations (15 exporters, 25 importers, 10 years). Fixed effects: exporter×year and importer×year (328 groups). Standard errors are clustered at the bilateral-pair level. IV_1 is the language-complementarity shift-share. IV_2 is the genetic-distance shift-share. The original uses `ivpoisson gmm` with 341 explicit regressors (328 FE dummies plus controls) and iterative GMM. The gap between the original overidentified estimate of 1.453 and the `ivppmlhdfe` estimate of 1.426 reflects one-step versus iterative GMM weighting. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

IVPPMLHDFE Results. Our `ivppmlhdfe` results appear in column (2) of Table 6. `ivppmlhdfe` absorbs the fixed effects via iterative projection and estimates the single parameter of interest. In the two just-identified cases, `ivppmlhdfe` matches the published estimates to the third decimal place: 1.538 (SE 0.575) with the language instrument and 1.367 (SE 0.582) with the genetic-distance instrument. In the overidentified specification, `ivppmlhdfe` returns 1.426 (SE 0.413), which differs from the published 1.453 by 0.027 because ivppmlhdfe uses one-step GMM rather than iterative GMM weighting. The overidentified estimate runs in under one second, a speedup of more than three orders of magnitude over the original's roughly 18-minute runtime.[9]

We are unable to apply the SPJ bias correction to this application, because the Class B SPJ strategy splits exporters and importers each into two groups, leaving roughly 600 observations and 180 fixed-effect groups per sub-panel, or about three observations per cell. After singleton dropping, the sub-panel models become rank-deficient. Nevertheless, in the two just-identified specifications ivppmlhdfe matches the published estimates to three decimal places. Without SPJ available, however, it is difficult to assess the extent of IPP bias in either the original or our own estimates.

Takeaways. Three takeaways emerge from this application. First, ivppmlhdfe exactly replicates the just-identified published estimates. Second, it is fast and does not require the user to enumerate fixed-effect dummies. Third, SPJ is not always feasible: when the sample is small relative to the number of fixed effects, sub-panel splitting can leave too few degrees of freedom for identification.

4.3 Application 3: Geographical Indications and Cheese Exports

Curzi and Huysmans (2022) estimate the effect of EU geographical indication (GI) protections on bilateral cheese exports. The specification includes pair×year fixed effects (8,392 groups) and product×year fixed effects (445 groups). The sample covers 25 EU exporters, 55 extra-EU trading partners, 30 CN8 cheese codes, and 16 years (2004–2019, $N = 241{,}150$), with the instrument being a cross-FTA, cross-product GI indicator (binary, 0.38% nonzero).

[9] Computation times are reported on a 12th Gen Intel Core i7-12700 CPU (2.10 GHz, 32 GB RAM at 4,400 MT/s) running Windows 11 and Stata 18. Exact timings are approximately 18 minutes for the original ivpoisson gmm overidentified specification with 341 explicit regressors ($N = 2{,}365$, 328 fixed-effect groups across two dimensions), and under 1 second for the corresponding ivppmlhdfe call on the same sample, which absorbs the fixed effects iteratively. Reported runtimes are hardware- and specification-dependent and should be read as an illustrative benchmark.

Original Specifications and Results. Curzi and Huysmans (2022) implement IV-PPML via a two-step procedure because `ivpoisson gmm` cannot absorb high-dimensional fixed effects natively. Step 1 runs `ppmlhdfe` with the endogenous Listed-GI variable included without instrumentation and saves the estimated fixed effects $\hat{\alpha}_g$. Step 2 feeds $\hat{\alpha}_g$ as a single scalar regressor into `ivpoisson gmm` together with the Listed-GI variable and the cross-FTA instrument. The paper reports two main IV-PPML results from this procedure: the average listed-GI effect (their Table 2, column 4) is 0.067 (SE 0.052, not statistically significant), and the ex-officio-protection effect (their Table 5, column 2)[10] is 0.373 (SE 0.150, significant at the 5% level). We replicate these results in column (1) of Table 7.

Table 7: Replication of Curzi and Huysmans (2022): Two-step plug-in vs. joint `ivppmlhdfe`

	Original (two-step)	`ivppmlhdfe`
Listed GI effect (CH Table 2)	0.067	0.289**
	(0.052)	(0.118)
Ex-officio enforcement (CH Table 5)	0.373**	0.651**
	(0.150)	(0.281)

Notes. $N = 241{,}150$ bilateral-product-year observations. Fixed effects: pair×year (8,392 groups) and product×year (445 groups). Standard errors are clustered at the exporter-importer-product level. The original two-step runs `ppmlhdfe` to save fixed effects, then `ivpoisson gmm` with those fixed effects as a single scalar regressor. `ivppmlhdfe` estimates the coefficient of interest and the fixed effects jointly within a single GMM problem. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

IVPPMLHDFE Results. Columns (1) and (2) of Table 7 report the original estimates alongside our `ivppmlhdfe` results. `ivppmlhdfe` avoids the two-step by jointly estimating the coefficient of interest and the fixed effects within a single GMM problem, profiling out the fixed effects at each IRLS iteration. For the listed-GI effect, `ivppmlhdfe` yields 0.289 (SE

[10] Ex-officio protection requires the trading partner's administration to actively monitor the market for breaches of geographical indication rights, rather than leaving enforcement to the rightsholder to pursue through legal action. Curzi and Huysmans (2022) hypothesize that stronger protection of this kind reduces infringement and therefore generates larger export gains for listed GIs than mere listing in the register.

0.118), more than four times the two-step estimate of 0.067 and statistically significant at the 5% level. For the ex-officio-protection effect, `ivppmlhdfe` yields 0.651 (SE 0.281), nearly double the two-step value of 0.373. The main substantive conclusion of Curzi and Huysmans (2022), that ex-officio protection matters more than mere listing, therefore remains unchanged, but the magnitudes more than double under `ivppmlhdfe`. We again cannot apply the SPJ correction, this time because the instrument is sparse: only 0.38% of observations take a nonzero value, the 8-panel strategy further dilutes this variation, and the smallest sub-panel quadrant retains only 0.14% nonzero observations, which destabilizes the sub-panel estimates.

Takeaways. `ivppmlhdfe` obviates the two-step procedure, and with it the standard-error corrections required by two-stage approaches to account for estimation uncertainty in the plug-in fixed effects. As in Application 2, the SPJ bias correction is infeasible, although the failure mechanism differs. Application 2 produces sub-panels with too few observations per fixed-effect cell, whereas the present application produces sub-panels with too little instrument variation to identify the parameter of interest.

5 Conclusion

This paper implements an instrumental-variable extension of the Poisson pseudo-maximum likelihood estimator that accommodates high-dimensional fixed effects. Our contribution is twofold. Methodologically, we document that the Poisson score cancellation, which renders standard PPML immune to leading-order incidental parameter bias, does not carry over to the IV case, and we propose split-panel jackknife corrections paired with bootstrap standard errors as a practical remedy. Monte Carlo simulations show that this combination reduces mean absolute bias by 42% and raises mean bootstrap confidence-interval coverage from 69% to 92% across three commonly used fixed-effect structures. Practically, we provide a fast, feasible estimator, `ivppmlhdfe`, that combines IV estimation with the iterative fixed-effect absorption methods of Correia et al. (2020) and achieves computational and functional parity with the widely used `ppmlhdfe` command.

We apply the estimator to three published studies. For Chen et al. (2025) and Chang et al. (2022), `ivppmlhdfe` reproduces the published estimates and validates the command as a more user-friendly replacement. In Curzi and Huysmans (2022), we show that the two-step plug-in procedure used in the original paper attenuates coefficients by a factor of two to four, although the authors' main conclusion remains unchanged. These latter two applications also highlight a practical limitation of SPJ: it requires that each sub-panel retain enough observations per fixed-effect cell and enough instrument variation to identify the parameter of interest. This requirement fails when the cross-section is small or the instrument is sparse.

Several directions for future work merit attention. First, our analysis focuses on the just-identified case with a single endogenous regressor. Extension to overidentified models, following the general framework of Fernández-Val and Lee (2013), would broaden the estimator's applicability. Second, while we demonstrate that the SPJ with bootstrapped standard errors procedure works well, we also see potential benefits from the development of an analytical bias correction procedure for the case of IV-PPML with high-dimensional fixed effects.

References

Jeffrey H. Bergstrand and Peter H. Egger. A knowledge-and-physical-capital model of international trade flows, foreign direct investment, and multinational enterprises. *Journal of International Economics*, 73(2):278–308, 2007.

Cosimo Beverelli. Pull factors for migration: The impact of migrant integration policies. *Economics and Politics*, September 2021. URL https://doi.org/10.1111/ecpo.12191.

Sebastien Bradley, Federico Carril-Caccia, and Yoto Yotov. Reassessing the Effects of Corporate Income Taxes on Mergers and Acquisitions Using Empirical Advances in the Gravity Literature. School of Economics Working Paper Series 2023-8, LeBow College of Business, Drexel University, December 2023.

Federico Carril-Caccia, Juliette Milgram-Baleix, and Jordi Paniagua. Does terrorism affect foreign greenfield investments? *Defence and Peace Economics*, 34(6):827–844, 2023.

Pao-Li Chang, Tomoki Fujii, and Hangyu Jin. Does country image affect international trade? Evidence from Japanese manga. *Management Science*, 68(4):2784–2809, 2022.

Jiafeng Chen and Jonathan Roth. Logs with zeros? Some problems and solutions. *Quarterly Journal of Economics*, 139(2):891–936, 2024.

M. Keith Chen, Katherine Feinerman, and Kareem Haggag. Flexible pay and labor supply: Evidence from Uber's Instant Pay. *Management Science*, pages 1–20, 2025. Articles in Advance.

Jonathan B. Cohn, Zack Liu, and Malcolm I. Wardlaw. Count (and count-like) data in finance. *Journal of Financial Economics*, 146(2):529–551, 2022.

Sergio Correia, Paulo Guimarães, and Thomas Zylkin. Fast Poisson estimation with high-dimensional fixed effects. *Stata Journal*, 20(1):95–115, 2020.

Daniele Curzi and Martijn Huysmans. The impact of protecting EU geographical indications in trade agreements. *American Journal of Agricultural Economics*, 104(1):364–384, 2022.

Jonathan I. Dingel and Felix Tintelnot. Spatial economics for granular settings. *Econometrica*, 94:407–464, 2026.

Peter Egger, Mario Larch, Kevin E. Staub, and Rainer Winkelmann. The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy*, 3(3):113–143, 2011.

Iván Fernández-Val and Jinyong Lee. Panel data models with nonadditive unobserved heterogeneity: Estimation and inference. *Quantitative Economics*, 4(3):453–481, 2013.

Iván Fernández-Val and Martin Weidner. Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192(1):291–312, 2016.

Koen Jochmans and Vincenzo Verardi. Instrumental-variable estimation of exponential-regression models with two-way fixed effects with an application to gravity equations. *Journal of Applied Econometrics*, 37(6):1121–1137, 2022.

G. Andrew Karolyi and Alvaro G. Taboada. Regulatory arbitrage and cross-border bank acquisitions. *Journal of Finance*, 70(6):2395–2450, 2015.

Brandon J Kinne. Network dynamics and the evolution of international cooperation. *American Political Science Review*, 107(4):766–785, November 2013.

Jesse LaBelle, Inmaculada Martinez-Zarzoso, Ana Maria Santacreu, and Yoto Yotov. Cross-border Patenting, Globalization, and Development. School of Economics Working Paper Series 2023-7, LeBow College of Business, Drexel University, December 2023.

Mario Larch and Yoto V. Yotov. Deep trade agreements and fdi in partial and general equilibrium: A structural estimation framework. *The World Bank Economic Review*, 39: 281–307, 2025.

Mario Larch, Joschka Wanner, and Yoto V. Yotov. Bi- and Unilateral trade effects of joining the Euro. *Economics Letters*, 171(C):230–234, 2018.

Mario Larch, Serge Shikher, and Yoto V. Yotov. Estimating gravity equations: Theory implications, econometric developments, and practical recommendations. *Review of International Economics*, 33(5):1066–1092, November 2025.

Wei Lin and Jeffrey M. Wooldridge. Testing and correcting for endogeneity in nonlinear unobserved effects models. In *Panel Data Econometrics*, pages 21–43. Academic Press, 2019.

Rogelio Mercado. Bilateral capital flows: Gravity, push and pull. *International Finance*, September 2022. URL https://doi.org/10.1111/infi.12421.

Jean-Frédéric Morin, Noemie Laurens, Clara Brandi, and Jakob Schwab. Using trade provisions to make environmental agreements more dynamic. *International Studies Quarterly*, 69(1):sqaf010, 2025.

John Mullahy. Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics*, 79(4):586–593, 1997.

Damiaan Persyn and Wouter Torfs. A gravity equation for commuting with an application to estimating regional border effects in belgium. *Journal of Economic Geography*, 16(1): 155–175, 03 2015.

Jacques Poot, Omoniyi Alimi, Michael P. Cameron, and David C. Mará. The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science. *Investigaciones Regionales - Journal of Regional Research*, pages 63–86, 2016.

João M.C. Santos Silva and Silvana Tenreyro. The log of gravity. *Review of Economics and Statistics*, 88(4):641–658, 2006.

João M.C. Santos Silva and Silvana Tenreyro. The log of gravity at 15. *Portuguese Economic Journal*, 21(3):423–437, 2022. doi: 10.1007/s10258-021-00203-w.

Silvana Tenreyro. On the trade impact of nominal exchange rate volatility. *Journal of Development Economics*, 82(2):485–508, 2007.

Martin Weidner and Thomas Zylkin. Bias and Consistency in Three-way Gravity Models. *Journal of International Economics*, 132(103513), 2021.

Frank A. G. Windmeijer and J. M. C. Santos Silva. Endogeneity in count data models: An application to demand for health care. *Journal of Applied Econometrics*, 12(3):281–294, 1997.

Appendix

A IPP bias orders and remedies by model class

This appendix derives the incidental-parameter bias orders reported for IV-PPML in Table 2, and shows that the leading order of the estimated standard error bias inherits the same order. Throughout we maintain the notation of Section 2.1: observations are indexed by $g = 1, \ldots, n$, with outcome $y_g \geq 0$, regressors $w_g = \left(x_g', d_g\right)' \in \mathbb{R}^{K+1}$, instruments $q_g = \left(x_g', z_g\right)' \in \mathbb{R}^{K+1}$, and fixed-effect sum ψ_g. The conditional mean is $\mu_g\left(\beta, \psi_g\right) = \exp\left(w_g'\beta + \psi_g\right)$, with residual $u_g\left(\beta, \psi_g\right) = y_g - \mu_g\left(\beta, \psi_g\right)$. The IV-PPML-HDFE estimator $\left(\hat{\beta}, \hat{\psi}\right)$ solves the just-identified GMM system

$$\frac{1}{n}\sum_{g=1}^{n} q_g u_g\left(\beta, \psi_g\right) = 0, \quad \sum_{g\in r} u_g\left(\beta, \psi_g\right) = 0 \quad \text{for every fixed-effect group } r. \tag{5}$$

We partition the fixed-effect parameters into M *families* (e.g., $\{\alpha_i\}$, $\{\gamma_t\}$, $\{\eta_{ij}\}$). Family m contains K_m parameters (one per cell), and each parameter is identified from h_m observations on average. By construction, $K_m h_m \asymp n$ for every family.

The roadmap of the appendix is as follows. Section A.1 states the maintained regularity conditions and the generic second-order expansion of $\hat{\beta}$. Section A.2 derives the family-wise bias expansion and records the resulting *p/n rule*: to leading order, the bias of a nonlinear just-identified GMM estimator with high-dimensional fixed effects is of order $\sum_{m=1}^{M} h_m^{-1}$, absent a score cancellation. Section A.3 explains why the Poisson score cancellation neutralizes specific family contributions for benchmark PPML but not for IV-PPML. Sections A.4–A.6 specialize the general expansion to the three model classes and thereby establish the second row of Table 2. Section A.7 derives the third row (standard-error bias). Section A.8 discusses the implications for consistency: for Classes A and C, both N (or N_c) and T must diverge jointly.

A.1 Setup, notation, and regularity

For any smooth function $f_g(\beta, \psi)$ we use subscripts for partial derivatives, e.g., $u_g^{\psi} \equiv \partial u_g / \partial \psi_g$, $u_g^{\psi\psi} \equiv \partial^2 u_g / \partial \psi_g^2$, and analogously for derivatives with respect to β. Under the exponential mean (1),

$$u_g^{\psi} = -\mu_g, \quad u_g^{\psi\psi} = -\mu_g, \quad u_g^{\psi\psi\psi} = -\mu_g, \quad u_g^{\beta} = -\mu_g w_g, \quad u_g^{\beta\psi} = -\mu_g w_g. \tag{6}$$

For each family m and cell c, write $\psi_{m,c}$ for the corresponding fixed-effect parameter. Let $\phi = (\psi_{m,c})$ collect all $p_\phi \equiv \sum_m K_m$ fixed effects, and let ϕ_0 denote the truth. Let $\hat{\phi}(\beta)$ solve the fixed-effect equations in (5) given β, and write $\hat{\phi}_0 = \hat{\phi}(\beta_0)$.

We maintain the following high-level regularity conditions, which are the just-identified IV analog of Assumption 4.1 of Fernández-Val and Weidner (2016) and Assumption A of Weidner and Zylkin (2021).

Assumption 1 (Regularity)**.**

1. $\mathbb{E}[y_g \mid w_g, \psi_{g,0}] = \mu_g(\beta_0, \psi_{g,0}) = \exp(w_g' \beta_0 + \psi_{g,0})$.

2. $\mathbb{E}[q_g u_g(\beta_0, \psi_{g,0}) \mid \mathcal{F}_g, \psi_{g,0}] = 0$, *where $\mathcal{F}_g$ is a sigma-field containing the admissible conditioning information (exogenous covariates, instruments, lagged residuals, and fixed effects).*

3. *No restriction is placed on the joint distribution of $(x_g, d_g, z_g, \psi_{g,0})$; in particular, $\psi_{g,0}$ may be arbitrarily correlated with d_g and may also be correlated with (x_g, z_g). What is ruled out is only residual correlation between the fixed-effect heterogeneity and the structural IV-PPML error: for every square-integrable measurable function a,*

$$\mathbb{E}[a(\psi_{g,0}) u_g(\beta_0, \psi_{g,0})] = 0,$$

or, equivalently, $\mathbb{E}[u_g(\beta_0, \psi_{g,0}) \mid \sigma(\psi_{g,0})] = 0$ whenever the conditional expectation is *well defined. Thus fixed effects may be correlated with the endogenous regressor, but they do not absorb residual variation left unexplained by the structural exponential mean.*

4. *For β in a neighborhood of β_0, the profile fixed-effect system admits a unique solution $\hat{\phi}(\beta)$. The profile Jacobian*

$$J_0 \equiv p\lim \frac{\partial}{\partial \beta'} \left(\frac{1}{n} \sum_g q_g u_g \left(\beta, \hat{\phi}(\beta) \right) \right) \Bigg|_{\beta=\beta_0}$$

exists and is nonsingular.

5. *The standard smoothness, moment, and weak-dependence conditions of Fernández-Val and Weidner (2016, Assumption 4.1) and Weidner and Zylkin (2021, Assumption A) hold, including (a) bounded $(8+\nu)$-moments of all derivatives of u_g up to fourth order, (b) α-mixing over the time dimension conditional on the fixed effects with mixing coefficients of polynomial decay, and (c) cross-sectional independence conditional on the fixed effects.*

6. *For every family m and cell c, the Schur-complement score, Hessian, and third partial of the concentrated objective in family m, evaluated at the truth, satisfy*

$$\bar{S}_{m,c,n} = O_p(h_m^{1/2}), \quad \bar{H}_{m,c,n} = h_m \bar{h}_{m,c} + o_p(h_m), \quad \bar{J}_{m,c,n} = h_m \bar{j}_{m,c} + o_p(h_m),$$

uniformly over (m, c), with $0 < \underline{h} \leq |\bar{h}_{m,c}| \leq \overline{h} < \infty$ and $|\bar{j}_{m,c}| \leq \bar{j} < \infty$.

Parts 1–2 are the correctly specified structural mean and IV-validity conditions, respectively. Part 2 is the conditional-moment restriction that validates the IV score: instruments, exogenous covariates, and admissible predetermined information may be used in q_g, while contemporaneously endogenous components of d_g need not be strictly exogenous. Part 3 makes explicit that the fixed effects are allowed to be arbitrarily correlated with the endogenous regressor and with the other observed variables, subject only to the residual-orthogonality requirement that $\psi_{g,0}$ is not correlated with $y_g - \mu_g(\beta_0, \psi_{g,0})$ (Windmeijer and Santos Silva, 1997). This is weaker than independence between fixed effects and regressors and is the condition needed for the fixed-effect score equations to remain valid. Parts 4–5 are the standard smoothness conditions under which the second-order expansion of the profile score admits a uniform bound on its remainder. Part 6 is the just-identified counterpart of the cell-size bounds used in Lemma D.1 of Fernández-Val and Weidner (2016) and Lemma S.1 of Weidner and Zylkin (2021); it rules out degenerate fixed-effect blocks.

A.2 General family-wise bias expansion and the p/n rule

We first record the general second-order expansion of $\hat{\beta}$ in terms of the incidental-parameter estimation error.

Lemma 1 (Family-wise expansion of the fixed effects)**.** *Under Assumption 1, for every family m and cell c,*

$$\hat{\psi}_{m,c,0} - \psi_{m,c,0} = h_m^{-1/2}\zeta_{m,c} + h_m^{-1}b_{m,c} + r_{m,c}, \quad \sup_{m,c}|r_{m,c}| = o_p\left(h_m^{-1}\right), \tag{7}$$

where $\mathbb{E}\left[\zeta_{m,c} \mid \mathcal{F}_{m,c}\right] = 0$*,* $\sup_{m,c}\mathbb{E}\left[\zeta_{m,c}^2\right] < \infty$*, and* $\sup_{m,c}\mathbb{E}\left|b_{m,c}\right| < \infty$*.*

Proof. Let $\delta_{m,c} = \hat{\psi}_{m,c,0} - \psi_{m,c,0}$. A second-order Taylor expansion of the (Schur-complement) fixed-effect equation $\bar{S}_{m,c,n}\left(\beta_0, \hat{\psi}_{m,c,0}\right) = 0$ around $\psi_{m,c,0}$ gives

$$0 = \bar{S}_{m,c,n}^0 + \bar{H}_{m,c,n}\delta_{m,c} + 0.5\bar{J}_{m,c,n}\delta_{m,c}^2 + \bar{R}_{m,c,n},$$

with $\sup_{m,c}\left|\bar{R}_{m,c,n}\right| \le Ch_m\left|\delta_{m,c}\right|^3$ by Assumption 1(v)–(vi) and smoothness. From $\bar{S}_{m,c,n}^0 = O_p\left(h_m^{1/2}\right)$ and $\bar{H}_{m,c,n} \asymp h_m$, the linearized equation gives $\delta_{m,c} = O_p\left(h_m^{-1/2}\right)$ uniformly, hence

$\bar{R}_{m,c,n} = o_p\left(h_m^{-1/2}\right)$. One Newton step yields

$$\delta_{m,c} = -\bar{H}_{m,c,n}^{-1}\bar{S}_{m,c,n}^{0} - 0.5\bar{H}_{m,c,n}^{-1}\bar{J}_{m,c,n}\left(\bar{H}_{m,c,n}^{-1}\bar{S}_{m,c,n}^{0}\right)^2 + o_p\left(h_m^{-1}\right),$$

which is (7) with $\zeta_{m,c} = -h_m^{1/2}\bar{H}_{m,c,n}^{-1}\bar{S}_{m,c,n}^{0}$ and $b_{m,c} = -0.5h_m\bar{H}_{m,c,n}^{-1}\bar{J}_{m,c,n}(\bar{H}_{m,c,n}^{-1}\bar{S}_{m,c,n}^{0})^2$. The moment bounds follow from Assumption 1(v)–(vi). □

Lemma 2 (Profile-score expansion). *Let $G_n^p(\beta) \equiv n^{-1}\sum_g q_g u_g\left(\beta, \hat{\psi}_g(\beta)\right)$ denote the profiled IV score. If Assumption 1 holds, there exist finite vectors $\left\{\Delta_m^B, \Delta_m^C, \Delta_m^V\right\}_{m=1}^M$ and $\left\{\Delta_{m\ell}^X\right\}_{1\le m<\ell\le M}$ such that*

$$\begin{aligned} G_n^p(\beta_0) = n^{-1/2}\mathcal{Z}_n + \sum_{m=1}^{M}\frac{\Delta_m^B + \Delta_m^C + \Delta_m^V}{h_m} + \sum_{1\le m<\ell\le M}\frac{\Delta_{m\ell}^X}{\sqrt{h_m h_\ell}} \\ + \ o_p\left(n^{-1/2} + \textstyle\sum_m h_m^{-1} + \sum_{m<\ell}(h_m h_\ell)^{-1/2}\right), \end{aligned} \tag{8}$$

where $\mathcal{Z}_n \xrightarrow{d} \mathcal{N}(0, \Omega_0)$ for a finite positive-semidefinite Ω_0. The four components are:

- *Δ_m^B: the higher-order bias term generated by $b_{m,c}$ in (7);*
- *Δ_m^C: the common-score/common-nuisance correlation term generated by the linear $\zeta_{m,c}$ term;*
- *Δ_m^V: the same-family quadratic term generated by $\zeta_{m,c}^2$;*
- *$\Delta_{m\ell}^X$: the cross-family quadratic interaction term from $\zeta_{m,c}\zeta_{\ell,d}$.*

Proof. Write $\delta_m = (\delta_{m,c})_c$ and expand $G_n(\beta_0, \phi)$ around ϕ_0 to second order:

$$G_n^p(\beta_0) = \frac{1}{n}\sum_g q_g u_g^0 + A_n\delta + 0.5\delta' B_n\delta + R_n,$$

where A_n and B_n collect the first and second derivatives of G_n with respect to ϕ evaluated at the truth. The first term on the right is a centered IV-score sample average and satisfies $n^{-1/2}\mathcal{Z}_n \xrightarrow{d} \mathcal{N}(0, \Omega_0)$ by Assumption 1(ii), (v). Substituting (7) into the linear and quadratic terms:

- The linear contribution in $h_m^{-1/2}\zeta_{m,c}$ has a centered component of order $O_p\left(n^{-1/2}\right)$ and a non-centered component of order h_m^{-1}, because each summand is of order $O_p\left(h_m^{1/2}/n\right)$, there are $K_m \asymp n/h_m$ summands, and the sum of their means is $O\left(h_m^{-1}\right)$. This is the correlation term Δ_m^C/h_m.

- The linear contribution in $h_m^{-1}b_{m,c}$ has mean of order $K_m h_m/\left(nh_m\right) = h_m^{-1}$, delivering Δ_m^B/h_m.

- The same-family quadratic term $\sum_c B_{mm,c,n}\delta_{m,c}^2$ contributes $O_p\left(h_m^{-1}\right)$, giving Δ_m^V/h_m.

- For $m \neq \ell$, the cross-family term $\sum_{c,d} B_{m\ell,cd,n}\delta_{m,c}\delta_{\ell,d}$ contributes $O_p\left(\left(h_m h_\ell\right)^{-1/2}\right)$, giving $\Delta_{m\ell}^X/\sqrt{h_m h_\ell}$.

Collecting terms yields (8). A fully rigorous derivation follows the argument of Theorem B.1 in Fernández-Val and Weidner (2016) with obvious changes for the just-identified GMM structure and the IV-specific expected-Hessian blocks. □

Proposition 1 (Family-wise bias of the IV-PPML-HDFE estimator)**.** *Under Assumption 1,*

$$\hat{\beta} - \beta_0 = O_p\left(n^{-1/2}\right) + \sum_{m=1}^{M} \frac{B_m}{h_m} + \sum_{1\le m<\ell\le M} \frac{B_{m\ell}}{\sqrt{h_m h_\ell}} + o_p\left(\textstyle\sum_m h_m^{-1}\right), \tag{9}$$

where $B_m = -J_0^{-1}\left(\Delta_m^B + \Delta_m^C + \Delta_m^V\right)$ *and* $B_{m\ell} = -J_0^{-1}\Delta_{m\ell}^X$*, with* $\Delta_m^{\cdot}$ *and* $\Delta_{m\ell}^X$ *as in Lemma 2. In particular, the leading-order incidental-parameter bias of* $\hat{\beta}$ *has order*

$$\mathrm{Bias}_{\mathrm{IPP}}\left(\hat{\beta}\right) = O\left(\sum_{m=1}^{M} h_m^{-1}\right), \tag{10}$$

because $\left(h_m h_\ell\right)^{-1/2} \le 0.5\left(h_m^{-1} + h_\ell^{-1}\right)$.

Proof. The estimator $\hat{\beta}$ solves $G_n^p\left(\hat{\beta}\right) = 0$. A first-order expansion around β_0 gives

$$0 = G_n^p\left(\beta_0\right) + J_0\left(\hat{\beta} - \beta_0\right) + o_p\left(\left\|\hat{\beta} - \beta_0\right\|\right).$$

Substituting (8) and premultiplying by J_0^{-1} yields (9). The bias order (10) follows by applying AM–GM to the cross-family term. $\square$

Remark 1 (p/n heuristic)**.** *Proposition 1 formalizes the p/n rule for incidental-parameter bias in nonlinear fixed-effect models, in the spirit of the expansions of Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021): to leading order, the IPP bias is the sum over fixed-effect families of the reciprocal of the average cell size,*

$$\mathrm{Bias}_{\mathrm{IPP}}(\hat{\beta}) \sim \sum_{m=1}^{M} \frac{1}{h_m} = \frac{(\#\ \textit{incidental parameters})}{(\#\ \textit{observations per parameter}) \cdot 1}.$$

The result (10) is purely a consequence of the cell-counting structure of the fixed-effect system in (5) and the validity of the second-order expansion; it does not rely on any special feature of the Poisson likelihood.

A.3 Why PPML score cancellation does not carry over to IV-PPML

Proposition 1 gives the generic leading-order bias. Whether the bias constants B_m are actually nonzero depends on whether a score cancellation eliminates them. We now explain why this cancellation occurs under benchmark PPML but generically fails under IV-PPML.

Benchmark PPML. Benchmark PPML corresponds to $q_g = w_g$ in (5). In that case the slope and fixed-effect moments are generated by the same log-likelihood, and (6) yields

$$u_g^{\beta\psi} = u_g^{\psi} w_g = (\text{FE moment gradient}) \times w_g. \tag{11}$$

This alignment is the algebraic source of the Poisson score cancellation. Specifically, the bias component $\Delta_m^B + \Delta_m^C$ in Lemma 2 involves the projection of q_g onto the linear span of the fixed-effect indicators in family m, weighted by μ_g. When $q_g = w_g$, the identity (11) combined with the fixed-effect FOCs $\sum_{g\in r} u_g = 0$ implies that this projection coincides with the population projection of w_g onto the same span, so the corresponding bias constant vanishes for each family whose fixed effect enters the likelihood through a score of the form $\partial_\psi \ell_g = u_g$. This is the two-way Poisson no-IPP result of Fernández-Val and Weidner (2016, Theorem 4.1 and Remark 3). For Class C, Weidner and Zylkin (2021, Proposition 3) further shows that the pair fixed effects η_{ij} can be profiled out in closed form via the shared fixed-effect FOC $\sum_t u_{ijt} = 0$, reducing the problem to a multinomial-type profile that carries only a residual $O\left(N_c^{-1}\right)$ bias from the directional effects.

IV-PPML. In IV-PPML, the slope moment replaces w_g with q_g while the fixed-effect moments are unchanged. The equation (11) becomes

$$u_g^{\beta\psi} = u_g^{\psi} w_g \neq u_g^{\psi} q_g \quad \text{whenever } d_g \neq z_g.$$

Because $q_g \neq w_g$ generically, the projection of q_g onto the fixed-effect span weighted by μ_g no longer coincides with the projection of the $u_g^{\beta\psi}$ weight, and the family-wise bias constants B_m in (9) are generically nonzero. Equivalently, the closed-form profile of η_{ij} (when present) continues to exist via the shared FE FOC, but the profiled slope moment $\sum_g q_g \tilde{u}_g$ is no longer a multinomial MLE score and thus does not inherit the no-IPP property of the benchmark PPML case. We therefore take (10) as a tight bound, rather than an upper bound, on the leading-order bias of the IV-PPML estimator.

Remark 2 (Interpretation of Table 2)**.** *Row "IPP bias (PPML)" of Table 2 reports what survives of* (10) *after the score-cancellation arguments of Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021). Row "IPP bias (IV-PPML)" reports* (10) *itself, because no analogous cancellation is available. The two rows therefore differ precisely by the family contributions that the PPML alignment* (11) *is able to zero out.*

A.4 Class A: individual and time fixed effects

For Class A, $g = (i, t)$, $n = NT$, and $\psi_g = \alpha_i + \gamma_t$. The fixed-effect families are

$$\mathcal{F}_\alpha = \{\alpha_i\}_{i=1}^N, \quad \mathcal{F}_\gamma = \{\gamma_t\}_{t=1}^T.$$

Each α_i is identified from the T observations $\{(i, t) : t = 1, \ldots, T\}$, and each γ_t from the N observations $\{(i, t) : i = 1, \ldots, N\}$. Hence

$$h_\alpha = T, \quad h_\gamma = N. \tag{12}$$

Proposition 2 (Class A bias order)**.** *Under Assumption 1, the Class A IV-PPML-HDFE estimator satisfies*

$$\hat{\beta}_A - \beta_0 = \frac{B_\alpha^A}{T} + \frac{B_\gamma^A}{N} + O_p\left((NT)^{-1/2}\right) + o_p\left(T^{-1} + N^{-1}\right), \tag{13}$$

for finite vectors B_α^A, B_γ^A that are generically nonzero. Hence $\text{Bias}_{\text{IPP}}(\hat{\beta}_A) = O(T^{-1} + N^{-1})$.

Proof. With $M = 2$ and cell sizes (12), Proposition 1 gives

$$\hat{\beta}_A - \beta_0 = O_p(n^{-1/2}) + \frac{B_\alpha^A}{T} + \frac{B_\gamma^A}{N} + \frac{B_{\alpha\gamma}^A}{\sqrt{NT}} + o_p(T^{-1} + N^{-1}).$$

Because $(NT)^{-1/2} \leq \frac{1}{2}(T^{-1} + N^{-1})$, the cross-family term is absorbed into the combined $O(T^{-1} + N^{-1})$ order, yielding (13). The constants B_α^A, B_γ^A are generically nonzero by the IV-PPML alignment argument of Section A.3. □

Remark 3 (Comparison with benchmark PPML)**.** *For benchmark PPML, the score cancellation of Fernández-Val and Weidner (2016, Example 2) implies $B_\alpha^A = B_\gamma^A = 0$ when the non-fixed-effect regressors are strictly exogenous, so that $\hat{\beta}_A^{\text{PPML}} - \beta_0 = O_p((NT)^{-1/2})$ and the PPML column of Table 2 reads* 0.

A.5 Class B: directional gravity fixed effects

For Class B, $g = (i, j, t)$ with $i \neq j$, $n = N_c(N_c - 1)T$, and $\psi_g = \alpha_{it} + \gamma_{jt}$. The families are

$$\mathcal{F}_\alpha = \{\alpha_{it}\}_{i,t}, \quad \mathcal{F}_\gamma = \{\gamma_{jt}\}_{j,t}.$$

For fixed (i, t), α_{it} is identified from the $N_c - 1$ observations $\{(i, j, t) : j \neq i\}$; symmetrically for γ_{jt}. Hence

$$h_\alpha = h_\gamma = N_c - 1 \asymp N_c. \tag{14}$$

Proposition 3 (Class B bias order)**.** *Under Assumption 1, the Class B IV-PPML-HDFE estimator satisfies*

$$\hat{\beta}_B - \beta_0 = \frac{B^B}{N_c} + O_p\left((N_c^2 T)^{-1/2}\right) + o_p(N_c^{-1}), \tag{15}$$

for a finite vector B^B that is generically nonzero. Hence $\mathrm{Bias}_{\mathrm{IPP}}(\hat{\beta}_B) = O(N_c^{-1})$.

Proof. With $M = 2$ and cell sizes (14), Proposition 1 gives

$$\hat{\beta}_B - \beta_0 = O_p(n^{-1/2}) + \frac{B_\alpha^B + B_\gamma^B}{N_c} + \frac{B_{\alpha\gamma}^B}{N_c} + o_p(N_c^{-1}).$$

Collecting the family-wise and cross-family contributions into a single constant $B^B = B_\alpha^B + B_\gamma^B + B_{\alpha\gamma}^B$ yields (15). The constant B^B is generically nonzero by the alignment argument of Section A.3. □

Remark 4 (Comparison with benchmark PPML)**.** *For benchmark PPML with $q_g = w_g$, the directional-effects analogue of the score cancellation yields $B^B = 0$ (Fernández-Val and Weidner, 2016, Theorem 4.1), so the PPML column for Class B in Table 2 is also* 0.

A.6 Class C: directional plus pair fixed effects

For Class C, $g = (i, j, t)$ with $i \neq j$, $n = N_c(N_c - 1)T$, and $\psi_g = \alpha_{it} + \gamma_{jt} + \eta_{ij}$. The families are

$$\mathcal{F}_\alpha = \{\alpha_{it}\}, \quad \mathcal{F}_\gamma = \{\gamma_{jt}\}, \quad \mathcal{F}_\eta = \{\eta_{ij}\}_{i \neq j}.$$

Each α_{it} and γ_{jt} is identified from $N_c - 1$ partner observations, while each η_{ij} is identified from the T observations $\{(i, j, t) : t = 1, \ldots, T\}$. Hence

$$h_\alpha = h_\gamma = N_c - 1 \asymp N_c, \quad h_\eta = T. \tag{16}$$

Proposition 4 (Class C bias order)**.** *Under Assumption 1, the Class C IV-PPML-HDFE estimator satisfies*

$$\hat{\beta}_C - \beta_0 = \frac{B_\eta^C}{T} + \frac{B_{\alpha\gamma}^C}{N_c} + O_p\left((N_c^2 T)^{-1/2}\right) + o_p(T^{-1} + N_c^{-1}), \tag{17}$$

for finite vectors $B_\eta^C, B_{\alpha\gamma}^C$ that are generically nonzero. Hence $\mathrm{Bias}_{\mathrm{IPP}}(\hat{\beta}_C) = O(T^{-1} + N_c^{-1})$.

Proof. With $M = 3$ and cell sizes (16), Proposition 1 gives

$$\hat{\beta}_C - \beta_0 = O_p(n^{-1/2}) + \frac{B_\alpha^C}{N_c} + \frac{B_\gamma^C}{N_c} + \frac{B_\eta^C}{T} + \sum_{m<\ell} \frac{B_{m\ell}^C}{\sqrt{h_m h_\ell}} + o_p(T^{-1} + N_c^{-1}).$$

Since $(N_c T)^{-1/2} \leq \frac{1}{2}(N_c^{-1} + T^{-1})$ and $N_c^{-1} \leq \frac{1}{2}(N_c^{-1} + N_c^{-1})$, all cross-family terms are absorbed into the combined $O(T^{-1} + N_c^{-1})$ order. Setting $B_{\alpha\gamma}^C = B_\alpha^C + B_\gamma^C + B_{\alpha\gamma}^{\mathrm{cross},C}$ yields (17).

The constant B_η^C is generically nonzero under IV-PPML for the following reason. Although the pair-level fixed-effect FOC $\sum_t u_{ijt} = 0$ admits a closed-form solution $\hat{\eta}_{ij}(\beta, \alpha, \gamma)$ identical to that used by Weidner and Zylkin (2021, Section 3.2) for benchmark PPML, the resulting profile slope moment $n^{-1} \sum_g q_g \tilde{u}_g$ with $\tilde{u}_{ijt} = y_{ijt} - (\sum_s y_{ijs}) \vartheta_{ijt}$, $\vartheta_{ijt} = \mu_{ijt} / \sum_s \mu_{ijs}$, is no longer a multinomial MLE score whenever $q_g \neq w_g$. The alignment (11) that is used by Weidner and Zylkin (2021) to kill the T^{-1} contribution from $\hat{\eta}_{ij}$ fails under

IV-PPML, so the pair-family bias channel $\Delta_\eta^B + \Delta_\eta^C + \Delta_\eta^V$ of Lemma 2 does not vanish, yielding $B_\eta^C \neq 0$. Similarly, $B_{\alpha\gamma}^C \neq 0$ by the same reasoning applied to the directional families, as in Class B. □

Remark 5 (Comparison with benchmark PPML)**.** *For benchmark PPML, the profile step of Weidner and Zylkin (2021, Section 3.2) reduces the three-way likelihood to a multinomial profile, and the T^{-1} contribution from $\hat{\eta}_{ij}$ is zeroed out by the alignment* (11)*. Only the directional-effect term $B_{\alpha\gamma}^{C,\mathrm{PPML}}/N_c$ survives. This matches Weidner and Zylkin (2021, Proposition 3) and the "IPP bias (PPML)" row for Class C in Table 2.*

A.7 Standard-error bias

We now establish the third row of Table 2. The `ivppmlhdfe` analytical standard errors are constructed from the just-identified sandwich variance

$$\hat{V}(\hat{\beta}) = \frac{1}{n}\hat{W}_n^{-1}\,\hat{\Omega}_n\,\hat{W}_n^{-1\prime}, \quad \hat{W}_n = \frac{1}{n}\sum_g q_g\,\hat{\mu}_g\,\tilde{w}_g', \quad \hat{\Omega}_n = \frac{1}{n}\sum_g \hat{S}_g\hat{S}_g', \tag{18}$$

where $\tilde{w}_g$ is the two-way (or three-way) demeaned version of w_g implicit in the IRLS step of Correia et al. (2020), and $\hat{S}_g = q_g\,\hat{u}_g$ is the estimated slope-moment contribution for observation g.

Following the finite-sample analysis of the Poisson sandwich estimator in Weidner and Zylkin (2021, Section 3.3), the bias of $\hat{\Omega}_n$ as an estimator of the true "meat" matrix $\Omega_n \equiv n^{-1}\sum_g \mathbb{E}[S_g S_g']$ decomposes as

$$\mathbb{E}\left[\hat{\Omega}_n - \Omega_n\right] = -\underbrace{\mathbb{C}_\beta}_{\text{from } \mathrm{Var}(\hat{\beta})} - \underbrace{\mathbb{C}_\phi}_{\text{from } \mathrm{Var}(\hat{\phi})} + o(\text{leading}). \tag{19}$$

The first term $\mathbb{C}_\beta$ is the standard finite-sample downward bias of sandwich-type variance estimators and is of order n^{-1}. The second term $\mathbb{C}_\phi$ is the *IPP contribution* to the variance bias, and it is of the same leading order as the IPP bias of the point estimator, as we now show.

Proposition 5 (SE-bias order)**.** *Under Assumption 1, the leading-order bias of* $\hat{\Omega}_n$ *induced by fixed-effect estimation error is*

$$\|\mathbb{C}_\phi\| = O\left(\sum_{m=1}^{M} \frac{1}{h_m}\right), \tag{20}$$

and, consequently, the leading-order bias of $\hat{V}(\hat{\beta})$ *as an estimator of* $\mathrm{Var}(\hat{\beta})$ *is of the same order. Specializing to Classes A–C using the cell sizes* (12)–(16) *gives*

$$\mathrm{Bias}(\hat{\mathrm{SE}}_A) = O(T^{-1} + N^{-1}), \quad \mathrm{Bias}(\hat{\mathrm{SE}}_B) = O(N_c^{-1}), \quad \mathrm{Bias}(\hat{\mathrm{SE}}_C) = O(T^{-1} + N_c^{-1}).$$

Proof. Write $S_g = q_g u_g$ and $\hat{S}_g = q_g \hat{u}_g$, with $\hat{u}_g = u_g(\hat{\beta}, \hat{\psi}_g)$. A second-order expansion of $\hat{u}_g$ around $(\beta_0, \psi_{g,0})$ gives

$$\hat{u}_g - u_g = -\mu_g\left(w_g'(\hat{\beta} - \beta_0) + (\hat{\psi}_g - \psi_{g,0})\right) + O_p\left(\|\hat{\beta} - \beta_0\|^2 + (\hat{\psi}_g - \psi_{g,0})^2\right), \tag{21}$$

using (6). Squaring and taking expectation,

$$\begin{aligned} \mathbb{E}[\hat{S}_g \hat{S}_g'] - \mathbb{E}[S_g S_g'] = &- q_g q_g'\, \mu_g^2\, w_g'\, \mathbb{E}[(\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)']\, w_g \\ &- \; q_g q_g'\, \mu_g^2\, \mathbb{E}[(\hat{\psi}_g - \psi_{g,0})^2] + \text{higher-order}. \end{aligned} \tag{22}$$

Averaging over g and using Assumption 1(v), the first term on the right yields

$$-\mathbb{C}_\beta = O_p(\|\mathrm{Var}(\hat{\beta})\|) = O_p(n^{-1}),$$

while the second term yields

$$-\mathbb{C}_\phi = -\frac{1}{n}\sum_g q_g q_g'\, \mu_g^2\, \mathbb{E}[(\hat{\psi}_g - \psi_{g,0})^2] + o_p(\cdot).$$

From Lemma 1, $\hat{\psi}_g - \psi_{g,0} = \sum_{m:g\in\mathcal{F}_m}(\hat{\psi}_{m,c_m(g),0} - \psi_{m,c_m(g),0})$, so

$$\begin{aligned}\mathbb{E}\left[\left(\hat{\psi}_g - \psi_{g,0}\right)^2\right] &= \sum_m \mathbb{E}\left[\left(\hat{\psi}_{m,c_m(g),0} - \psi_{m,c_m(g),0}\right)^2\right] + \text{cross-family terms} \\ &= \sum_m \frac{\sigma^2_{m,c_m(g)}}{h_m} + O\left(\sum_{m<\ell}\frac{1}{\sqrt{h_m h_\ell}}\right),\end{aligned}$$

where $\sigma^2_{m,c} = \mathbb{E}[\zeta^2_{m,c}]$ is bounded by Assumption 1(v). Substituting back,

$$\|\mathbb{C}_\phi\| = O\left(\sum_m h_m^{-1}\right) + O\left(\sum_{m<\ell}(h_m h_\ell)^{-1/2}\right) = O\left(\sum_m h_m^{-1}\right),$$

using arithmetic mean-geometric mean inequality. This is (20).

Because $\hat{W}_n$ is a sample average of bounded functions of the estimated mean $\hat{\mu}_g$, standard arguments give $\hat{W}_n - W_n = O_p((\sum_m h_m^{-1})^{1/2})$, which contributes only a higher-order correction to $\hat{V}(\hat{\beta}) - V(\hat{\beta})$. Hence the leading-order bias of $\hat{V}$ is driven by $\mathbb{C}_\phi$ and inherits the order $\sum_m h_m^{-1}$.

The class-specific orders follow by substituting the cell sizes (12)–(16). □

Remark 6 (Interpretation)**.** *Proposition 5 shows that the estimated IV-PPML standard error is downward-biased by the same order as the IPP bias of the point estimator. The mechanism is identical to the two-way PPML bias documented by Weidner and Zylkin (2021, Section 3.3): the fitted scores $\hat{S}_g$ are "over-shrunk" toward zero because they have been orthogonalized to the fixed-effect spans whose parameters have been noisily estimated. In our IV-PPML setting, the source of the orthogonalization is the same fixed-effect system, and the resulting bias is of order $\sum_m h_m^{-1}$. For Classes A and C this translates to $O(T^{-1} + N^{-1})$; for Class B, to $O(N_c^{-1})$.*

A.8 Consistency, asymptotic bias, and the role of T

Propositions 2–4 have three immediate consequences for the asymptotic behavior of the IV-PPML-HDFE estimator. Since standard IV-PPML suffers an IPP bias whose order does not vanish under fixed-T or fixed-N asymptotics (unlike benchmark PPML, where the score cancellation often sends the leading-order bias to zero), the asymptotic framework itself must be strong enough to drive the bias to zero.

1. Consistency requires both N (or N_c) and T to diverge in Classes A and C. For Class A, (13) gives $\text{Bias}_{\text{IPP}}(\hat{\beta}_A) = B_\alpha^A/T + B_\gamma^A/N + o(\cdot)$. If T is held fixed while $N \to \infty$, the B_α^A/T component does *not* vanish, and $\hat{\beta}_A - \beta_0 \not\to 0$; the estimator is inconsistent. Symmetrically, $\hat{\beta}_A$ is inconsistent when N is held fixed and $T \to \infty$. Consistency therefore requires $N, T \to \infty$ jointly. For Class C, the same conclusion follows from (17) with (N, T) replaced by (N_c, T): both B_η^C/T and $B_{\alpha\gamma}^C/N_c$ must vanish, which requires $N_c \to \infty$ and $T \to \infty$ jointly.

2. Class B is consistent under fixed-T asymptotics. For Class B, (15) gives $\text{Bias}_{\text{IPP}}(\hat{\beta}_B) = B^B/N_c + o(N_c^{-1})$. Consistency requires only $N_c \to \infty$; T may be held fixed.

3. Even when the estimator is consistent, its asymptotic distribution is generally off-center. Consider Class A under the asymptotic framework $N, T \to \infty$ with $N/T \to \kappa^2$, $0 < \kappa < \infty$. The rescaled estimator satisfies

$$\begin{aligned}\sqrt{NT}\left(\hat{\beta}_A - \beta_0\right) &= \sqrt{NT}\left(\frac{B_\alpha^A}{T} + \frac{B_\gamma^A}{N}\right) + \sqrt{NT}O_p\left((NT)^{-1/2}\right) + o_p(1) \\ &\xrightarrow{d} \mathcal{N}\left(\kappa B_\alpha^A + \kappa^{-1}B_\gamma^A, V_A\right),\end{aligned}$$

for some variance V_A, because $\sqrt{NT}/T = \sqrt{N/T} \to \kappa$ and $\sqrt{NT}/N = \sqrt{T/N} \to \kappa^{-1}$. The asymptotic distribution is therefore *not* centered at the truth: it has an asymptotic bias of order one in the limit. The same phenomenon occurs for Class C under $N_c, T \to \infty$ with

$N_c/T \to \kappa^2$, and for Class B under $N_c \to \infty$ with T either fixed or jointly diverging. This is the large-N, T asymptotic-bias problem of Fernández-Val and Weidner (2016) and Weidner and Zylkin (2021), and it is the feature that motivates the split-panel jackknife corrections in Section 2.3. The SPJ formulas recenter the asymptotic distribution at β_0 by eliminating the leading $1/h_m$ bias contributions, while the pair bootstrap of Section 2.3 delivers valid inference that simultaneously corrects the downward SE bias of Proposition 5.

4. Summary. Collecting Propositions 2–4 and Proposition 5, we have proven the three rows of Table 2 associated with the IV-PPML estimator. The first row (benchmark PPML) follows from Fernández-Val and Weidner (2016, Theorem 4.1 and Remark 3) and Weidner and Zylkin (2021, Proposition 3), as discussed in Remark 2. Table 2 can therefore be restated compactly as follows: under IV-PPML, the leading IPP bias is $\sum_m h_m^{-1}$ for each class, and the leading SE bias inherits the same order. Under benchmark PPML, the score cancellation eliminates the contributions from every family whose fixed effect enters only through u_g and whose FOC admits a closed-form profile, leaving 0 for Classes A and B and $O(N_c^{-1})$ for Class C.

B Jackknife Bias Correction: Derivations

This appendix derives the split-panel jackknife (SPJ) bias-correction formulas presented in Section 2.3. Throughout, we maintain the notation of Section 2.1 and Appendix A: observations are indexed by $g = 1, \ldots, n$, with regressors $w_g = \left(x_g', d_g\right)'$, instruments $q_g = \left(x_g', z_g\right)'$, fixed-effect sum ψ_g, conditional mean $\mu_g = \exp\left(w_g'\beta + \psi_g\right)$, and residual $u_g = y_g - \mu_g$. The derivations extend the arguments of Fernández-Val and Weidner (2016) for two-way panel models and Weidner and Zylkin (2021) for three-way gravity structures to the IV-PPML setting.

B.1 General bias structure

Proposition 1 of Appendix A establishes that the leading-order incidental-parameter bias of the IV-PPML-HDFE estimator admits the family-wise expansion

$$\mathbb{E}\left[\hat{\beta}\right] - \beta_0 = \sum_{m=1}^{M} \frac{B_m}{h_m} + o\left(\sum_{m=1}^{M} h_m^{-1}\right), \tag{23}$$

where M is the number of fixed-effect families, h_m is the average number of observations identifying each parameter in family m, and B_m is a finite vector that depends on the data-generating process but not on the sample dimensions. This additive structure is the algebraic foundation of the SPJ correction. If one re-estimates β on a sub-panel that halves h_m while leaving the remaining cell sizes unchanged, the contribution of family m to the bias doubles from B_m/h_m to $2B_m/h_m$, while the other family contributions are preserved. An appropriate linear combination of the full-sample estimator and the sub-panel estimators therefore cancels every leading-order bias term simultaneously. The remainder of this appendix specializes this logic to each of the three model classes defined in Section 2.1, then records a simplified formula for specifications with a single interacted fixed-effect family.

B.2 Class A: individual and time fixed effects ($\alpha_i + \gamma_t$)

For Class A, $g = (i, t)$, $n = NT$, and the two fixed-effect families have cell sizes $h_\alpha = T$ and $h_\gamma = N$. Proposition 2 specializes (23) to

$$\mathbb{E}\left[\hat{\beta}\right] - \beta_0 = \frac{B_\alpha^A}{T} + \frac{B_\gamma^A}{N} + o\left(T^{-1} + N^{-1}\right), \tag{24}$$

where B_α^A collects the bias contribution from estimating the individual effects $\{\alpha_i\}_{i=1}^N$ and B_γ^A the contribution from the time effects $\{\gamma_t\}_{t=1}^T$.

Time split. Partition the sample into two temporal halves of length $T/2$ each. On either half, every individual effect α_i is identified from $T/2$ observations instead of T, while each time effect γ_t continues to be identified from N observations. The bias on either half therefore equals

$$\frac{B_\alpha^A}{T/2} + \frac{B_\gamma^A}{N} = \frac{2B_\alpha^A}{T} + \frac{B_\gamma^A}{N}.$$

Averaging across the two halves yields

$$\mathbb{E}\left[\bar{\beta}_{T/2}\right] - \beta_0 = \frac{2B_\alpha^A}{T} + \frac{B_\gamma^A}{N} + o\left(T^{-1} + N^{-1}\right).$$

Cross-sectional split. Randomly partition the N individuals into two groups of size $N/2$. On either group, every retained individual effect α_i is identified from the full T observations, while each time effect γ_t is identified from only $N/2$ observations. The bias on either half becomes

$$\frac{B_\alpha^A}{T} + \frac{B_\gamma^A}{N/2} = \frac{B_\alpha^A}{T} + \frac{2B_\gamma^A}{N}.$$

Averaging yields

$$\mathbb{E}\left[\bar{\beta}_{N/2}\right] - \beta_0 = \frac{B_\alpha^A}{T} + \frac{2B_\gamma^A}{N} + o\left(T^{-1} + N^{-1}\right).$$

Combined correction. Consider the linear combination

$$\hat{\beta}_{\mathrm{SPJ}} = c_0\hat{\beta} + c_1\bar{\beta}_{T/2} + c_2\bar{\beta}_{N/2}.$$

Unbiasedness at β_0 and cancellation of the two leading-order bias channels require

$$\begin{aligned} c_0 + c_1 + c_2 &= 1, \\ c_0 + 2c_1 + c_2 &= 0 \quad (\text{cancel } B_\alpha^A/T), \\ c_0 + c_1 + 2c_2 &= 0 \quad (\text{cancel } B_\gamma^A/N). \end{aligned}$$

The unique solution is $c_0 = 3$ and $c_1 = c_2 = -1$, which yields

$$\hat{\beta}_{\mathrm{SPJ}} = 3\hat{\beta} - \bar{\beta}_{T/2} - \bar{\beta}_{N/2}. \tag{25}$$

The bias of $\hat{\beta}_{\text{SPJ}}$ is $o\left(T^{-1} + N^{-1}\right)$, so both leading-order channels in (24) are eliminated.

B.3 Class B: two-way gravity fixed effects $(\alpha_{it} + \gamma_{jt})$

For Class B, $g = (i, j, t)$ with $i \neq j$, $n = N_c\,(N_c - 1)\,T$, and the two directional families share the common cell size $h_\alpha = h_\gamma = N_c - 1 \asymp N_c$. Proposition 3 specializes (23) to

$$\mathbb{E}\left[\hat{\beta}\right] - \beta_0 = \frac{B^B}{N_c} + o\left(N_c^{-1}\right), \tag{26}$$

where $B^B = B^B_\alpha + B^B_\gamma + B^B_{\alpha\gamma}$ collects the combined bias contributions from the exporter-time family $\{\alpha_{it}\}$, the importer-time family $\{\gamma_{jt}\}$, and the cross-family term. Because both directional families have the same cell-size order, the natural split operates on the country dimension.

Country split. Following Weidner and Zylkin (2021), randomly partition the N_c countries into two groups a and b of roughly equal size. Form the four directed sub-panels $(a \to a)$, $(a \to b)$, $(b \to a)$, and $(b \to b)$. In each sub-panel, the exporter pool and the importer pool are both restricted to a set of cardinality $N_c/2$, so that the within-panel cell sizes satisfy $h_\alpha = h_\gamma \asymp N_c/2$. Substituting this into (26) gives

$$\mathbb{E}\left[\hat{\beta}_s\right] - \beta_0 = \frac{2B^B}{N_c} + o\left(N_c^{-1}\right), \quad s \in \{aa, ab, ba, bb\}.$$

Averaging across the four sub-panels yields

$$\mathbb{E}\left[\bar{\beta}_{\text{country}}\right] - \beta_0 = \frac{2B^B}{N_c} + o\left(N_c^{-1}\right), \quad \bar{\beta}_{\text{country}} = \tfrac{1}{4}\left(\hat{\beta}_{aa} + \hat{\beta}_{ab} + \hat{\beta}_{ba} + \hat{\beta}_{bb}\right).$$

Combined correction. Setting $\hat{\beta}_{\mathrm{SPJ}} = c_0\hat{\beta} + c_1\bar{\beta}_{\mathrm{country}}$ and imposing $c_0 + c_1 = 1$ together with $c_0 + 2c_1 = 0$ gives $c_0 = 2$ and $c_1 = -1$:

$$\hat{\beta}_{\mathrm{SPJ}} = 2\hat{\beta} - \tfrac{1}{4}\left(\hat{\beta}_{aa} + \hat{\beta}_{ab} + \hat{\beta}_{ba} + \hat{\beta}_{bb}\right). \tag{27}$$

The bias of $\hat{\beta}_{\mathrm{SPJ}}$ is $o\left(N_c^{-1}\right)$.

Remark. The four-panel construction $\{(a \to a), (a \to b), (b \to a), (b \to b)\}$ has two useful properties. First, within each of the four sub-panels the exporter and importer pools are simultaneously restricted to a subset of size $N_c/2$, so that h_α and h_γ are halved in parallel and the within-panel bias admits the clean representation $2B^B/N_c$ on which the correction in (27) relies. Second, the four sub-panels exactly partition the directed country-pair space, so that averaging over them pools all $N_c\left(N_c - 1\right)T$ observations and produces a sub-panel-averaged estimator whose variance is of the same order as that of $\hat{\beta}$. A restricted two-panel construction such as $\{(a \to a), (b \to b)\}$ would share the first property but not the second, since it would discard the cross-group observations and therefore inflate the variance of $\bar{\beta}_{\mathrm{country}}$.

B.4 Class C: three-way gravity fixed effects ($\alpha_{it} + \gamma_{jt} + \eta_{ij}$)

For Class C, $g = (i, j, t)$ with $i \neq j$, $n = N_c\left(N_c - 1\right)T$, and the three fixed-effect families have cell sizes $h_\alpha = h_\gamma \asymp N_c$ and $h_\eta = T$. Proposition 4 specializes (23) to

$$\mathbb{E}\left[\hat{\beta}\right] - \beta_0 = \frac{B^C_{\alpha\gamma}}{N_c} + \frac{B^C_\eta}{T} + o\left(N_c^{-1} + T^{-1}\right), \tag{28}$$

where $B^C_{\alpha\gamma}$ is the combined contribution of the directional families $\{\alpha_{it}\}$ and $\{\gamma_{jt}\}$ (absorbing the directional cross-family term as in Class B), and B^C_η is the contribution of the pair family $\{\eta_{ij}\}_{i \neq j}$. The bias has two channels, one operating through N_c and one through T, and the SPJ correction must cancel both.

Eight-panel construction. We cross the Class B country split with a Class A-style time split. The country split randomly partitions the N_c countries into two groups a and b of roughly equal size and produces four directed sub-panels as in (27). The time split partitions the T periods into two halves of length $T/2$. Combining the two splits yields $4 \times 2 = 8$ cells. We define

- $\bar{\beta}_{\text{country}}$: the average of $\hat{\beta}$ across the four country sub-panels, each estimated on the full time span;
- $\bar{\beta}_{\text{time}}$: the average of $\hat{\beta}$ across the two temporal halves, each estimated on the full country set;
- $\bar{\beta}_{\text{8cell}}$: the average of $\hat{\beta}$ across all eight country-by-time cells.

Biases of the four ingredients. The country split halves h_α and h_γ while leaving h_η unchanged; the time split halves h_η while leaving h_α and h_γ unchanged; and the product split halves all three cell sizes. Substituting these cell-size changes into (28) gives

$$\begin{aligned}
\mathbb{E}\left[\hat{\beta}\right] - \beta_0 &= \frac{B^C_{\alpha\gamma}}{N_c} + \frac{B^C_{\eta}}{T} + o\left(N_c^{-1} + T^{-1}\right), \\
\mathbb{E}\left[\bar{\beta}_{\text{country}}\right] - \beta_0 &= \frac{2B^C_{\alpha\gamma}}{N_c} + \frac{B^C_{\eta}}{T} + o\left(N_c^{-1} + T^{-1}\right), \\
\mathbb{E}\left[\bar{\beta}_{\text{time}}\right] - \beta_0 &= \frac{B^C_{\alpha\gamma}}{N_c} + \frac{2B^C_{\eta}}{T} + o\left(N_c^{-1} + T^{-1}\right), \\
\mathbb{E}\left[\bar{\beta}_{\text{8cell}}\right] - \beta_0 &= \frac{2B^C_{\alpha\gamma}}{N_c} + \frac{2B^C_{\eta}}{T} + o\left(N_c^{-1} + T^{-1}\right).
\end{aligned}$$

Combined correction. We seek a linear combination

$$\hat{\beta}_{\text{SPJ}} = c_0\hat{\beta} + c_1\bar{\beta}_{\text{country}} + c_2\bar{\beta}_{\text{time}} + c_3\bar{\beta}_{\text{8cell}}$$

that is unbiased at β_0 and that cancels the two leading-order bias channels:

$$\begin{aligned} c_0 + c_1 + c_2 + c_3 &= 1, \\ c_0 + 2c_1 + c_2 + 2c_3 &= 0 \quad (\text{cancel } B^C_{\alpha\gamma}/N_c), \\ c_0 + c_1 + 2c_2 + 2c_3 &= 0 \quad (\text{cancel } B^C_{\eta}/T). \end{aligned}$$

With three equations in four unknowns, the system is underdetermined and admits a one-parameter family of solutions. Subtracting the third equation from the second yields $c_1 = c_2$, which ensures that the correction treats the two bias channels symmetrically. Setting $c_3 = 1$, which preserves the inclusion-exclusion interpretation discussed below, yields $c_0 = 4$ and $c_1 = c_2 = -2$:

$$\hat{\beta}_{\text{SPJ}} = 4\hat{\beta} - 2\bar{\beta}_{\text{country}} - 2\bar{\beta}_{\text{time}} + \bar{\beta}_{\text{8cell}}. \tag{29}$$

The bias of $\hat{\beta}_{\text{SPJ}}$ is $o\left(N_c^{-1} + T^{-1}\right)$, so both leading-order channels in (28) are eliminated.

Inclusion-exclusion interpretation. Formula (29) admits a natural inclusion-exclusion reading. The full-sample estimator $\hat{\beta}$ is corrected by subtracting, with weight two, the excess bias introduced by halving along each dimension separately (through $\bar{\beta}_{\text{country}}$ and $\bar{\beta}_{\text{time}}$), and the overlap that is double-counted by these two subtractions is restored through the eight-cell average $\bar{\beta}_{\text{8cell}}$. This is the three-way analog of the two-way inclusion-exclusion formula (25) used in Class A.

B.5 Interacted fixed effects: cross-sectional-only SPJ

When the fixed-effect structure reduces to a single interacted family, such as city $\times$ cohort $\times$ day in the replication of Chen et al. (2025), the p/n expansion (23) specializes to a single term

$$\mathbb{E}\left[\hat{\beta}\right] - \beta_0 = \frac{B}{h} + o\left(h^{-1}\right),$$

where B is the single-family bias constant and h is the average number of observations per interacted cell. Because each interacted group pools observations across the cross-sectional units within that group, the relevant splitting dimension is the cross-section N. A single cross-sectional split therefore suffices, and the coefficients $c_0 = 2$ and $c_1 = -1$ solve $c_0+c_1 = 1$ and $c_0 + 2c_1 = 0$ exactly as in Class B, giving

$$\hat{\beta}_{\text{SPJ}} = 2\hat{\beta} - \bar{\beta}_{N/2}, \tag{30}$$

where $\bar{\beta}_{N/2}$ averages the estimator across two random cross-sectional halves. This is the formula used in the Chen et al. (2025) replication reported in Section 4.1, where the split is on drivers.

B.6 Practical considerations

Random splits. The country partition used in Classes B and C and the individual partition used in Class A both rely on a random draw, which introduces additional variability in the SPJ point estimate relative to a hypothetical deterministic split. In principle, one should average across multiple random partitions to reduce this variability. In our Monte Carlo simulations, we use a single random partition per replication for computational tractability. Weidner and Zylkin (2021) average across 200 random partitions in their empirical work, a choice that we revisit in the robustness checks.

Uneven dimensions. When N, T, or N_c is odd, the two halves differ by one unit. The SPJ formulas remain valid provided both halves are large enough for estimation, since the $O\left(h_m^{-1}\right)$ bias characterization underlying (23) holds for any split ratio bounded away from zero and one.

Singleton dropping. Splitting the sample can create singleton fixed-effect groups, that is, cells with only one observation, that were not singletons in the full sample. Following standard practice, such groups are dropped before estimation, which reduces the effective sample size within each sub-panel. If too many observations are lost in this way, the sub-panel estimates become unstable and the SPJ correction can fail. This is the mechanism behind the infeasibility of SPJ in Applications 2 and 3 reported in Section 4.

C Monte Carlo Design and Full Results

C.1 Data generating process

The outcome is generated as

$$y_g = \exp\left(x_{1g}\beta_1 + x_{2g}\beta_2 + \psi_g\right) v_g,$$

where $x_{2g} \sim \mathcal{N}(0,1)$ is exogenous and ψ_g collects the fixed effects specific to each model class (see below).

Endogeneity. The endogenous regressor x_{1g} and the multiplicative error v_g are linked through a shared shock e_g:

$$\begin{aligned} x_{1g} &= \pi_z\, z_g + e_g, \quad z_g \sim \text{i.i.d. } \mathcal{N}(0,1), \quad e_g \sim \text{i.i.d. } \mathcal{N}(0,1), \\ \ln v_g &= -\tfrac{\sigma_v^2}{2} + \sigma_v\left(\rho_{ev}\, e_g + \sqrt{1-\rho_{ev}^2}\, u_g\right), \quad u_g \sim \text{i.i.d. } \mathcal{N}(0,1). \end{aligned}$$

Because e_g enters both x_{1g} and $\ln v_g$, we have $\operatorname{Cov}(x_1, v) \neq 0$ whenever $\rho_{ev} \neq 0$. The instrument z_g is independent of (e_g, u_g, v_g) and all fixed effects, satisfying the exclusion restriction.

Fixed-effect classes.

- **Class A** $(\alpha_i+\gamma_t)$: individual and time FE. $\alpha_i = \sigma_\alpha\left(\rho_{\alpha x}\, \bar{x}_{1i} + \sqrt{1-\rho_{\alpha x}^2}\, \xi_i\right)$, correlated with x_1. Grid: $N \in \{50, 100, 200, 500\}$, $T \in \{10, 20, 30\}$.

- **Class B** ($\alpha_{it} + \gamma_{jt}$): exporter-year and importer-year FE (two-way gravity). Grid: $N_c \in \{20, 30, 40, 50\}$, $T \in \{10, 20, 30\}$.
- **Class C** ($\alpha_{it}+\gamma_{jt}+\eta_{ij}$): pair, exporter-year, and importer-year FE (three-way gravity). Grid: $N_c \in \{20, 30, 40, 50\}$, $T \in \{10, 20, 30\}$.

Parameter values. $\beta_1 = 0.5$, $\beta_2 = 0.3$, $\sigma_v = 1.5$, $\sigma_\alpha = 0.3$, $\rho_{\alpha x} = 0.4$, $\rho_{ev} = 0.5$, $\pi_z = 0.8$. R = 1,000 replications per cell. For the exogenous PPML benchmark, we set $\rho_{ev} = 0$ and estimate with `ppmlhdfe` using robust SE.

C.2 Full results: `ivppmlhdfe` vs. `ppmlhdfe`

Table C1: Monte Carlo: `ivppmlhdfe` vs. `ppmlhdfe` ($T = 10$, $\beta_1 = 0.5$)

				`ivppmlhdfe`			`ppmlhdfe` (w/ endog.)			`ppmlhdfe` (no endog.)		
Class	FE structure	N	T	Bias	SE/SD	Cov	Bias	SE/SD	Cov	Bias	SE/SD	Cov
A	$\alpha_i + \gamma_t$	50	10	−.061	0.35	81.5	.446	0.57	0.6	−.007	0.66	80.5
		100	10	−.031	0.68	82.7	.449	0.60	0.0	−.004	0.74	85.2
		200	10	−.042	0.68	81.0	.447	0.58	0.0	−.006	0.74	86.4
		500	10	−.028	0.71	83.2	.454	0.61	0.0	−.002	0.75	85.1
B	$\alpha_{it} + \gamma_{jt}$	20	10	−.031	0.64	75.3	.448	0.54	0.0	−.004	0.66	78.8
		30	10	−.026	0.70	78.5	.448	0.58	0.0	−.004	0.73	84.9
		40	10	−.021	0.70	79.0	.448	0.60	0.0	−.002	0.74	84.5
		50	10	−.017	0.74	80.5	.450	0.62	0.0	−.001	0.78	88.0
C	$\alpha_{it} + \gamma_{jt} + \eta_{ij}$	20	10	−.036	0.59	68.6	.452	0.55	0.0	−.004	0.60	76.3
		30	10	−.035	0.60	59.4	.451	0.56	0.0	−.003	0.65	79.1
		40	10	−.033	0.61	54.0	.450	0.53	0.0	−.004	0.63	77.8
		50	10	−.034	0.65	46.0	.449	0.56	0.0	−.004	0.68	80.3

Notes. See Table 3 notes. R = 1,000 replications per cell and B = 1,000 pair-bootstrap draws per replication.

Table C2: Monte Carlo: `ivppmlhdfe` vs. `ppmlhdfe` ($T = 30$, $\beta_1 = 0.5$)

				`ivppmlhdfe`			`ppmlhdfe` (w/ endog.)			`ppmlhdfe` (no endog.)		
Class	FE structure	N	T	Bias	SE/SD	Cov	Bias	SE/SD	Cov	Bias	SE/SD	Cov
A	$\alpha_i + \gamma_t$	50	30	−.026	0.71	83.3	.442	0.60	0.1	−.006	0.75	87.0
		100	30	−.024	0.72	83.5	.450	0.58	0.0	−.002	0.73	85.5
		200	30	−.019	0.76	87.0	.451	0.63	0.0	−.001	0.78	87.5
		500	30	−.014	0.79	86.8	.453	0.65	0.0	−.000	0.83	88.8
B	$\alpha_{it} + \gamma_{jt}$	20	30	−.026	0.66	71.2	.452	0.56	0.0	−.001	0.69	81.3
		30	30	−.024	0.66	65.8	.448	0.55	0.0	−.003	0.69	81.8
		40	30	−.019	0.72	71.1	.449	0.57	0.0	−.002	0.73	84.0
		50	30	−.017	0.69	68.9	.450	0.62	0.0	−.001	0.76	86.7
C	$\alpha_{it} + \gamma_{jt} + \eta_{ij}$	20	30	−.036	0.63	56.2	.445	0.56	0.0	−.006	0.65	77.1
		30	30	−.031	0.64	47.1	.443	0.54	0.0	−.006	0.64	75.5
		40	30	−.030	0.65	39.3	.443	0.55	0.0	−.006	0.69	76.1
		50	30	−.029	0.65	30.5	.443	0.56	0.0	−.005	0.69	76.8

Notes. See Table 3 notes. $R = 1{,}000$ replications per cell and $B = 1{,}000$ pair-bootstrap draws per replication.

C.3 Full results: SPJ with bootstrap

Table C3: IV-PPML vs. SPJ with bootstrap SE ($T = 10$)

Class	N	T	Estimator	Bias	Bias red.	SD	SE	SE/SD	Cov (%)
A	50	10	IV-PPML	−.061		.420	.147	0.35	81.5
			SPJ	−.017	73%	.301	.210	**0.70**	**95.0**
	100	10	IV-PPML	−.031		.141	.096	0.68	82.7
			SPJ	−.007	77%	.179	.147	**0.82**	**94.5**
	200	10	IV-PPML	−.042		.112	.077	0.68	81.0
			SPJ	−.027	36%	.145	.110	**0.75**	**92.2**
	500	10	IV-PPML	−.028		.076	.054	0.71	83.2
			SPJ	−.010	63%	.094	.077	**0.83**	**92.6**
B	20	10	IV-PPML	−.031		.070	.044	0.64	75.3
			SPJ	−.025	20%	.096	.099	**1.03**	**98.2**
	30	10	IV-PPML	−.026		.049	.034	0.70	78.5
			SPJ	−.016	38%	.064	.059	**0.93**	**96.0**
	40	10	IV-PPML	−.021		.040	.028	0.70	79.0
			SPJ	−.011	49%	.050	.044	**0.88**	**92.3**
	50	10	IV-PPML	−.017		.033	.025	0.74	80.5
			SPJ	−.006	65%	.041	.036	**0.88**	**93.8**
C	20	10	IV-PPML	−.036		.058	.034	0.59	68.6
			SPJ	−.039	−10%	.097	.116	**1.19**	**99.3**
	30	10	IV-PPML	−.035		.040	.024	0.60	59.4
			SPJ	−.031	12%	.061	.065	**1.06**	**95.4**
	40	10	IV-PPML	−.033		.032	.019	0.61	54.0
			SPJ	−.024	26%	.047	.046	**0.97**	**92.8**
	50	10	IV-PPML	−.034		.025	.016	0.65	46.0
			SPJ	−.025	27%	.037	.036	**0.98**	**89.9**

Notes. See Table 4 notes. R = 1,000 replications per cell and B = 1,000 pair-bootstrap draws per replication.

Table C4: IV-PPML vs. SPJ with bootstrap SE ($T = 30$)

Class	N	T	Estimator	Bias	Bias red.	SD	SE	SE/SD	Cov (%)
A	50	30	IV-PPML	−.026		.119	.085	0.71	83.3
			SPJ	−.005	82%	.150	.114	**0.76**	**92.5**
	100	30	IV-PPML	−.024		.095	.068	0.72	83.5
			SPJ	−.008	68%	.114	.087	**0.76**	**91.4**
	200	30	IV-PPML	−.019		.071	.054	0.76	87.0
			SPJ	−.007	62%	.085	.067	**0.79**	**92.2**
	500	30	IV-PPML	−.014		.048	.038	0.79	86.8
			SPJ	−.004	74%	.055	.046	**0.85**	**92.2**
B	20	30	IV-PPML	−.026		.039	.026	0.66	71.2
			SPJ	−.020	23%	.057	.057	**1.00**	**97.0**
	30	30	IV-PPML	−.024		.030	.020	0.66	65.8
			SPJ	−.016	35%	.039	.035	**0.88**	**90.0**
	40	30	IV-PPML	−.019		.023	.016	0.72	71.1
			SPJ	−.009	52%	.029	.026	**0.89**	**89.0**
	50	30	IV-PPML	−.017		.021	.014	0.69	68.9
			SPJ	−.008	53%	.025	.021	**0.83**	**85.2**
C	20	30	IV-PPML	−.036		.034	.021	0.63	56.2
			SPJ	−.035	2%	.052	.060	**1.15**	**96.2**
	30	30	IV-PPML	−.031		.024	.016	0.64	47.1
			SPJ	−.024	24%	.035	.035	**0.99**	**86.5**
	40	30	IV-PPML	−.030		.019	.012	0.65	39.3
			SPJ	−.020	33%	.026	.025	**0.98**	**80.6**
	50	30	IV-PPML	−.029		.016	.010	0.65	30.5
			SPJ	−.018	38%	.022	.020	**0.92**	**74.9**

Notes. See Table 4 notes. $R = 1{,}000$ replications per cell and $B = 1{,}000$ pair-bootstrap draws per replication.

D Implementation and Stress Tests

We provide three parallel implementations of `ivppmlhdfe`. The primary implementation is a Stata/Mata command that builds on `reghdfe` and `ivreg2` for the inner linear solves and follows the `ppmlhdfe` architecture for fixed-effect absorption. A companion native Julia package implements the same algorithm using `FixedEffects.jl` for the absorption step. It is the preferred backend for large-scale Monte Carlo work because it avoids per-call startup costs. A third option is a Stata-to-Julia bridge that exposes the native Julia backend through the Stata command line, so users who prefer Stata syntax can still obtain the Julia speed when the dataset is large.

To ensure reliability, we run an extensive stress-test suite that exercises separation detection, collinearity handling, convergence, and edge cases in the fixed-effect structure. Following `ppmlhdfe`, we detect and drop groups in which all outcomes are zero before estimation, and we apply the standard separation-detection passes to remove observations and regressors that would otherwise produce a degenerate likelihood. We further detect and drop collinear regressors and instruments, including factor-variable dummies that are fully absorbed by the fixed effects. Non-convergence and runaway divergence are raised as explicit errors with clear diagnostic messages rather than returning silently incorrect estimates.

The stress-test suite includes (i) bit-identical regression checks against `ppmlhdfe` on the same data without endogeneity, (ii) cross-backend consistency checks between the Stata and Julia implementations, and (iii) edge-case tests with singleton fixed-effect groups, weakly identified instruments, and cascading separation patterns. Results match `ppmlhdfe` to working precision on all common specifications, and the two backends agree to within eight significant digits across the full test grid.