Persistent Zeros: The Extensive Margin of Trade*

Julian Hinz† Amrei Stammann‡ Joschka Wanner§

October 13, 2019

Abstract

The extensive margin of bilateral trade exhibits a high level of persistence that cannot be explained by geography or trade policy. We combine a heterogeneous firms model of international trade with bounded productivity with features from the firm dynamics literature to derive expressions for an exporting country's participation in a specific destination market in a given period. The model framework asks for a dynamic binary choice estimator with two or three sets of high-dimensional fixed effects. To mitigate the incidental parameter problem associated with nonlinear fixed effects models, we characterize and implement suitable bias corrections. Extensive simulation experiments confirm the desirable statistical properties of the bias-corrected estimators. Empirically, taking two sources of persistence — true state dependence and unobserved heterogeneity — into account using a dynamic specification, along with appropriate fixed effects and bias corrections, changes the estimated effects considerably: out of the most commonly studied potential determinants (joint WTO membership, common regional trade agreement, and shared currency), only sharing a common currency retains a significant effect on whether two countries trade with each other at all in our preferred estimation.

JEL Classification Codes: C13, C23, C55, F14, F15

Key Words: Dynamic binary choice, extensive margin, high-dimensional fixed effects, incidental parameter bias correction, trade policy

---

**Acknowledgements*: Will be added. *Note*: An R implementation of the estimators developed in this paper will be provided on CRAN and is currently available from the authors upon request.

†*Affiliation:* Kiel Institute for the World Economy & Kiel Centre for Globalization. *Address:* Kiellinie 66, 24105 Kiel, Germany

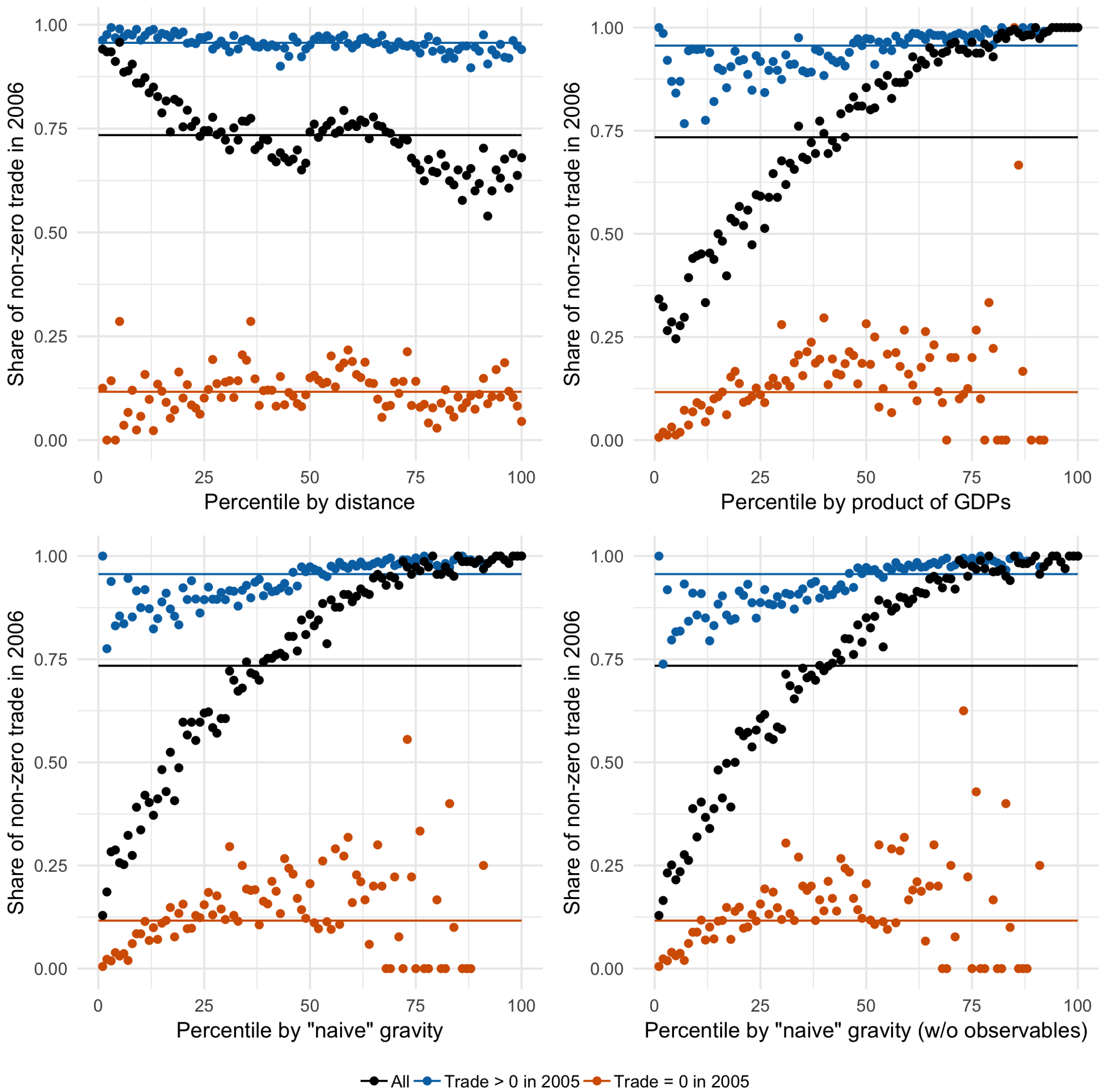
‡*Affiliation:* Heinrich-Heine-University Düsseldorf. *Address:* Universitätsstraße 1, 40225 Düsseldorf, Germany.

§*Affiliation:* University of Bayreuth. *Address:* Universitätsstraße 30, 95447 Bayreuth, Germany.

1 Introduction

What induces country pairs to trade? In 2006, still more than one quarter of potential bilateral trade relations reported zero trade flows. Figure 1 breaks down the share of nonzero trade flows in 2006 along the percentiles of four different ad-hoc indicators of "trade potential": bilateral distance; product of GDPs; "naive" gravity, i.e. the product of GDPs divided by their bilateral distance; and the latter when excluding country pairs in FTAs, with common currencies or common colonial history. The x-axis indicates the potential trade volume, i.e. the joint economic size and/or proximity of any two countries. All four plots paint a common picture: the black dots, covering all country

Figure 1: Determinants of the Extensive Margin of Trade — Gravity and Persistence.



pairs, show a strong general relationship between trade potential and actual nonzero trade. The blue and red dots split the country pairs according to whether the two did or did not engage in trade in the previous year. The clearly separated pattern for the two groups highlights a remarkable persistence of trade relations, even after controlling for differences in trade potential in terms of distance, size, and bilateral trade policy. More than 75 percent of those country pairs in the lowest percentile of trade potential trade again in 2006, provided they already did so in 2005. On the other hand, even comparably large and close pairs are likely not to trade in 2006 if they did not trade in 2005 either.[1,2]

In this paper we examine the determinants of the extensive margin of international trade, explicitly taking its persistence into account. We combine a heterogeneous firms model of international trade with bounded productivity with features from the firm dynamics literature to derive expressions for an exporting country's participation in a specific destination market in a given period. These expressions depend on partly unobserved (i) exporter-time, (ii) destination-time, and (iii) exporter-destination specific components, as well as on (iv) whether the exporter has already served the market in the previous period, and on (v) exporter-destination-time specific gravity-type trade cost determinants. We estimate the model making use of recent advances in the estimation of binary choice estimators with high-dimensional fixed effects to address (i)-(iii). The inclusion of fixed effects in a binary choice setting induces an incidental parameter problem that is potentially aggravated by the dynamics introduced by (iv). To mitigate this bias, we characterize and implement new analytical and jackknife bias corrections for coefficients and estimates of average partial effects in our specifications with two- and three-way fixed effects. Extensive simulation experiments demonstrate the desirable statistical properties of our proposed bias-corrected two- and three-way fixed effect logit and probit estimators. The empirical results provide evidence that both unobserved bilateral factors and true state dependence due to entry dynamics contribute strongly to the high persistence. Taking this persistence into account changes the coefficients considerably: out of the most commonly studied potential determinants (joint WTO membership, common regional trade agreement, and shared currency), only sharing a common currency has a significant effect on whether two countries trade with each

[1]Note that throughout the paper, "country pair" refers to a *directed* pair of countries, i.e. Germany-France and France-Germany are two distinct country pairs.

[2]The years 2005–2006 are the last available in our data set. A very similar pattern emerges for other points in time (see Figure 6 in Appendix A where the same graph is reproduced for the years 1990–1991). If longer time intervals are considered, a similar picture remains, but the relationship becomes considerably weaker (see Figure 7 in Appendix A for the years 1997–2006).

other at all.

Our paper builds on recent insights from three flourishing strands of literature. First, our paper is related to the literature on the extensive margin of international trade. A number of theoretical frameworks have sought to propose mechanisms behind the decisions of firms to export, and their aggregate implications of zero or nonzero trade flows at the country pair level. Analogous to the intensive margin counterpart, these theories have established *gravity*-like determinants, such as two countries' bilateral distance, a free trade agreement, a common currency and joint membership in the WTO. Egger and Larch (2011) and Egger, Larch, Staub, and Winkelmann (2011) append an extensive margin to an Anderson and Wincoop (2003)-type model by assuming export participation to be determined by (homogeneous) firms weighing operating profits and bilateral fixed costs of exporting. This results in a two-part model in which, given a country's participation in exporting to any given destination, trade flows follow structural gravity. Helpman, Melitz, and Rubinstein (2008) build a model of international trade with heterogeneous firms. Here, the volume of trade between two countries can change either because incumbent firms expand their operations, or because of new competitors entering into a market. Eaton, Kortum, and Sotelo (2013) move away from the arguably simplifying notion of a continuum of firms and develop a model of a finite set of heterogeneous firms. Here, no firm may export to a given market because of their individual efficiency draws. Our model proposed in this paper directly builds on Helpman, Melitz, and Rubinstein (2008) and extends it by features from the literature on firm dynamics. In this firm-level literature, Das, Roberts, and Tybout (2007) develop a dynamic discrete-choice model in which current export participation depends on previous exporting, and hence sunk costs, and observable characteristics of profits from exporting (in line with previous empirical evidence by Roberts and Tybout, 1997; Bernard and Jensen, 2004). Alessandria and Choi (2007) extend this line of research and develop a general equilibrium framework that takes sunk costs and "period-by-period" fixed costs into account, showing that, contrary to previous partial equilibrium evidence, aggregate effects are negligible for the US. More recent works have looked at *new* exporter dynamics (Ruhl and Willis, 2017), emphasizing that sunk costs may be relatively smaller and continuation costs relatively larger than previously assumed. Bernard, Bøler, Massari, Reyes, and Taglioni (2017) stand somewhat in contrast to this finding, showing that first and second year growth rates may suffer from a bias as a result of different entry dates throughout the year. Berman, Rebeyrol, and Vicard (2019) note the important role of "demand learning" and firms' updating of their future demand and market participation. In a similar vein, Piveteau (2019) develops a model in which new firms accumulate consumers — or fail to do so — determining entry and exit. While these newer models feature rich firm-level predictions, they require tailor-made econometric models for their estimation. Our model abstracts from the specific role of *new* firms and has the advantage of yielding an econometric specification and demanding an estimator that remains general and flexible to be applied in other contexts.

Second, our paper builds on advances in the literature on the gravity equation and the *intensive* margin of international trade. With the advent of what has now been coined *structural* gravity (Head and Mayer, 2014), the gravity framework has gained rich microfoundations. Anderson and Wincoop (2003) and Eaton and Kortum (2002) each formulate an underlying structure for exporting and importing countries that in estimations can easily be captured by appropriate two-way country(-time) fixed effects, as first noted by Feenstra (2004) and Redding and Venables (2004). Although not theoretically motivated, since Baier and Bergstrand (2007) it has furthermore become standard to include country pair fixed effects to tackle unobservable bilateral characteristics. Estimating the model introduced in this paper similarly calls for *at least* two sets of fixed effects, specific to exporters and importers in a given year. Additionally, and following Baier and Bergstrand (2007), there is no reason to believe that bilateral unobservables should not be a problem in the context of the extensive margin. Our preferred estimation of the model thus includes the "full set" of fixed effects that has become standard in the estimation of gravity models of the intensive margin of trade: exporter-year, importer-year and bilateral fixed effects that leave only bilateral-time-specific variation for the estimation of parameters of interest.

Third, the paper builds on and contributes to the literature on the econometrics of generalized linear models (GLMs) with fixed effects. Recent advances in this literature have made it possible to go beyond ordinary linear models in the context of high-dimensional fixed effects by providing fast and feasible algorithms (see Guimarães and Portugal (2010), Stammann (2018), and Hinz, Hudlet, and Wanner (2019)).[3] As known since Neyman and Scott (1948), the inclusion of fixed effects potentially introduces an incidental parameter problem, leading to inconsistent estimates. In the last few years, there have been a number of advances to deal with this problem, and

[3] Stammann, Heiß, and McFadden (2016) have shown in the context of binary choice models with individual fixed effects that a weighted version of the Frisch-Waugh-Lovell theorem (Frisch and Waugh (1933), Lovell (1963)) can be incorporated in a standard Newton-Raphson optimization procedure. This result paved the way to derive a computationally efficient algorithm for all GLMs with high-dimensional multi-way fixed effects (see Stammann (2018)). More recently, Hinz, Hudlet, and Wanner (2019) offer a different way to partial out fixed effects using a modification of the Gauss-Seidel algorithm proposed by Guimarães and Portugal (2010).

a variety of approaches have been proposed (see Fernández-Val and Weidner (2018) for a recent overview). Fernández-Val and Weidner (2016) develop analytical and jackknife bias corrections for nonlinear maximum likelihood estimators in static and dynamic models with individual and time effects for structural parameters and average partial effects. In Fernández-Val and Weidner (2018) they generalize their previous findings and show that the order of the bias induced by fixed effects in a wide family of models translates into a simple heuristic p/n, with n being the sample size and p the number of estimated parameters. Recently, Czarnowske and Stammann (2019) show how analytical bias corrections can be efficiently implemented in a high-dimensional fixed effects setting using the methods described by Stammann (2018). Our paper is complementary to computational and econometric contributions on the estimation of the intensive margin of trade. Larch, Wanner, Yotov, and Zylkin (2019) present a feasible procedure to estimate pseudo-poisson (PPML) models with three high-dimensional fixed effects. Correia, Guimarães, and Zylkin (2019) generalize this estimation procedure to arbitrary sets of fixed effects. Weidner and Zylkin (2019) investigate the incidental parameter problem in three-way fixed effects PPML models under fixed T asymptotics and suggest an appropriate jackknife bias correction. We contribute to this literature by characterizing and implementing analytical and jackknife bias corrections for our specific two- and three-way fixed effects in the context of binary choice models. This helps us mitigate the bias induced by estimating our theory-consistent model, requiring exporter-time (it), importer-time (jt), and in our preferred specification bilateral fixed effects (ij).

The remainder of the paper is structured as follows. In Section 2 we build a dynamic model of the extensive margin of international trade. The model yields aggregate predictions that can be structurally estimated using a probit model with high-dimensional fixed effects. In Section 3 we describe the estimator and bias correction procedure. We show its performance in Monte Carlo simulations in Section 4, before finally estimating the theoretical model in Section 5. Section 6 concludes.

2 An Empirical Model of the Extensive Margin of Trade

As a theoretical foundation for our econometric specification, we consider a stylized dynamic Melitz (2003)-type heterogeneous firms model of international trade. Following Helpman, Melitz, and Rubinstein (2008, henceforth HMR) we assume a bounded productivity distribution, like a truncated Pareto in HMR's case. We deviate from HMR by explicitly stating a time dimension and, unlike in the standard Melitz setting, separate fixed exporting costs into costs of entering a new market and costs of selling in a given market (as in Alessandria and Choi, 2007; Das, Roberts, and Tybout, 2007).

There are N countries, indexed by i and j, each of which consumes and produces a continuum of products. The representative consumer in j receives utility according to a CES utility function:

$$u_{jt} = \left(\int_{\omega \in \Omega_{jt}} (\xi_{ijt})^{\frac{1}{\sigma}} q_{jt}(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}} \quad \text{with} \quad \sigma > 1. \tag{1}$$

where $q_{jt}(\omega)$ is j's consumption of product ω in period t, Ω_{jt} is the set of products available in j, σ is the elasticity of substitution across products, and ξ_{ijt} is a log-normally distributed idiosyncratic demand shock (with $\mu_\xi = 0$ and $\sigma_\xi = 1$) for goods from country i in country j and period t (similar to Eaton, Kortum, and Kramarz, 2011). Demand in country j for good ω depends on this demand shock, j's overall expenditure E_{jt}, and the good price $p_{jt}(\omega)$ relative to the overall price level as captured by the price index P_{jt}:

$$q_{jt}(\omega) = \frac{p_{jt}(\omega)^{-\sigma}}{P_{jt}^{1-\sigma}} \xi_{ijt} E_{jt}.$$

$$\text{with} \qquad P_{jt} = \left(\int_{\omega \in \Omega_{jt}} \xi_{ijt} p_{jt}(\omega)^{1-\sigma} d\omega \right)^{\frac{1}{1-\sigma}}.$$

Each country has a fixed continuum of potentially active firms that have different productivities drawn from the distribution $G_{it}(\varphi)$, where $\varphi \in (0, \varphi_{it}^*]$. The productivity distribution evolves over time and firms' ranks within the productivity distribution can also change from period to period, though firms that in the last period did not export to a market already served by a domestic competitor are assumed not to directly jump to being the country's most productive firm in the next period.[4] Each period, a firm can decide to pay a fixed cost f_{it}^{prod} and start production of a differentiated variety using labour l as its only input, such that $l_t(\omega) = f_{it}^{prod} + q_t(\omega)/\varphi_t(\omega)$. A firm's marginal cost of providing one unit of its good to market j consists of iceberg trade costs τ_{ijt} and labour costs $w_{it}/\varphi_t(\omega)$. Firms compete with each other in monopolistic competition and charge a constant markup over marginal costs. Therefore, the price of a good ω produced in i and sold in j is:

[4]Note that we could in principle also allow for new firm entry into the pool of potential producers without changing our final expression for the extensive margin as long as the new entrants cannot become the country's most productive firm right away.

$$p_{ijt}(\omega) = \frac{\sigma}{\sigma - 1} \frac{\tau_{ijt} w_{it}}{\varphi_t(\omega)}.$$

A firm's *operating* profits in market j are hence given by:

$$\tilde{\pi}_{ijt}(\omega) = \frac{1}{\sigma} \left(\frac{\sigma}{\sigma - 1} \frac{\tau_{ijt} w_{it}}{\varphi_t(\omega)} \right)^{1-\sigma} P_{jt}^{\sigma-1} \xi_{ijt} E_{jt}.$$

If a firm wants to export to a market j in period t, it has to pay a fixed exporting cost f_{ijt}^{exp}. The exporting fixed cost is higher by a market entry cost factor $f^{entry} \geq 1$ if the firm has not been active in the respective market in the previous period. For tractability, the entry cost factor is assumed to be constant across countries and time. Capturing the export decision by a binary variable $y_{ijt}(\omega)$, i.e. equal to one if the firm decides to serve market j in period t, we can formalize a firm's *realized* profits in market j as follows:

$$\pi_{ijt}(\omega) = y_{ijt}(\omega) \left\{ \tilde{\pi}_{ijt}(\omega) - f_{ijt}^{exp} (f^{entry})^{[1-y_{ij(t-1)}(\omega)]} \right\}.$$

In the absence of entry costs, a firm would simply compare its operating profits to the fixed exporting cost and decide to serve a market if the former are greater than the latter. With market entry costs, a firm might be willing to incur a loss in the current period if expected future profits from that same market outweigh the initial loss. Firms discount future profits at a rate δ per period. To keep things tractable and allow us to derive a theory-consistent estimation expression below, we assume that firms expect their future operating profits from and fixed costs of serving a given market to be equal to today's values, i.e. $\mathbb{E}_t[\tilde{\pi}_{ij(t+s)}] = \tilde{\pi}_{ijt}$ and $\mathbb{E}_t[f_{ij(t+s)}^{exp}] = f_{ijt}^{exp}$ $\forall s \in \mathbb{N}$.[5] The current value of today's and all future operating profits from market j is then given by $\sum_{s=0}^{\infty}(1-\delta)^s \tilde{\pi}_{ijt} = \frac{\tilde{\pi}_{ijt}}{\delta}$. A firm will decide to serve a destination market if these discounted expected profits exceed the sum of today's and discounted future fixed costs of entry and exporting, given by

$$f_{ijt}^{exp} (f^{entry})^{(1-y_{ij(t-1)}(\omega))} + \sum_{s=1}^{\infty} (1-\delta)^s f_{ijt}^{exp} = \frac{f_{ijt}^{exp}}{\delta} \left(1 + \delta(f^{entry} - 1)\right)^{(1-y_{ij(t-1)}(\omega))}.$$

Given this model setup, the question whether a country exports to another country *at all* can be considered by looking at the most productive firm (with φ_t^*) only. Denoting that firm's product by ω^*, we can capture the aggregate extensive margin by the binary variable y_{ijt} as follows:

[5] Note that our final expression for the extensive margin also holds if firms instead expect their operating profits from serving an export market to grow at a constant rate $\bar{g} < \delta$.

$$y_{ijt} = y_{ijt}(\omega^*) = \begin{cases} 1 & \text{if} \quad \frac{\left(\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\frac{\tau_{ijt}w_{it}}{\varphi^*_{it}}\right)^{1-\sigma} P_{jt}^{\sigma-1}\xi_{ijt}E_{jt}\right)}{f_{ijt}^{exp}(1+\delta(f^{entry}-1))^{(1-y_{ij(t-1)})}} \geq 1, \\ 0 & \text{else}. \end{cases} \tag{2}$$

Country i is hence more likely to export to country j in period t if (i) bilateral variable trade costs are lower; (ii) wages in i, and hence production costs, are lower; (iii) the productivity of the most productive firm is higher, again reducing production costs; (iv) competitive pressure, inversely captured by the price index, in j is lower, corresponding to the idea of inward multilateral resistance coined by Anderson and Wincoop (2003) in the intensive margin context; (v) the market in j is larger; (vi) bilateral fixed costs of exporting are smaller; or (vii) i's most productive firm already served market j in the previous period and therefore does not have to pay the market entry cost. Note that (i) to (iv) all act via higher operating profits and depend on the elasticity of substitution between goods. The higher this elasticity, the stronger the reaction of profits to changes in any of these factors. At the same time, a higher elasticity reduces the mark-up firms can charge and hence makes it generally harder to earn enough profits to mitigate the fixed costs of exporting. Further note that the importance of the entry costs depends on the discount factor. Intuitively, if agents are more patient, the one-time entry costs matter less compared to the repeatedly earned profits.

In order to turn equation (2) into the empirical expression that we will bring to the data, we take the natural logarithm and group all exporter-time and importer-time specific components and capture them with corresponding sets of fixed effects. Further, we need to specify the fixed and variable trade costs. In keeping with the existing literature, we model them as a linear combination of different observable bilateral variables, such as geographical distance, whether i and j are both WTO members, or whether i and j share a common currency. In our most general specification, we additionally include country pair fixed effects. Following Baier and Bergstrand (2007), this is common practice in the estimation of the determinants of the intensive margin of trade in order to avoid endogeneity due to unobserved heterogeneity. Further, these bilateral fixed effects may capture (part of) the strong persistence documented above.[6] We then arrive at the following econometric model:

[6]If the trade costs further include any exporter(-time) or importer(-time) specific components, these are captured by the aforementioned corresponding sets of fixed effects.

$$y_{ijt} = \begin{cases} 1 & \text{if} \quad \kappa + \lambda_{it} + \psi_{jt} + \beta_y y_{ij(t-1)} + \mathbf{x}'_{ijt}\boldsymbol{\beta}_x + \mu_{ij} \geq \zeta_{ijt}, \\ 0 & \text{else}, \end{cases} \tag{3}$$

where $\kappa = -\sigma \log(\sigma) - (1-\sigma)\log(\sigma - 1) - \log(1 + \delta(f^{entry} - 1))$, $\lambda_{it} = (1-\sigma)(\log(w_{it}) - \log(\varphi^*_{it}))$, $\psi_{jt} = (\sigma - 1)\log(P_{jt}) + \log(E_{jt})$, $\beta_y = \log(1 + \delta(f^{entry} - 1))$, $\mathbf{x}'_{ijt}\boldsymbol{\beta}_x + \mu_{ij} = (1-\sigma)\log(\tau_{ijt}) - \log(f^{exp}_{ijt})$, and $\zeta_{ijt} = -\log(\xi_{ijt}) \sim \mathcal{N}(0, 1)$. The error term distribution implies that a probit estimator is the appropriate choice to estimate our model. Alternatively, we could deviate from Eaton, Kortum, and Kramarz (2011) and assume a log-logistic distribution for the idiosyncratic demand shocks, which would lead to a logit specification.

Our theoretical framework implies a flexible empirical specification that can reconcile the extensive margin estimation with the stylized fact presented in Section 1. Note that we chose to make a number of simplifying assumptions in order to achieve the clear theory-consistent interpretation of specification (3). An alternative interpretation of equation (3) as a reduced-from representation of a more elaborate and realistic model (similar e.g. to how Roberts and Tybout, 1997, motivate their empirical consideration) is equally justifiable. At the same time, while our model is written along the lines of Helpman, Melitz, and Rubinstein (2008), which remains the benchmark for the empirical assessment of the (aggregate) extensive margin of trade, it is not decisive for our empirical specification that zero trade flows result from a truncated productivity distribution instead of a discrete number of firms (as in Eaton, Kortum, and Sotelo, 2013) or from fixed exporting costs in a Krugman (1980)-type homogeneous firms setting (as in Egger and Larch, 2011; Egger, Larch, et al., 2011).

3 Binary Response Estimators with High-Dimensional Fixed Effects

Having set up the empirical framework, we now turn to the estimation procedure. As equation (3) demands two- or three-way fixed effects to capture unobservable characteristics, we describe how to implement suitable binary choice estimators. In a first step, we review a recent procedure for estimating probit and logit models with high-dimensional fixed effects. In a second step, we characterize appropriate bias correction techniques to address the induced incidental parameter problem.

Table 1: Expressions and Derivatives for Logit and Probit Models

	Logit	Probit
F_{ijt}	$(1+\exp(-\eta_{ijt}))^{-1}$	$\Phi(\eta_{ijt})$
$\partial_{\eta} F_{ijt}$	$F_{ijt}(1-F_{ijt})$	$\phi(\eta_{ijt})$
$\partial_{\eta^2} F_{ijt}$	$\partial_{\eta} F_{ijt}(1-2F_{ijt})$	$-\eta_{ijt}\phi(\eta_{ijt})$
ν_{ijt}	$(y_{ijt}-F_{ijt})/\partial_{\eta} F_{ijt}$	$(y_{ijt}-F_{ijt})/\partial_{\eta} F_{ijt}$
H_{ijt}	1	$\partial_{\eta} F_{ijt}/(F_{ijt}(1-F_{ijt}))$
ω_{ijt}	$\partial_{\eta} F_{ijt}$	$H_{ijt}\partial_{\eta} F_{ijt}$
$\partial_{\eta} \ell_{ijt}$	$y_{ijt}-F_{ijt}$	$H_{ijt}(y_{ijt}-F_{ijt})$

Note: $\eta_{ijt} = \mathbf{x}'_{ijt}\boldsymbol{\beta}+\lambda_{it}+\psi_{jt}$ or $\eta_{ijt} = \mathbf{x}'_{ijt}\boldsymbol{\beta}+\lambda_{it}+\psi_{jt}+\mu_{ij}$ is the linear predictor.

3.1 Feasible Estimation

In this subsection, we sketch how to estimate structural parameters, average partial effects (APEs), and the corresponding standard errors in a binary response setting in the presence of high-dimensional fixed effects. Let $\mathbf{Z} = [\mathbf{D}, \mathbf{X}]$, where $\mathbf{D}$ is the dummy matrix corresponding to the fixed effects and $\mathbf{X}$ is a matrix of further regressors. Note that $\mathbf{X}$ may also include predetermined variables. Further, let $\boldsymbol{\alpha}$ denote the vector of fixed effects, $\boldsymbol{\beta}$ the vector of structural parameters, and $\boldsymbol{\theta} = [\boldsymbol{\alpha}', \boldsymbol{\beta}']'$. The log-likelihood contribution of the ijt-th observation is

$$\ell_{ijt}(\boldsymbol{\beta}, \boldsymbol{\alpha}_{ijt}) = y_{ijt}\log(F_{ijt}) + (1-y_{ijt})\log(1-F_{ijt}),$$

where $\boldsymbol{\alpha}_{ijt} = [\lambda_{it}, \psi_{jt}]'$ in the case of two-way fixed effects and $\boldsymbol{\alpha}_{ijt} = [\lambda_{it}, \psi_{jt}, \mu_{ij}]'$ in the case of three-way fixed effects.[7] Further, F_{ijt} is either the logistic or the standard normal cumulative distribution function. See Table 1 for the relevant expressions and derivatives.

The standard approach to estimate binary choice models is to maximize the following log-likelihood function:

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \ell_{ijt}(\boldsymbol{\beta}, \boldsymbol{\alpha}_{ijt})$$

[7]Note that we use for brevity notation for balanced data.

using Newton's method. The update in the $(r-1)$-th iteration is

$$\boldsymbol{\theta}^r - \boldsymbol{\theta}^{r-1} = (\mathbf{Z}'\widehat{\boldsymbol{\Omega}}\mathbf{Z})^{-1}\mathbf{Z}'\widehat{\boldsymbol{\Omega}}\hat{\boldsymbol{\nu}} \; , \tag{4}$$

where $\mathbf{Z}'\widehat{\boldsymbol{\Omega}}\mathbf{Z}$ and $\mathbf{Z}'\widehat{\boldsymbol{\Omega}}\hat{\boldsymbol{\nu}}$ denote the Hessian and gradient of the log-likelihood, respectively, and $\widehat{\boldsymbol{\Omega}}$ is a diagonal weighting matrix with diag$(\widehat{\boldsymbol{\Omega}}) = \widehat{\boldsymbol{\omega}}$.

The brute-force computation of equation (4) quickly becomes computationally demanding, if not impossible.[8] Thus Stammann (2018) suggests a straightforward strategy called pseudo-demeaning, which mimics the well-known within transformation for linear regression models. The approach allows us to update the structural parameters without having to explicitly update the incidental parameters, which leads to the following concentrated version of equation (4)

$$\boldsymbol{\beta}^r - \boldsymbol{\beta}^{r-1} = \left((\widehat{\mathbb{M}}\mathbf{X})'\widehat{\boldsymbol{\Omega}}(\widehat{\mathbb{M}}\mathbf{X})\right)^{-1}(\widehat{\mathbb{M}}\mathbf{X})'\widehat{\boldsymbol{\Omega}}(\widehat{\mathbb{M}}\hat{\boldsymbol{\nu}}) \; , \tag{5}$$

where $(\widehat{\mathbb{M}}\mathbf{X})'\widehat{\boldsymbol{\Omega}}(\widehat{\mathbb{M}}\hat{\boldsymbol{\nu}})$ is the concentrated gradient, $(\widehat{\mathbb{M}}\mathbf{X})'\widehat{\boldsymbol{\Omega}}(\widehat{\mathbb{M}}\mathbf{X})$ is the concentrated Hessian, and $\widehat{\mathbb{M}} = \mathbf{I}_{IJT} - \widehat{\mathbb{P}} = \mathbf{I}_{IJT} - \mathbf{D}(\mathbf{D}'\widehat{\boldsymbol{\Omega}}\mathbf{D})^{-1}\mathbf{D}'\widehat{\boldsymbol{\Omega}}$ is known as the residual projection that partials out the fixed effects. After convergence of the optimization routine, the standard errors associated with the structural parameters can be computed from the inverse of the concentrated Hessian.

Since the computation of $\widehat{\mathbb{M}}$ itself is problematic even in moderately large data sets, Stammann (2018) proposes to calculate the centered variables $\widehat{\mathbb{M}}\hat{\boldsymbol{\nu}}$ and $\widehat{\mathbb{M}}\mathbf{X}$ using the method of alternating projections (MAP), which only requires repeatedly performing group-specific one-way weighted within transformations. This approach is feasible, since these within transformations translate into simple scalar transformations (see Stammann, Heiß, and McFadden, 2016).[9] Note that all expressions containing $\widehat{\mathbb{M}}$ or $\widehat{\mathbb{P}}$ can be calculated efficiently based on the MAP.

Next, we address the estimation of APEs. An estimator for the APEs is

[8]In a balanced data set ($I = J = N$) with two-way fixed effects the routine requires to estimate $\approx 2NT$ fixed effects associated with a $2NT \times 2NT$ Hessian. In the case of three-way fixed effects, the number of parameters to be estimated is even $\approx N(N-1) \times 2NT$. In a trade panel data set with 200 countries and 50 years, the number of fixed effects in the latter case amounts to 59800 parameters.

[9]For further details, we refer the reader to Appendix B.1, where we sketch the MAP for our application of two-way and three-way models, and provide the entire optimization routine corresponding to equation (5).

$$\hat{\delta}_k = \frac{1}{IJT} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\Delta}_{ijt}^{k} \,,$$

where the partial effect of the k-th regressor $\widehat{\Delta}_{ijt}^{k}$ is either $\widehat{\Delta}_{ijt}^{k} = \partial \widehat{F}_{ijt} / \partial x_{ijtk}$ in the case of a continuous regressor or $\hat{\Delta}_{ijt}^{k} = \widehat{F}_{ijt}|_{x_{ijtk=1}} - \widehat{F}_{ijt}|_{x_{ijtk=0}}$ in the case of a binary regressors. Another question that arises in the context of APEs is how to calculate appropriate standard errors, even in the case of high-dimensional fixed effects. A possible candidate is the delta method, but in its standard form it requires the entire covariance matrix, which we do not obtain using the pseudo-demeaning approach. However, as outlined in Fernández-Val and Weidner (2016) and Czarnowske and Stammann (2019) in the context of individual and time fixed effects, it is possible to use a concentrated version of the delta method. In the following we present the feasible covariance estimators for our two-way and three-way error structure.[10] An appropriate covariance estimator for the APEs of the two-way fixed effects model is

$$\widehat{\mathbf{V}}^{\delta} = \frac{1}{I^2 J^2 T^2} \left(\underbrace{\left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\overline{\boldsymbol{\Delta}}}_{ijt} \right) \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\overline{\boldsymbol{\Delta}}}_{ijt} \right)^{\prime}}_{v_1} + \underbrace{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\boldsymbol{\Gamma}}_{ijt} \widehat{\boldsymbol{\Gamma}}_{ijt}^{\prime}}_{v_2} \right) , \tag{6}$$

and of the three-way error component model

$$\widehat{\mathbf{V}}^{\delta} = \frac{1}{I^2 J^2 T^2} \left(\underbrace{\left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\overline{\boldsymbol{\Delta}}}_{ijt} \right) \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\overline{\boldsymbol{\Delta}}}_{ijt} \right)^{\prime}}_{v_1} + \underbrace{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\boldsymbol{\Gamma}}_{ijt} \widehat{\boldsymbol{\Gamma}}_{ijt}^{\prime}}_{v_2} \right.$$

$$\left. + \underbrace{2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{s>t}^{T} \widehat{\overline{\boldsymbol{\Delta}}}_{ijt} \widehat{\boldsymbol{\Gamma}}_{ijs}^{\prime}}_{v_3} \right) , \tag{7}$$

where in both cases $\widehat{\overline{\boldsymbol{\Delta}}}_{ijt} = \widehat{\boldsymbol{\Delta}}_{ijt} - \hat{\boldsymbol{\delta}}$, $\widehat{\boldsymbol{\Delta}}_{ijt} = [\widehat{\Delta}_{ijt}^{1}, \ldots, \widehat{\Delta}_{ijt}^{m}]^{\prime}$, $\hat{\boldsymbol{\delta}} = [\hat{\delta}_1, \ldots, \hat{\delta}_m]^{\prime}$, and

$$\widehat{\boldsymbol{\Gamma}}_{ijt} = \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \partial_{\beta} \widehat{\boldsymbol{\Delta}}_{ijt} - \left(\widehat{\mathbb{P}} \mathbf{X} \right)_{ijt} \partial_{\eta} \widehat{\boldsymbol{\Delta}}_{ijt} \right)^{\prime} \widehat{\mathbf{A}}^{-1} \left(\widehat{\mathbb{M}} \mathbf{X} \right)_{ijt} \hat{\omega}_{ijt} \hat{\boldsymbol{\nu}}_{ijt} - \left(\widehat{\mathbb{P}} \widehat{\boldsymbol{\Psi}} \right)_{ijt} \partial_{\eta} \hat{\ell}_{ijt} \,,$$

with $\widehat{\mathbf{A}} = (\widehat{\mathbb{M}} \mathbf{X})^{\prime} \widehat{\boldsymbol{\Omega}} (\widehat{\mathbb{M}} \mathbf{X})$, $\widehat{\boldsymbol{\Psi}}_{ijt} = \partial_{\eta} \widehat{\boldsymbol{\Delta}}_{ijt} / \hat{\omega}_{ijt}$, and $\partial_{\eta} \hat{\ell}_{ijt}$ defined in Table 1. To clarify

[10] The corresponding asymptotic distribution of the estimators is provided in Appendix B.3.

notation, $\partial_\iota g(\cdot)$ denotes the first order partial derivative of an arbitrary function $g(\cdot)$ with respect to some parameter ι. Note, that the term v_2 refers to the concentrated delta method. The terms v_1 and v_3 are in the spirit of Fernández-Val and Weidner (2016) to improve the finite sample properties. These are, on the one hand, the variation induced by estimating sample instead of population means (v_1). On the other hand, if we are concerned about the strict exogeneity assumption (as we are in the case of dynamic three-way error structure models), the covariance between the estimation of sample means and parameters is another factor that should be incorporated (v_3). These computationally efficient covariance estimators can be readily applied not only to uncorrected APE estimators, but also to the bias-corrected APE estimators, which we will introduce below.

3.2 Incidental Parameter Bias Correction

As many nonlinear estimators, standard fixed effects versions of the logit and probit models suffer from the well-known incidental parameter problem first identified by Neyman and Scott (1948). The problem stems from the necessity to estimate many nuisance parameters, which contaminate the estimator of the structural parameters and average partial effects. It can be further amplified by the inclusion of a lagged dependent variable.[11] Fernández-Val and Weidner (2018) derive the order of the bias induced by incidental parameters to be given by $bias \sim p/n$, where p and n are the numbers of parameters and observations, respectively. The literature suggests different types of bias corrections to reduce this incidental parameter bias. Jackknife corrections, like the leave-one-out jackknife proposed by Hahn and Newey (2004), or the split-panel jackknife (SPJ) introduced by Dhaene and Jochmans (2015), are the simplest approaches to obtain a bias correction, at the expense of being computationally costly. In contrast to analytical corrections, their application only requires knowledge of the order of the bias to form appropriate subpanels that are used to reestimate the model and to form an estimator of the bias terms. For analytical bias correction (ABC), it is necessary to derive the asymptotic distribution of the maximum likelihood estimator (MLE), in order to obtain an explicit expression of the asymptotic bias. This is then used to form a suitable estimator for the bias terms. Fernández-Val and Weidner (2016) propose analytical and split-panel jackknife bias corrections for structural parameters and APEs in the context of nonlinear models with individual and time fixed effects. In the following two subsections, we adapt and extend the bias corrections of Fernández-Val and Weidner

[11] Note that this induces an incidental parameter problem, even in the linear three-way fixed effects setting (see Nickell, 1981) — and hence in our case also affects a linear probability model specification.

(2016) to our two-way and three-way error component.[12]

3.2.1 Two-way fixed effects

The two-way fixed effects case with exporter-time and importer-time fixed effects is closely related to the two-way fixed effects models with a classical panel structure and individual and time fixed effects or with a pseudo-panel ij-structure and exporter and importer fixed effects as discussed by Fernández-Val and Weidner (2016) and Cruz-Gonzalez, Fernández-Val, and Weidner (2017), respectively. It is straightforward to see that in our case the overall bias consists of two components that are due to the inclusion of importer-time and exporter-time fixed effects, respectively, and takes the form $B_1/I + B_2/J$.[13]

The form of the bias suggests to separately split the panel by I and J, leading to the following split-panel corrected estimator for the structural parameters:

$$\widehat{\boldsymbol{\beta}}^{sp} = 3\widehat{\boldsymbol{\beta}}_{I,J,T} - \widehat{\boldsymbol{\beta}}_{I/2,J,T} - \widehat{\boldsymbol{\beta}}_{I,J/2,T}, \quad \text{with} \tag{8}$$

$$\widehat{\boldsymbol{\beta}}_{I/2,J,T} = \frac{1}{2}\left[\widehat{\boldsymbol{\beta}}_{\{i:i\leq\lceil I/2\rceil\},J,T} + \widehat{\boldsymbol{\beta}}_{\{i:i\geq\lfloor I/2+1\rfloor\},J,T}\right],$$

$$\widehat{\boldsymbol{\beta}}_{I,J/2,T} = \frac{1}{2}\left[\widehat{\boldsymbol{\beta}}_{I,\{j:j\leq\lceil J/2\rceil,T\}} + \widehat{\boldsymbol{\beta}}_{I,\{j:j\geq\lfloor J/2+1\rfloor,T\}}\right],$$

where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions. To clarify the notation, the subscript $\{i : i \leq \lceil I/2\rceil\}, J, T$ denotes that the estimator is based on a subsample, which contains all importers and time periods, but only the first half of all exporters.

In order to form the appropriate analytical bias correction, we need to specify the asymptotic distribution of the MLE, which we show in Appendix B.3. The analytical bias-corrected estimator $\tilde{\boldsymbol{\beta}}^a$ is formed from estimators of the leading bias terms that are subtracted from the MLE of the full sample $\widehat{\boldsymbol{\beta}}_{I,J,T}$. More precisely:

[12]We do not elaborate on the leave-one-out jackknife bias correction because the large number of fixed effects in our panel structure makes it unnecessarily computationally demanding.

[13]See Appendix B.3. We also report the appropriate Neyman and Scott (1948) variance example in Appendix B.2 as an illustration.

$$\tilde{\boldsymbol{\beta}}^a = \hat{\boldsymbol{\beta}}_{I,J,T} - \frac{\widehat{\mathbf{B}}_1^\beta}{I} - \frac{\widehat{\mathbf{B}}_2^\beta}{J}, \quad \text{with} \quad \widehat{\mathbf{B}}_1^\beta = \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{B}}_1, \widehat{\mathbf{B}}_2^\beta = \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{B}}_2, \quad \text{and}$$

$$\widehat{\mathbf{B}}_1 = -\frac{1}{2JT}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{\sum_{i=1}^{I}\widehat{H}_{ijt}\partial_{\eta^2}\widehat{F}_{ijt}\left(\widehat{\mathbb{M}}\mathbf{X}\right)_{ijt}}{\sum_{i=1}^{I}\hat{\omega}_{ijt}},$$

$$\widehat{\mathbf{B}}_2 = -\frac{1}{2IT}\sum_{i=1}^{I}\sum_{t=1}^{T}\frac{\sum_{j=1}^{J}\widehat{H}_{ijt}\partial_{\eta^2}\widehat{F}_{ijt}\left(\widehat{\mathbb{M}}\mathbf{X}\right)_{ijt}}{\sum_{j=1}^{J}\hat{\omega}_{ijt}},$$

$$\widehat{\mathbf{W}} = \frac{1}{IJT}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\hat{\omega}_{ijt}\left(\widehat{\mathbb{M}}\mathbf{X}\right)_{ijt}\left(\widehat{\mathbb{M}}\mathbf{X}\right)_{ijt}',$$

where $\partial_{\iota^2}g(\cdot)$ denotes the second order partial derivative of an arbitrary function $g(\cdot)$ with respect to some parameter ι. The explicit expressions of H_{ijt} and $\partial_{\eta^2}F_{ijt}$ are reported in Table 1.

The split-panel jackknife estimator works similarly with APEs as with structural parameters. We simply replace in formula (8) the estimators for the structural parameters with estimators for the APEs. The following analytically bias-corrected estimator for the APEs is formed based on the asymptotic distribution presented in Appendix B.3:

$$\tilde{\boldsymbol{\delta}}^a = \hat{\boldsymbol{\delta}} - \frac{\widehat{\mathbf{B}}_1^\delta}{I} - \frac{\widehat{\mathbf{B}}_2^\delta}{J}, \quad \text{with}$$

$$\widehat{\mathbf{B}}_1^\delta = \frac{1}{2JT}\sum_{j=1}^{J}\sum_{t=1}^{T}\frac{\sum_{i=1}^{I}-\widehat{H}_{ijt}\partial_{\eta^2}\widehat{F}_{ijt}\left(\widehat{\mathbb{P}}\widehat{\boldsymbol{\Psi}}\right)_{ijt}+\partial_{\eta^2}\widehat{\boldsymbol{\Delta}}_{ijt}}{\sum_{i=1}^{I}\hat{\omega}_{ijt}},$$

$$\widehat{\mathbf{B}}_2^\delta = \frac{1}{2IT}\sum_{i=1}^{I}\sum_{t=1}^{T}\frac{\sum_{j=1}^{J}-\widehat{H}_{ijt}\partial_{\eta^2}\widehat{F}_{ijt}\left(\widehat{\mathbb{P}}\widehat{\boldsymbol{\Psi}}\right)_{ijt}+\partial_{\eta^2}\widehat{\boldsymbol{\Delta}}_{ijt}}{\sum_{j=1}^{J}\hat{\omega}_{ijt}}.$$

Note that all quantities are evaluated at bias-corrected structural parameters and the corresponding estimates of the fixed effects, where the latter can be obtained by reestimating the model using an offset algorithm as in Czarnowske and Stammann (2019). The covariance can be estimated according to equation (6).

3.2.2 Three-way fixed effects

Having adapted the two-way fixed effects bias correction of Fernández-Val and Weidner (2016) to the ijt-panel setting, we now move on to the more difficult case of extending the consideration to three-way fixed effects. Fernández-Val and Weidner (2018) conjecture, based on their previously discussed formula, $bias \sim p/n$, that the bias is of order $(IT + JT + IJ)/(IJT)$ and of the form $B_1/I + B_2/J + B_3/T$. Intuitively, the inclusion of dyadic fixed effects induces another bias of order $1/T$ because there are only T informative observations per additionally included parameter. We support their conjecture by providing the appropriate Neyman and Scott (1948) variance example in Appendix B.2 and propose novel analytical and jackknife bias corrections for three-way fixed effects models.

For the split-panel jackknife bias correction, this bias structure implies that we add an additional splitting dimension, leading to the following estimator for the structural parameters:

$$\widehat{\boldsymbol{\beta}}^{sp} = 4\widehat{\boldsymbol{\beta}}_{I,J,T} - \widehat{\boldsymbol{\beta}}_{I/2,J,T} - \widehat{\boldsymbol{\beta}}_{I,J/2,T} - \widehat{\boldsymbol{\beta}}_{I,J,T/2}, \quad \text{with} \tag{9}$$

$$\widehat{\boldsymbol{\beta}}_{I/2,J,T} = \frac{1}{2}\Big[\widehat{\boldsymbol{\beta}}_{\{i:i\leq\lfloor I/2\rfloor,J,T\}} + \widehat{\boldsymbol{\beta}}_{\{i:i\geq\lceil I/2+1\rceil,J,T\}}\Big],$$

$$\widehat{\boldsymbol{\beta}}_{I,J/2,T} = \frac{1}{2}\Big[\widehat{\boldsymbol{\beta}}_{\{I,j:j\leq\lfloor J/2\rfloor,T\}} + \widehat{\boldsymbol{\beta}}_{\{I,j:j\geq\lceil J/2+1\rceil,T\}}\Big],$$

$$\widehat{\boldsymbol{\beta}}_{I,J,T/2} = \frac{1}{2}\Big[\widehat{\boldsymbol{\beta}}_{\{I,J,t:t\leq\lfloor T/2\rfloor\}} + \widehat{\boldsymbol{\beta}}_{\{I,J,t:t\geq\lceil T/2+1\rceil\}}\Big].$$

Combining insights from the classical panel structure in Fernández-Val and Weidner (2016), the pseudo-panel setting in Cruz-Gonzalez, Fernández-Val, and Weidner (2017), and the three-way fixed effects conjecture by Fernández-Val and Weidner (2018), we formulate a conjecture for the asymptotic MLE distribution in the three-way setting (which we present in Appendix B.3) and propose to extend the analytical two-way bias correction by a third part $\widehat{\mathrm{B}}_3$, such that

$$\tilde{\boldsymbol{\beta}}^a = \hat{\boldsymbol{\beta}}_{I,J,T} - \frac{\widehat{\mathbf{B}}_1^\beta}{I} - \frac{\widehat{\mathbf{B}}_2^\beta}{J} - \frac{\widehat{\mathbf{B}}_3^\beta}{T}, \quad \text{with} \quad \widehat{\mathbf{B}}_3^\beta = \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{B}}_3$$

$$\widehat{\mathbf{B}}_3 = -\frac{1}{2IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{t=1}^{T}\hat{\omega}_{ijt}\right)^{-1}\left(\sum_{t=1}^{T}\widehat{H}_{ijt}\partial_{\eta^2}\widehat{F}_{ijt}\left(\widehat{\mathbb{M}}\mathbf{X}\right)_{ijt}\right.$$
$$\left. +2\sum_{l=1}^{L}(T/(T-L))\sum_{t=l+1}^{T}\partial_\eta\hat{\ell}_{ijt-l}\hat{\omega}_{ijt}\left(\widehat{\mathbb{M}}\mathbf{X}\right)_{ijt}\right).$$

L is a bandwidth parameter and is used for the estimation of spectral densities (Hahn and Kuersteiner, 2007). In a model where all regressors are exogenous, L is set to zero, such that the second part in the numerator of $\widehat{\mathbf{B}}_3$ vanishes and all three estimators of the bias terms are symmetric. Otherwise, for instance in the dynamic model, Fernández-Val and Weidner (2016) suggest conducting a sensitivity analysis with $L \in \{1, 2, 3, 4\}$.

Again, for the APEs the split-panel jackknife estimator is formed by replacing the estimators for the structural parameters with estimators for the APEs in formula (9). The analytically bias-corrected estimator, based on our conjecture for the asymptotic distribution provided in Appendix B.3, is given by

$$\tilde{\boldsymbol{\delta}}^a = \hat{\boldsymbol{\delta}} - \frac{\widehat{\mathbf{B}}_1^\delta}{I} - \frac{\widehat{\mathbf{B}}_2^\delta}{J} - \frac{\widehat{\mathbf{B}}_3^\delta}{T}, \quad \text{with}$$

$$\widehat{\mathbf{B}}_3^\delta = \frac{1}{2IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{t=1}^{T}\hat{\omega}_{ijt}\right)^{-1}\left(\sum_{t=1}^{T}-\widehat{H}_{ijt}\partial_{\eta^2}\widehat{F}_{ijt}\left(\widehat{\mathbb{P}}\widehat{\boldsymbol{\Psi}}\right)_{ijt} + \partial_{\eta^2}\widehat{\boldsymbol{\Delta}}_{ijt}\right.$$
$$\left. +2\sum_{l=1}^{L}(T/(T-l))\sum_{t=l+1}^{T}\partial_\eta\hat{\ell}_{ijt-l}\hat{\omega}_{ijt}\left(\widehat{\mathbb{M}}\widehat{\boldsymbol{\Psi}}\right)_{ijt}\right).$$

The last part in the numerator of $\widehat{\mathbf{B}}_3^\delta$ is again dropped if all regressors are assumed to be strictly exogenous. As previously mentioned, standard errors can still be obtained from equation (7).

4 Monte Carlo Simulations

In this section, we conduct extensive simulation experiments to investigate the properties of different estimators for both the structural parameters and the APEs. The estimators we study are MLE, ABC, SPJ and a (bias-corrected) ordinary least squares fixed effects estimator (LPM).[14] Our main focus are the biases and inference accuracies. To this end, we compute the relative bias and standard deviation (SD) in percent, the ratio between standard error and standard deviation (SE/SD), the relative root mean square error (RMSE) in percent, and the coverage probabilities (CPs) at a nominal level of 95 percent.

For the simulation experiments we adapt the design for a dynamic probit model of Fernández-Val and Weidner (2016) to our ijt-panel structure for the two cases with two- and three-way fixed effects.[15]

4.1 Two-way fixed effects

The simulations in this section correspond to a theory-consistent estimation of the extensive margin outlined in section 2, taking into account unobserved time-varying exporter- and importer-specific terms as well as dynamics, but not allowing for bilateral unobserved heterogeneity. Specifically, we generate data according to

$$y_{ijt} = \mathbf{1}[\beta_y y_{ijt-1} + \beta_x x_{ijt} + \lambda_{it} + \psi_{jt} \geq \epsilon_{ijt}] \ ,$$

$$y_{ij0} = \mathbf{1}[\beta_x x_{ij0} + \lambda_{i0} + \psi_{j0} \geq \epsilon_{ij0}] \ ,$$

where $i = 1, \ldots, N$, $j = 1, \ldots, N$, $t = 1, \ldots, T$, $\lambda_{it} \sim \text{iid.}\, \mathcal{N}(0, 1/16)$, $\psi_{jt} \sim \text{iid.}\, \mathcal{N}(0, 1/16)$, and $\epsilon_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 1)$.[16] Further, $x_{ijt} = 0.5 x_{ijt-1} + \lambda_{it} + \psi_{jt} + \nu_{ijt}$, where $\nu_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 0.5)$, $x_{ij0} \sim \text{iid.}\, \mathcal{N}(0, 1)$. To get an impression of how the different statistics evolve with changing panel dimensions, we consider all possible combinations of $N \in \{50, 100, 150\}$ and $T \in \{10, 20, 30, 40, 50\}$. For each of these combinations we generate $1,000$ samples.

Tables 8 – 13 in Appendix C.1 report the extensive simulation results for the exogenous and predetermined regressors, respectively. The left panels contain the results of the structural parameters and the right panels the results of the APEs. In the following, we focus on the biases and coverage probabilities for $N \in \{50, 150\}$, which we visualize in Figures 2 and 3 for better comprehensibility.

First of all, we start analyzing the properties of the different estimators for the structural

[14] Details on LPM and our suggested bias correction in this context are given in Appendix B.4.

[15] Further simulation experiments including static panel models are presented in Appendix D.

[16] Since $\{\lambda_{it}\}_{IT}$ and $\{\psi_{jt}\}_{JT}$ are independent sequences, and λ_{it} and ψ_{jt} are independent for all it, jt, we follow Fernández-Val and Weidner (2016) and incorporate this information in the covariance estimator for the APEs. The explicit expression is provided in the Appendix B.3.

Figure 2: Dynamic: Two-way Fixed Effects – Predetermined Regressor

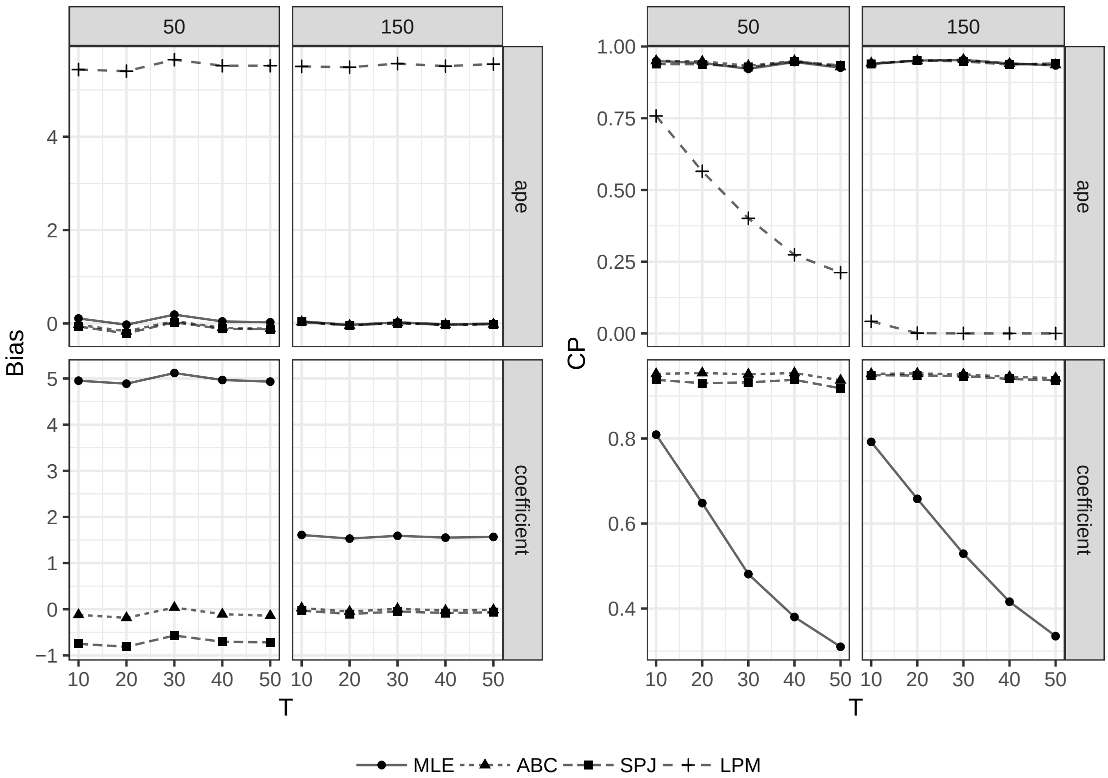
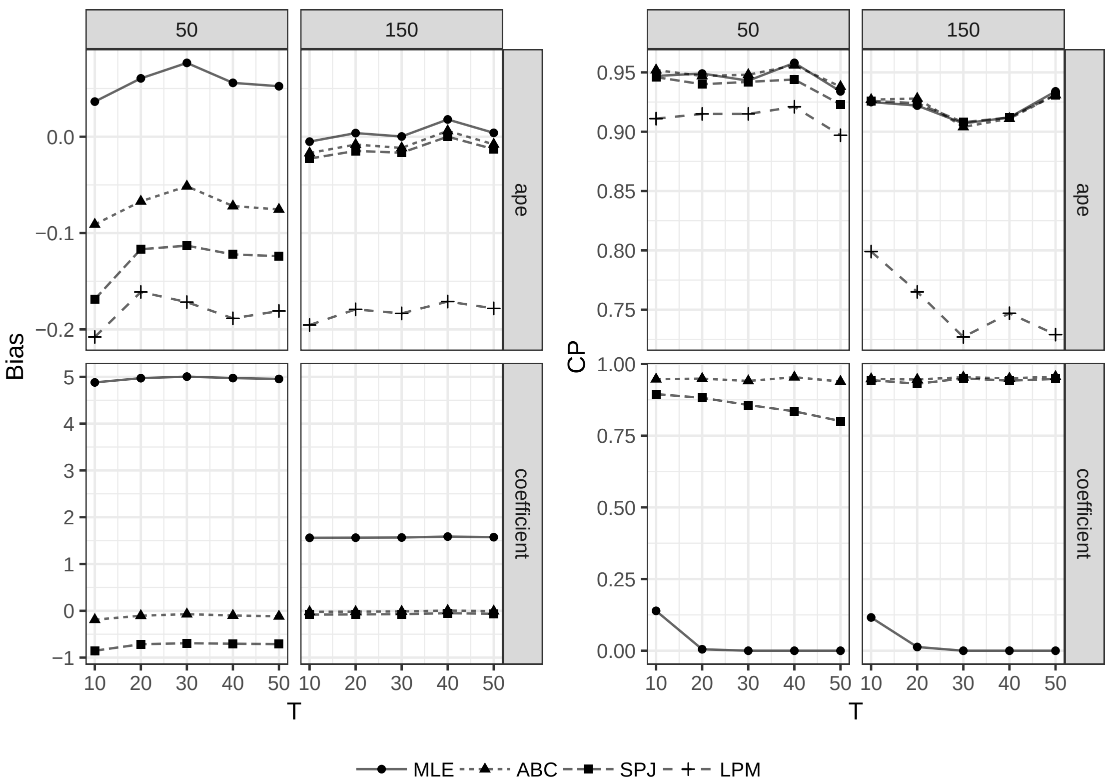


Figure 3: Dynamic: Two-way Fixed Effects – Exogenous Regressor



parameters. MLE exhibits persistent biases that do not fade with increasing T but with increasing N. This result is as expected, since MLE is fixed T consistent as shown in Appendix B.3. Further, its CPs are too low and decreasing in T. The bias-corrected estimators clearly perform better than MLE. First, they reduce the bias considerably. ABC shows basically no bias for any considered sample size. SPJ performs slightly worse. Second, the bias corrections also dramatically improve the coverage probabilities. Whereas the CPs of ABC are close to the nominal value in all cases, the CPs of SPJ are somewhat too low for the exogenous regressor in the case of $N = 50$.

Next, we turn to the estimators of the APEs, where we now also consider LPM. It turns out that MLE, as well as the two bias-corrected estimators, are essentially unbiased. This is particularly noteworthy for MLE, since it exhibits a non-negligible bias for the structural parameters. Remarkably, LPM displays persistent biases that — differently to the nonlinear estimators — do not vanish with larger N. The bias is very small for the exogenous regressor but for the predetermined regressor it ranges between 5 and 6 percent.[17] These persistent biases also explain that LPM delivers too small CPs that decrease in T. Contrary, the CPs of the three nonlinear estimators are close to the nominal value in most cases.

All in all, our two-way fixed effects simulation results demonstrate that the bias-corrected estimators work extremely well in this context — for both structural parameters and APEs and both bias and coverage probabilities. Between the two, the analytical correction slightly outperforms the split-panel jackknife correction. If the interest lies only in APEs, the MLE estimator works well, too, but for the structural parameters it shows bias and essentially useless coverage probabilities. LPM performs clearly worse than the probit estimators and should — given the availability of the nonlinear alternatives — only be used with great caution.

4.2 Three-way fixed effects

The simulations in this section correspond to our preferred empirical specification for the extensive margin of international trade, in which we not only take into account the theoretically motivated it- and jt-fixed effects, but additionally allow for bilateral unobserved heterogeneity. In this three-way error structure environment, we generate data according to

[17] We found that the predicted probabilities of LPM exceed the boundaries of the unit interval considerably. This, in turn, affects the APEs for binary regressors, since they are based on differences of predicted probabilities.

$$y_{ijt} = \mathbf{1}[\beta_y y_{ijt-1} + \beta_x x_{ijt} + \lambda_{it} + \psi_{jt} + \mu_{ij} \geq \epsilon_{ijt}] \ ,$$

$$y_{ij0} = \mathbf{1}[\beta_x x_{ij0} + \lambda_{i0} + \psi_{j0} + \mu_{ij} \geq \epsilon_{ij0}] \ ,$$

where $i = 1, \ldots, N$, $j = 1, \ldots, N$, $t = 1, \ldots, T$, $\beta_y = 0.5$, $\beta_x = 1$, $\lambda_{it} \sim \text{iid.}\, \mathcal{N}(0, 1/24)$, $\psi_{jt} \sim \text{iid.}\, \mathcal{N}(0, 1/24)$, $\mu_{ij} \sim \text{iid.}\, \mathcal{N}(0, 1/24)$, and $\epsilon_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 1)$.[18] The exogenous regressor is modeled as an AR-1 process, $x_{ijt} = 0.5 x_{ijt-1} + \lambda_{it} + \psi_{jt} + \mu_{ij} + \nu_{ijt}$, where $\nu_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 0.5)$ and $x_{ij0} \sim \text{iid.}\, \mathcal{N}(0, 1)$. Again, we consider different sample sizes, specifically $N \in \{50, 100, 150\}$ and $T \in \{10, 20, 30, 40, 50\}$ and generate $1,000$ data sets for each.

Tables 17 – 16 in Appendix C.2 summarize the extensive simulation results for both regressors. For ABC and LPM we report two different choices of the bandwidth parameter, $L = 1$ and $L = 2$. Here, we again focus on the biases and coverage probabilities for $N \in \{50, 150\}$ which are shown in Figures 4 and 5.

We start by considering the different estimators for the structural parameters. For both kinds of regressors, MLE exhibits a severe bias that decreases with increasing T. However, even with $T = 50$, the estimator shows a distortion of 11 percent in the case of the predetermined regressor and 5 percent in the case of the exogenous regressor. We also find that the inference is not valid, since the CPs are zero or close to zero. The bias corrections bring a substantial improvement. First, they reduce the bias considerably. For example, the MLE estimator of the predetermined regressor shows a distortion of 63 percent for $T = 10$ and $N = 150$. ABC reduces the bias to 8 percent and SPJ to 20 percent. In the case of the exogenous regressor, MLE exhibits a bias of 23 percent, whereas ABC has a bias of 1 percent and SPJ of 7 percent. Irrespective of the type of the regressor, both bias-corrected estimators also converge quickly to the true parameter value with growing T. Second, the bias corrections improve the CPs. For the exogenous regressor the CPs of ABC are close to the desired level of 95 percent for all T, whereas SPJ remains far away from 95 percent even at $T = 50$. In the case of the predetermined regressor, the CPs of both corrections approach the nominal level when T rises. This happens faster for ABC.

We again proceed with the APEs, where we also consider LPM as an alternative estimator.

[18] We again follow Fernández-Val and Weidner (2016) and incorporate the information that $\{\lambda_{it}\}_{IT}$, $\{\psi_{jt}\}_{JT}$, and $\{\mu_{ij}\}_{IJ}$ are independent sequences, and λ_{it}, ψ_{jt}, and μ_{ij} are independent for all it, jt, ij in the covariance estimator for the APEs. The explicit expression is provided in Appendix B.3.

Figure 4: Dynamic: Three-way Fixed Effects – Predetermined Regressor

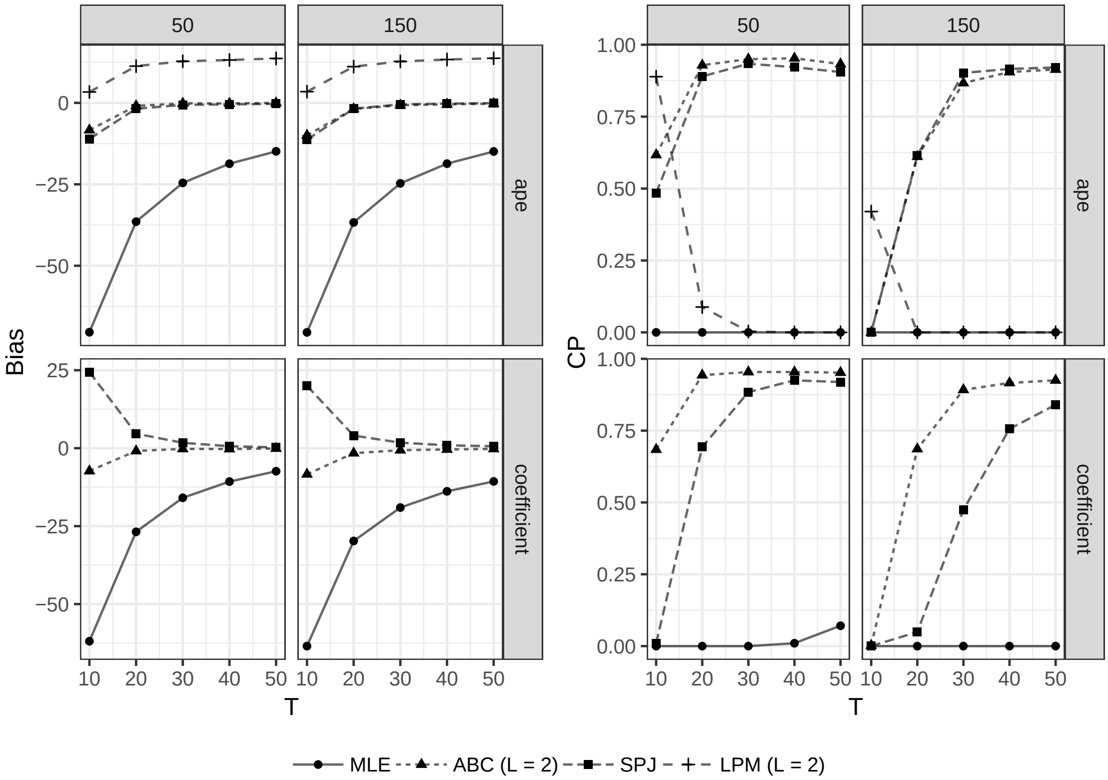
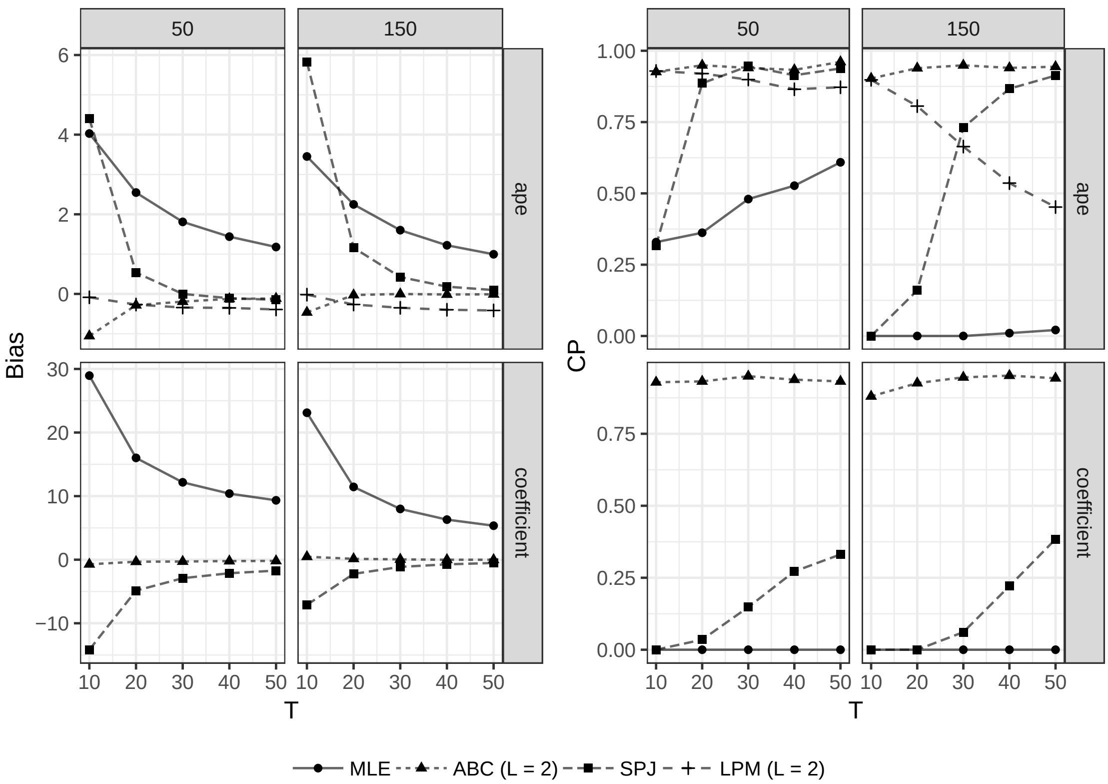


Figure 5: Dynamic: Three-way Fixed Effects – Exogenous Regressor



Overall, we obtain similar findings as for the structural parameters. MLE is distorted over all settings, but the bias decreases as T increases. The distortion is especially severe in the case of the predetermined regressor. Even at $T = 50$, MLE suffers a bias of 15 percent. The bias corrections bring a substantial reduction in this case. Whereas ABC shows only a small distortion of 1 percent in the case of the exogenous regressor at $T = 10$, SPJ is even more heavily distorted than MLE. However, with increasing T, both SPJ and ABC quickly converge to the true APE. Furthermore, unlike ABC, SPJ needs a sufficiently large number of time periods to get its CPs close to 95 percent. For the predetermined regressor, these convergence processes last longer for both bias corrections. Looking at LPM in the case of the exogenous regressors, it produces almost unbiased estimates irrespective of T, but its CPs fall dramatically with increasing T. Moreover, in the case of the predetermined regressor, we observe an increase in the bias up to 14 percent with increasing T.[19] These results illustrate the superiority of ABC and SPJ over LPM.

Overall, our three-way fixed effects simulation results confirm the conjecture of Fernández-Val and Weidner (2018) about the general form and lend support to our conjecture for the specific structure of the bias terms in the three-way fixed effects specification. First, we find that the bias corrections indeed substantially mitigate the bias. Second, as already found in other studies, analytical bias corrections clearly outperform split-panel jackknife bias corrections (see among others Fernández-Val and Weidner, 2016, and Czarnowske and Stammann, 2019). For samples with shorter time horizons, ABC is often less distorted and its dispersion is generally lower. This is also reflected by better CPs. Further, our three-way fixed effects simulation results suggest that estimates based on MLE or LPM should be treated with great caution. Generally, in the three-way fixed effects setting, a sufficiently large number of time periods appears to be crucial to obtain reliable results, even for the bias-corrected estimators.

[19] A similar behaviour of LPM has been observed by Czarnowske and Stammann (2019) in the context of a dynamic probit model with individual and time fixed effects. To ensure that the bias correction presented in Appendix B.4 in our three-way fixed effects specification is implemented correctly we have tested it in a data generation process for classical linear models, i.e. without binary dependent variables, and found that it works as intended. The undesirable behavior in our simulation design for the probit model is driven by the fact that, because of the autoregressive process of x, the predicted probabilities of LPM exceed the boundaries of the unit interval more and more frequently as T increases. This is particularly reflected in the APEs for binary regressors, since they are based on differences of predicted probabilities.

5 Determinants of the Extensive Margin of Trade

Having described the estimation and bias correction procedures, we now turn to the estimation of the determinants of the extensive margin of international trade outlined in section 2.

Recall equation (3) that relates the incidence of nonzero aggregate trade flows to exporter-time and importer-time specific characteristics, as well as trade in the previous period, next to fixed and variable trade costs:

$$y_{ijt} = \begin{cases} 1 & \text{if} \quad \kappa + \lambda_{it} + \psi_{jt} + \beta_y y_{ij(t-1)} + \mathbf{x}'_{ijt}\boldsymbol{\beta}_x \geq \zeta_{ijt}, \\ 0 & \text{else}\,. \end{cases}$$

This yields the following probit model:

$$\Pr(y_{ijt} = 1 | \mathbf{x}_{ijt}, y_{ij(t-1)}, \lambda_{it}, \psi_{jt}) = F\left(\mathbf{x}'_{ijt}\boldsymbol{\beta}_x + \beta_y y_{ij(t-1)} + \lambda_{it} + \psi_{jt}\right)\,, \tag{10}$$

in case we assume to capture bilateral variables and fixed trade costs entirely with observables, or:

$$y_{ijt} = \begin{cases} 1 & \text{if} \quad \kappa + \lambda_{it} + \psi_{jt} + \beta_y y_{ij(t-1)} + \mathbf{x}'_{ijt}\boldsymbol{\beta}_x + \mu_{ij} \geq \zeta_{ijt}, \\ 0 & \text{else} \end{cases}$$

and:

$$\Pr(y_{ijt} = 1 | \mathbf{x}_{ijt}, y_{ij(t-1)}, \lambda_{it}, \psi_{jt}, \mu_{ij}) = F\left(\mathbf{x}'_{ijt}\boldsymbol{\beta}_x + \beta_y y_{ij(t-1)} + \lambda_{it} + \psi_{jt} + \mu_{ij}\right)\,, \tag{11}$$

in case we include a time-invariant bilateral fixed effect to capture unobservable country pair characteristics. $y_{ij(t-1)}$ is the lagged dependent variable, $\mathbf{x}$ is a vector of observable bilateral variables, β_y and β_x are the corresponding parameters. We largely follow Helpman, Melitz, and Rubinstein (2008) and the wider literature on the determinants of the *intensive* margin of trade (compare Head and Mayer, 2014) in the choice of these variables: distance, a common land border, the same origin of the legal system, common language, previous colonial ties, a joint currency, an existing free trade agreement, or joint membership in the WTO. In terms of data, we turn to the comprehensive gravity dataset provided alongside Head, Mayer, and Ries (2010), which encompasses information on trade flows and these variables of interest from 1948 – 2006.

Table 2: Probit Estimation: Coefficients

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	-	-	1.664***	-	1.140***
	[-]	[-]	[1.719]	[-]	[1.057]
	(-)	(-)	(0.004)	(-)	(0.005)
log(Distance)	-	-0.800***	-0.528***	-	-
	[-0.656***]	[-0.821]	[-0.546]	[-]	[-]
	(0.003)	(0.003)	(0.004)	(-)	(-)
Land border	-	0.207***	0.118***	-	-
	[0.260***]	[0.214]	[0.124]	[-]	[-]
	(0.014)	(0.016)	(0.018)	(-)	(-)
Legal	-	0.137***	0.089***	-	-
	[0.090***]	[0.141]	[0.093]	[-]	[-]
	(0.004)	(0.004)	(0.005)	(-)	(-)
Language	-	0.426***	0.280***	-	-
	[0.380***]	[0.436]	[0.289]	[-]	[-]
	(0.005)	(0.006)	(0.007)	(-)	(-)
Colonial ties	-	0.657***	0.487***	-	-
	[0.190***]	[0.702]	[0.542]	[-]	[-]
	(0.02)	(0.031)	(0.036)	(-)	(-)
Currency Union	-	0.631***	0.424***	0.303***	0.214***
	[0.381***]	[0.649]	[0.443]	[0.335]	[0.255]
	(0.012)	(0.015)	(0.017)	(0.032)	(0.034)
FTA	-	0.543***	0.359***	0.073*	0.038
	[0.508***]	[0.552]	[0.364]	[0.072]	[0.033]
	(0.017)	(0.019)	(0.021)	(0.038)	(0.04)
WTO	-	0.152***	0.101***	0.052***	0.039**
	[0.286***]	[0.154]	[0.104]	[0.058]	[0.048]
	(0.005)	(0.008)	(0.009)	(0.016)	(0.017)
Fixed effects	i, j, t	it, jt	it, jt	it, jt, ij	it, jt, ij
Sample size	1204671	1204671	1171794	1204671	1171794
- perf. class.	12298	147760	141537	370617	374067
Deviance	8.891×10^5	7.019×10^5	5.183×10^5	4.76×10^5	4.189×10^5

Notes: Uncorrected coefficients in square brackets. Standard errors in parenthesis.

We report the bias-corrected coefficients in Table 2 and the corresponding average partial effects in Table 3.[20] For each uncorrected and (analytically) bias-corrected coefficients and average partial effects we also report the uncorrected one in square brackets, as well

[20]While the error term distribution assumed in section 2 suggests a probit estimator, we also estimate equations 10 and 11 with a logit estimator and show the corresponding results in Tables 26 and 27 in Appendix E. The coefficients and average partial effects are similar to those estimated with the probit model.

Table 3: Probit Estimation: Average partial effects

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	-	-	0.346***	-	0.179***
	[-]	[-]	[0.344]	[-]	[0.138]
	(-)	(-)	(0.003)	(-)	(0.052)
log(Distance)	-	-0.135***	-0.066***	-	-
	[-0.136***]	[-0.135]	[-0.066]	[-]	[-]
	(0.005)	(0.005)	(0.001)	(-)	(-)
Land border	-	0.035***	0.015***	-	-
	[0.054***]	[0.035]	[0.015]	[-]	[-]
	(0.004)	(0.004)	(0.003)	(-)	(-)
Legal	-	0.023***	0.011***	-	-
	[0.019***]	[0.023]	[0.011]	[-]	[-]
	(0.001)	(0.001)	(0.001)	(-)	(-)
Language	-	0.071***	0.035***	-	-
	[0.078***]	[0.071]	[0.035]	[-]	[-]
	(0.003)	(0.001)	(0.001)	(-)	(-)
Colonial ties	-	0.107***	0.061***	-	-
	[0.039***]	[0.111]	[0.066]	[-]	[-]
	(0.004)	(0.007)	(0.005)	(-)	(-)
Currency Union	-	0.103***	0.053***	0.038***	0.024***
	[0.078***]	[0.103]	[0.054]	[0.037]	[0.025]
	(0.004)	(0.003)	(0.002)	(0.005)	(0.009)
FTA	-	0.090***	0.045***	0.009	0.004
	[0.103***]	[0.088]	[0.044]	[0.008]	[0.003]
	(0.005)	(0.004)	(0.003)	(0.007)	(0.006)
WTO	-	0.026***	0.013***	0.006**	0.004
	[0.061***]	[0.026]	[0.013]	[0.006]	[0.005]
	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)
Fixed effects	i, j, t	it, jt	it, jt	it, jt, ij	it, jt, ij
Sample size	1204671	1204671	1171794	1204671	1171794
- perf. class.	12298	147760	141537	370617	374067
Deviance	8.891×10^5	7.019×10^5	5.183×10^5	4.76×10^5	4.189×10^5

Notes: Uncorrected average partial effects in square brackets. Standard errors in parenthesis.

as the standard error in parenthesis below. In column (1) we first mimic the specification estimated by Helpman, Melitz, and Rubinstein (2008).[21] Their specification includes

[21] Helpman, Melitz, and Rubinstein (2008) use a dataset that ranges from 1970 to 1997. They also include dummy variables for whether both countries are landlocked or islands, or follow the same religion. Hence our coefficients deviate somewhat from theirs, while remaining qualitatively similar.

exporter, importer and time fixed effects.[22] All coefficients have the expected sign, i.e. a negative impact of distance on the probability to trade, while all other variables are estimated to have a positive impact. Note the strong and highly significant impact of a common currency, free trade agreement or joint membership of the WTO. Ceteris paribus, each is estimated to increase the probability of nonzero flows by between 6 and 10 percentage points. Column (2) introduces a stricter set of fixed effects, namely at the exporter-time and importer-time level. Most coefficients and average partial effects are similar to those in column (1). This changes in column (3), which keeps the same fixed effects, but adds a lagged dependent variable. Assuming no unobservable bilateral heterogeneity, as in equation (10), this specification correctly estimates the model set up in section 2. The first thing to note is the highly significant coefficient for the lagged dependent variable, which reflects the strong impact of previous nonzero trade flows on current ones. Ceteris paribus, the average partial effect shows a 34 percentage points higher probability of nonzero trade, given the two countries were also engaged in trade in the previous year. The second observation is that essentially all coefficients are remarkably smaller than those in column (2), and average partial effects are reduced by about 50 percent across the board. This result underlines the need to explicitly take persistence into account. Note, however, that the APEs of the two specifications are not directly comparable, because the static model forces immediate effects and long-run dynamic adjustments into a single estimate.

Column (4) then takes one step back and one forward. While not including the lagged dependent variable in the estimation, it introduces a bilateral fixed effect that controls for bilateral unobserved heterogeneity. This follows the important insight by Baier and Bergstrand (2007), who show that controlling for unobserved bilateral heterogeneity produces a considerably different estimated impact of free trade agreements, among other variables, on the intensive margin of trade. While now an identification of many of the variables of interest is no longer possible because of their time invariance, this specification reveals a much reduced estimated impact of the time-varying variables. The impact of a common currency on the probability of exporting is reduced to 3.8 percentage points, while those of a common free trade agreement and WTO are decreased to less than 1 percentage point. This result highlights the importance of controlling for unobserved country pair heterogeneity and possible endogeneity. Finally, in column (5) we present our preferred specification, estimating equation (11). The estimation now

[22]Note that following Fernández-Val and Weidner (2018) the incidental bias problem is small enough to ignore in this setting with i, j and t fixed effects, since the order of the bias is $1/IT + 1/JT + 1/IJ$, which in our case becomes negligible small since I, J and T are large.

Table 4: Probit vs. OLS Estimation: Average Partial Effects with Three-way Fixed Effects

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	-	-	0.444***	0.474***	0.179***
	(-)	(-)	(0.001)	(0.001)	(0.052)
Currency Union	0.009***	0.038***	0.008***	0.008**	0.024***
	(0.003)	(0.005)	(0.003)	(0.003)	(0.009)
FTA	-0.121***	0.009	-0.065***	-0.062***	0.004
	(0.003)	(0.007)	(0.002)	(0.002)	(0.006)
WTO	0.017***	0.006**	0.008***	0.008***	0.004
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)
Estimator	OLS	Probit	OLS	OLS	Probit
bias corrected	-	true	false	true	true
Sample size	1204671	1204671	1171794	1171794	1171794

Notes: All columns include Origin $\times$ Year and Destination $\times$ Year fixed effects. Standard errors in parenthesis.

includes the "full set" of fixed effects, i.e. exporter-time, importer-time and bilateral fixed effect, in addition to the lagged dependent variable.[23] Again, the coefficient on the latter is highly significant, entailing an average partial effect of about 18 percentage points. Importantly, the only remaining statistically significant average partial effect is estimated for a common currency at 2.4 percentage points. The impact of a free trade agreement or joint membership of the WTO are statistically insignificant.

Contrasting the results from column (5) to those of column (1), which currently constitutes the de-facto standard of estimating the determinants of the extensive margin of trade, underlines the importance of (i) appropriate exporter-time and importer-time fixed effects that capture all country-time specific variation; (ii) country pair fixed effects that capture all unobserved bilateral heterogeneity and address endogeneity concerns, analogous to Baier and Bergstrand (2007) on the intensive margin; (iii) dynamics, in that country pairs that have previously traded are significantly more likely to do so than otherwise comparable country pairs. This corroborates the stylized facts from section 1, which showed country pairs that had previously engaged in trade to be likely to do so again in the next year. Failing to observe any of these three insights produces widely different estimates.

[23] Note that in the analytical bias correction we set the bandwidth parameter to $L = 2$. We report results for $L \in \{0, 1, 2, 3, 4\}$ in Tables 28 to 33 in Appendix E. The results remain robust with $L = 1 - 4$.

Another important insight is that the magnitude of the incidental parameter problem — at least in this specific setting — is not as severe as one might have feared. The most significant impact is observed on the coefficient for the lagged dependent variable, which in Table 2 column (5) differs by about 10 percent, and even almost 24 percent in the respective average partial effect reported in Table 3 column (5). However, this does not carry through to other variables, in particular for average partial effects. As shown in simulations in section 4, this may not come as a big surprise. In this application we consider a panel that covers 57 years, meaning the relatively large T inhibits a strong bias (e.g. compare Figure 5). As shown in the simulations, the bias is more severe in settings with fewer time periods and should be handled appropriately.

To show the superiority of using suitable binary choice estimators with high-dimensional fixed effects we also contrast the results to estimating equations (10) and (11) with a linear probability model. Table 4 shows that OLS with the same set of three-way fixed effects produces estimates that are far off the probit ones.[24] Columns (1) and (2) compare estimates without, columns (3) to (5) those with a lagged dependent variable.[25] Figure 8 underlines this impression: the LPM produces up to 28 percent of fitted probabilities < 0 or > 1. This result highlights that binary choice estimators with high-dimensional fixed effects cannot easily be mimicked by an OLS estimator.

6 Conclusion

In this paper we reexamine the determinants of the extensive margin of international trade. We set up a model that exhibits a dynamic component and allows for time-invariant unobserved bilateral trade cost factors, generating persistence — a feature in the data that has so far been given little attention. We estimate the model using a probit estimator with high-dimensional fixed effects. As fixed effects create an incidental parameter problem in binary choice settings, we characterize and implement bias corrections for estimations with appropriate two- and three-way fixed effects. Finally, we show that our estimates of the determinants of the extensive margin of trade differ significantly from previous ones. This highlights the importance of true state dependence and unobserved heterogeneity and therefore strongly supports the use of our bias-corrected dynamic fixed effects estimator.

[24] As for the probit estimates, we also report the bias-corrected LPM estimates with different bandwidth parameters in Table 33. All in all, the results remain robust with $L = 1 - 4$. We also report estimates for two-way fixed effects in Table 32 in Appendix E.

[25] In column (3) we ignore and in column (4) we apply the appropriate bias correction for the LPM with endogenous regressor, as detailed in Appendix B.4.

The extensive margin of trade obviously extends beyond the aggregate level, warranting further research at lower levels of aggregation, in particular in the context of firms. While our model's prediction and its empirical specification rely on some abstractions, it provides a very tractable and flexible framework that can be estimated with recently established estimation procedures, when combined with the bias correction technique we introduce.

Appendix

A Stylized facts

Figure 6: Determinants of the Extensive margin of Trade — Gravity and Persistence (1990–1991).

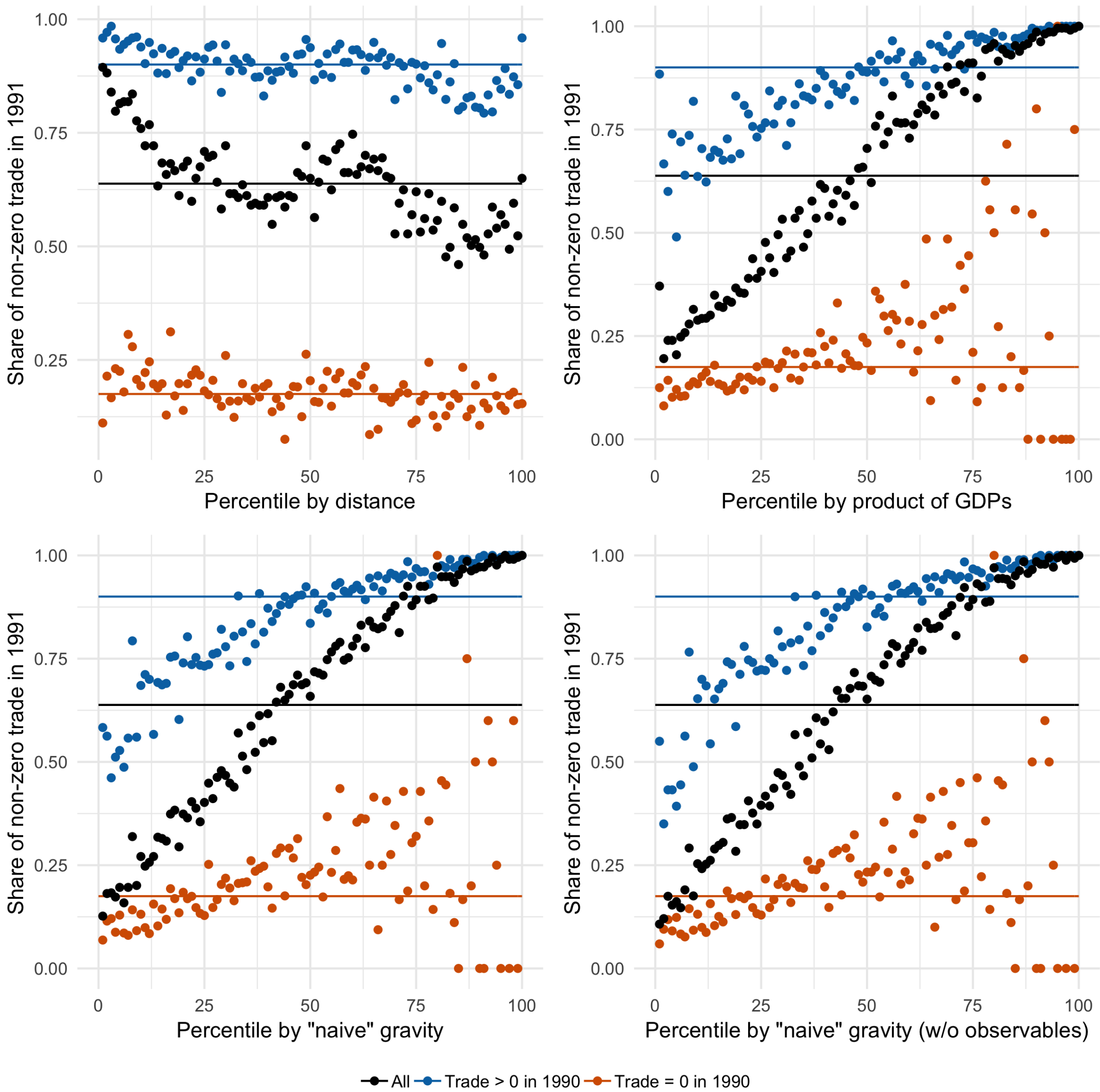
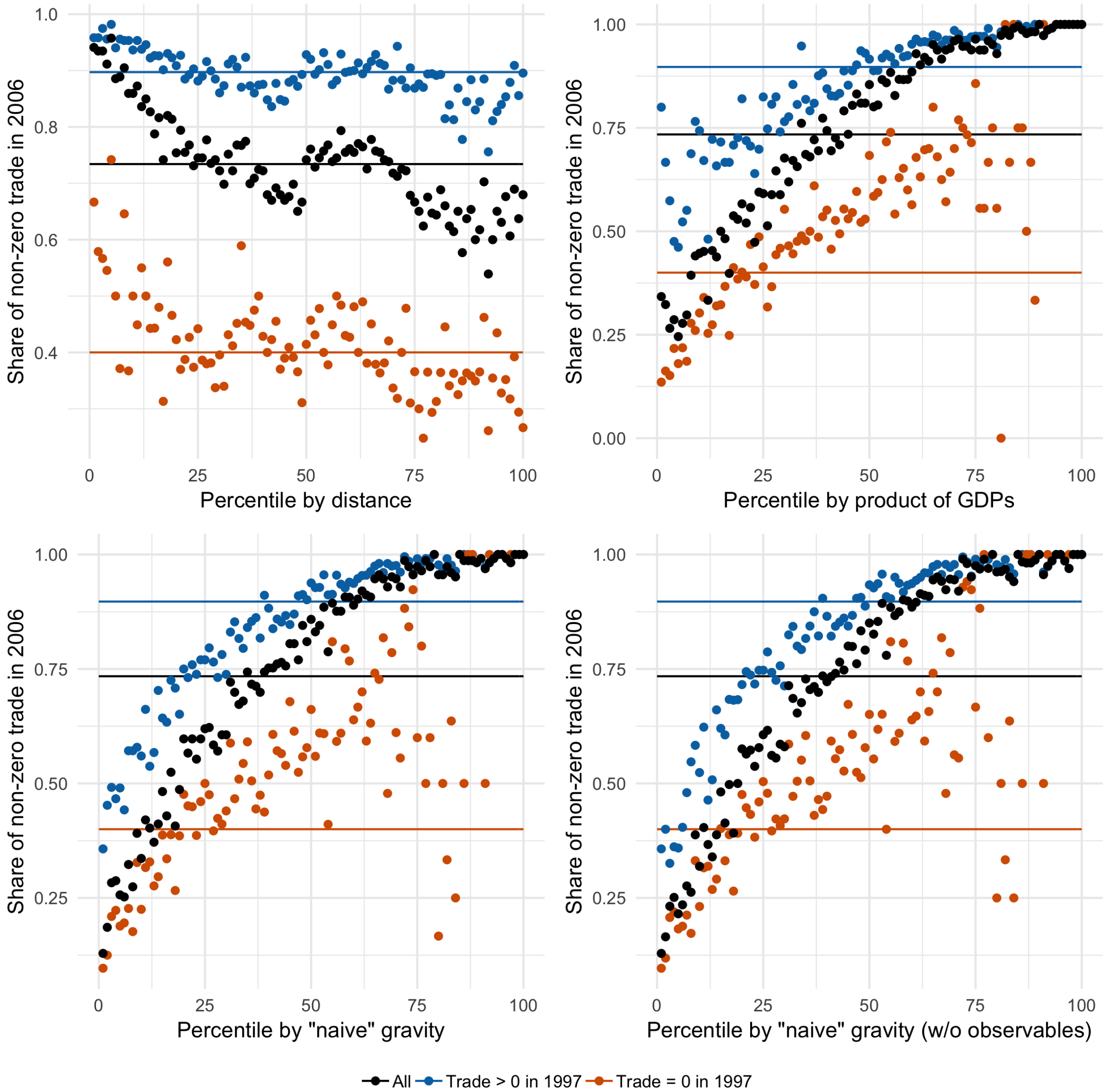


Figure 7: Determinants of the Extensive Margin of Trade — Gravity and Persistence (1997–2006).



B Computational and Econometric Details

B.1 Computational Details

In this section we briefly demonstrate how the method of alternating projections (MAP) works in the context of logit and probit models with a two- or three-way error component, and how it can be efficiently embedded into a standard Newton-Raphson optimization routine (see Stammann, 2018, for further details).

First, note that $\mathbb{M}\,\mathbf{v}$ is essentially a weighted within transformation, where $\mathbf{v}$ is an arbitrary $n \times 1$ vector, and $\mathbb{M} = \mathbf{I}_n - \mathbb{P} = \mathbf{I}_n - \mathbf{D}(\mathbf{D}'\boldsymbol{\Omega}\mathbf{D})^{-1}\mathbf{D}'\boldsymbol{\Omega}$. The computation of $\mathbb{M}$ is problematic even in moderately large data sets, and since $\mathbb{M}$ is non-sparse, there is also no general scalar expression to compute $\mathbb{M}\,\mathbf{v}$. Thus Stammann (2018) proposes to calculate $\mathbb{M}\,\mathbf{v}$ using a simple iterative approach based on the MAP tracing back to Von Neumann (1950) and Halperin (1962).[26] Let $\mathbf{D}_k$, denote the dummy variables corresponding to the k-th group, $k \in \{1, 2, 3\}$. Further, let $\mathbb{M}_{\mathbf{D}_k}\,\mathbf{v}$, with $\mathbb{M}_{\mathbf{D}_k} = \mathbf{I}_n - \mathbf{D}_k(\mathbf{D}_k'\boldsymbol{\Omega}\mathbf{D}_k)^{-1}\mathbf{D}_k'\boldsymbol{\Omega}$. The corresponding scalar expressions of $\mathbb{M}_{\mathbf{D}_k}\,\mathbf{v}$ are summarized in Table (5).

Table 5: Scalar Transformations

group	$\mathbb{M}_{\mathbf{D}_k}\,\mathbf{v}$
importer-time ($k = 1$)	$\mathbf{v}_{ijt} - \frac{\sum_{j=1}^{J} \omega_{ijt} v_{ijt}}{\sum_{j=1}^{J} \omega_{ijt}}$
exporter-time ($k = 2$)	$\mathbf{v}_{ijt} - \frac{\sum_{i=1}^{I} \omega_{ijt} v_{ijt}}{\sum_{i=1}^{I} \omega_{ijt}}$
dyadic ($k = 3$)	$\mathbf{v}_{ijt} - \frac{\sum_{t=1}^{T} \omega_{ijt} v_{ijt}}{\sum_{t=1}^{T} \omega_{ijt}}$

The MAP can be summarized by algorithm 1, where $K = 2$ in the case of two-way fixed effects and $K = 3$ in the case of three-way fixed effects. Thus, the MAP only requires to repeatedly apply weighted one-way within transformations (see Stammann, 2018)). The entire optimization routine is sketched by algorithm 2.

[26] The MAP has been introduced to econometrics by Guimarães and Portugal (2010) and Gaure (2013) in the context of linear models with multi-way fixed effects.

Algorithm 1 MAP: Neumann-Halperin

Initialize $\mathbb{M}\,\mathbf{v} = \mathbf{v}$.
repeat
  for $k = 1, \ldots, K$ **do**
    Compute $\mathbb{M}_{\mathbf{D}_k}\,\mathbb{M}\,\mathbf{v}$ and update $\mathbb{M}\,\mathbf{v}$ such that $\mathbb{M}\,\mathbf{v} = \mathbb{M}_{\mathbf{D}_k}\,\mathbb{M}\,\mathbf{v}$
until convergence.

Algorithm 2 Efficient Newton-Raphson using the MAP

Initialize $\boldsymbol{\beta}^0$, $\boldsymbol{\eta}^0$, and $r = 0$.
repeat
  Set $r = r + 1$.
  Given $\hat{\boldsymbol{\eta}}^{r-1}$ compute $\hat{\boldsymbol{\nu}}$ and $\widehat{\boldsymbol{\Omega}}$.
  Given $\hat{\boldsymbol{\nu}}$ and $\widehat{\boldsymbol{\Omega}}$ compute $\widehat{\mathbb{M}}\hat{\boldsymbol{\nu}}$ and $\widehat{\mathbb{M}}\mathbf{X}$ using the MAP
  Compute $\boldsymbol{\beta}^r - \boldsymbol{\beta}^{r-1} = \left((\widehat{\mathbb{M}}\mathbf{X})'\widehat{\boldsymbol{\Omega}}(\widehat{\mathbb{M}}\mathbf{X})\right)^{-1}(\widehat{\mathbb{M}}\mathbf{X})'\widehat{\boldsymbol{\Omega}}(\widehat{\mathbb{M}}\hat{\boldsymbol{\nu}})$
  Compute $\hat{\boldsymbol{\eta}}^r = \hat{\boldsymbol{\eta}}^{r-1} + \hat{\boldsymbol{\nu}} - \widehat{\mathbb{M}}\hat{\boldsymbol{\nu}} + \widehat{\mathbb{M}}\mathbf{X}(\boldsymbol{\beta}^r - \boldsymbol{\beta}^{r-1})$
until convergence.

B.2 Neyman-Scott Variance Example

In this section we study two variants of the classical Neyman and Scott (1948) variance example to support the form of the bias terms, and to illustrate the functionality of the bias corrections. To the best of our knowledge, the variance example of Neyman and Scott (1948) has not been investigated for our specific error components. We start with the more general three-way fixed effects case, which nests the two-way error structure.

B.2.1 Three-way Fixed Effects

Let $i = 1, \ldots, I$, $j = 1, \ldots, J$ and $t = 1, \ldots, T$. Consider the following linear three-way fixed effects model

$$y_{ijt} = \mathbf{x}_{ijt}'\boldsymbol{\beta} + \lambda_{it} + \psi_{jt} + \mu_{ij} + u_{ijt} \; . \tag{12}$$

According to Balazsi, Matyas, and Wansbeek (2018), the appropriate within transformation corresponding to equation (12) is given by

$$z_{ijt} - \bar{z}_{ij\cdot} - \bar{z}_{\cdot jt} - \bar{z}_{i\cdot t} + \bar{z}_{\cdot\cdot t} + \bar{z}_{\cdot j\cdot} + \bar{z}_{i\cdot\cdot} - \bar{z}_{\cdot\cdot\cdot} \; ,$$

where $\bar{z}_{ij\cdot} = \frac{1}{T}\sum_{t=1}^{T} z_{ijt}$, $\bar{z}_{\cdot jt} = \frac{1}{I}\sum_{i=1}^{I} z_{ijt}$, $\bar{z}_{i\cdot t} = \frac{1}{J}\sum_{j=1}^{J} z_{ijt}$, $\bar{z}_{\cdot\cdot t} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} z_{ijt}$,

$\bar{z}_{\cdot j \cdot} = \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=1}^{T} z_{ijt}$, $\bar{z}_{i\cdot\cdot} = \frac{1}{JT}\sum_{j=1}^{J}\sum_{t=1}^{T} z_{ijt}$, and $\bar{z}_{\cdot\cdot\cdot} = \frac{1}{IJT}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} z_{ijt}$.

This result is helpful to study the following variant of the Neyman and Scott (1948) variance example

$$y_{ijt}|\boldsymbol{\lambda}, \boldsymbol{\psi}, \boldsymbol{\mu} \sim \mathcal{N}(\lambda_{it} + \psi_{jt} + \mu_{ij}, \beta) \ ,$$

where we can now easily form the uncorrected variance estimator

$$\hat{\beta}_{I,J,T} = \frac{1}{IJT}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}(y_{ijt} - \bar{y}_{ij\cdot} - \bar{y}_{\cdot jt} - \bar{y}_{i\cdot t} + \bar{y}_{\cdot\cdot t} + \bar{y}_{\cdot j\cdot} + \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2 \tag{13}$$

and the (degrees-of-freedom)-corrected counterpart

$$\hat{\beta}^{cor}_{I,J,T} = \frac{IJT}{(I-1)(J-1)(T-1)}\hat{\beta}_{I,J,T} \ .$$

Taking the expectation of (13) (conditional on the fixed effects) yields

$$\begin{aligned} \bar{\beta}_{I,J,T} = \mathbb{E}_\alpha[\hat{\beta}_{I,J,T}] &= \beta^0\left(\frac{(I-1)(J-1)(T-1)}{IJT}\right) \qquad (14) \\ &= \beta^0\left(1 - \frac{1}{I} - \frac{1}{J} - \frac{1}{T} + \frac{1}{IT} + \frac{1}{JT} + \frac{1}{IJ} - \frac{1}{IJT}\right) \ , \end{aligned}$$

where β^0 is the true variance parameter. Thus, the three leading bias terms, which drive the main part of the asymptotic bias, are $\overline{\mathbf{B}}^{\beta}_{1,\infty} = -\beta^0$, $\overline{\mathbf{B}}^{\beta}_{2,\infty} = -\beta^0$, and $\overline{\mathbf{B}}^{\beta}_{3,\infty} = -\beta^0$.

Analytical Bias Correction

Using equation (14), we can form the analytically bias-corrected estimator

$$\tilde{\beta}^{a}_{I,J,T} = \hat{\beta}_{I,J,T} - \frac{\widehat{\mathbf{B}}^{\beta}_{1,I,J,T}}{I} - \frac{\widehat{\mathbf{B}}^{\beta}_{2,I,J,T}}{J} - \frac{\widehat{\mathbf{B}}^{\beta}_{3,I,J,T}}{T} \ , \tag{15}$$

where we set $\widehat{\mathbf{B}}^{\beta}_{1,I,J,T} = -\hat{\beta}_{I,J,T}$, $\widehat{\mathbf{B}}^{\beta}_{2,I,J,T} = -\hat{\beta}_{I,J,T}$, and $\widehat{\mathbf{B}}^{\beta}_{3,I,J,T} = -\hat{\beta}_{I,J,T}$ to reduce the order of the bias in equation (14) at costs of introducing higher order terms (see equation (17)). Thus, we can rewrite the analytically bias-corrected estimator (15)

$$\tilde{\beta}^{a}_{I,J,T} = \hat{\beta}_{I,J,T}\left(1 + \frac{1}{I} + \frac{1}{J} + \frac{1}{T}\right) \ . \tag{16}$$

Taking the expectation of (16) yields

$$\bar{\beta}^a_{I,J,T} = \mathbb{E}_\alpha[\tilde{\beta}^a_{I,J,T}] = \beta^0 \left(1 - \frac{1}{I} - \frac{1}{J} - \frac{1}{T} + \frac{1}{IT} + \frac{1}{JT} + \frac{1}{IJ} - \frac{1}{IJT}\right)\left(1 + \frac{1}{I} + \frac{1}{J} + \frac{1}{T}\right) \tag{17}$$

$$= \beta^0 \left(1 - \frac{1}{IT} - \frac{1}{JT} - \frac{1}{T^2} - \frac{3}{IJ} + \frac{1}{I^3} + \frac{1}{J^3} + \frac{4}{IJT} + \frac{1}{IT^2} + \frac{1}{JT^2} - \frac{1}{I^3 T} - \frac{1}{J^3 T} - \frac{1}{IJT^2}\right) .$$

Split-Panel Jackknife

As an alternative to equation (16) we can also form the following SPJ estimator

$$\hat{\beta}^{spj}_{I,J,T} = 4\hat{\beta}_{I,J,T} - \hat{\beta}_{I/2,J,T} - \hat{\beta}_{I,J/2,T} - \hat{\beta}_{I,J,T/2} ,$$

where $\hat{\beta}_{I/2,J,T}$ denotes the half panel estimator based on splitting the panel by exporters. This estimator also reduces the order of the bias in equation (14) as we see from its expected value

$$\bar{\beta}^{spj}_{I,J,T} = E_\phi[\hat{\beta}^{spj}_{I,J,T}] = 4\bar{\beta}_{I,J,T} - \bar{\beta}_{I/2,J,T} - \bar{\beta}_{I,J/2,T} - \bar{\beta}_{I,J,T/2} \tag{18}$$

$$= \beta^0 \left(1 - \frac{1}{IT} - \frac{1}{JT} - \frac{1}{IJ} + \frac{2}{IJT}\right) .$$

Numerical Results

Table 6 shows numerical results for the uncorrected and the bias-corrected estimators in finite samples, where we assume symmetry, i.e. $I = J = N$. The results demonstrate that the bias corrections are effective in reducing the bias.

Table 6: Bias - Three-way Fixed Effects

N	T	$(\bar{\beta}_{I,J,T} - \beta^0)/\beta^0$	$(\bar{\beta}^a_{I,J,T} - \beta^0)/\beta^0$	$(\bar{\beta}^{spj}_{I,J,T} - \beta^0)/\beta^0$
10	10	-0.271	-0.052	-0.028
25	10	-0.171	-0.021	-0.009
25	25	-0.115	-0.009	-0.005
50	10	-0.136	-0.015	-0.004
50	25	-0.078	-0.004	-0.002
50	50	-0.059	-0.002	-0.001

B.2.2 Two-way Fixed Effects

In the following we briefly review the example with two-way fixed effects

$$y_{ijt}|\boldsymbol{\lambda}, \boldsymbol{\psi} \sim \mathcal{N}(\lambda_{it} + \psi_{jt}, \beta) \ .$$

Since it is a subcase of three-way fixed effects example, all previous results simplify by dropping the terms that exhibit T.
The uncorrected variance estimator is[27]

$$\hat{\beta}_{I,J,T} = \frac{1}{IJT} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} (y_{ijt} - \bar{y}_{\cdot jt} - \bar{y}_{i \cdot t} + \bar{y}_{\cdot\cdot t})^2$$

and the (degrees-of-freedom)-corrected variance estimator is

$$\hat{\beta}^{cor}_{I,J,T} = \frac{IJ}{(I-1)(J-1)} \hat{\beta}_{I,J,T} \ .$$

Taking the expected value yields

$$\begin{aligned} \bar{\beta}_{I,J,T} = \mathbb{E}_\alpha[\hat{\beta}_{I,J,T}] &= \beta^0 \left(\frac{(I-1)^2}{IJ} \right) \\ &= \beta^0 \left(1 - \frac{1}{I} - \frac{1}{J} + \frac{1}{IJ} \right) \ . \end{aligned} \tag{19}$$

Analytical Bias Correction

Based on equation (19) we can form the following analytically bias-corrected estimator

$$\tilde{\beta}^a_{I,J,T} = \hat{\beta}_{I,J,T} \left(1 + \frac{1}{I} + \frac{1}{J} \right) \ ,$$

which has the expected value

$$\bar{\beta}^a_{I,J,T} = \mathbb{E}_\alpha[\tilde{\beta}^a_{I,J,T}] = \beta^0 \left(1 - \frac{3}{IJ} + \frac{1}{I^3} + \frac{1}{J^3} \right) \ .$$

Split-Panel Jackknife

A suitable split-panel jackknife estimator is

[27] We draw on the appropriate demeaning formula for the two-way fixed effects model $y_{ijt} = \boldsymbol{x}'_{ijt}\boldsymbol{\beta} + \lambda_{it} + \psi_{jt} + u_{ijt}$, which is given by $z_{ijt} - \bar{z}_{\cdot jt} - \bar{z}_{i \cdot t} + \bar{z}_{\cdot\cdot t}$.

$$\hat{\beta}^{spj}_{I,J,T} = 4\hat{\beta}_{I,J,T} - \hat{\beta}_{I/2,J,T} - \hat{\beta}_{I,J/2,T} \ ,$$

which has the expected value

$$\begin{aligned}\bar{\beta}^{spj}_{I,J,T} = \mathbb{E}_\alpha[\hat{\beta}^{spj}_{I,J,T}] &= 3\bar{\beta}_{I,J,T} - \bar{\beta}_{I/2,J,T} - \bar{\beta}_{I,J/2,T} \\ &= \beta^0 \left(1 - \frac{1}{IJ}\right) \ .\end{aligned}$$

Numerical Results

The numerical results in Table 7 demonstrate that the bias corrections work.

Table 7: Bias - Two-way Fixed Effects

N	$(\bar{\beta}_{I,J,T} - \beta^0)/\beta^0$	$(\bar{\beta}^a_{I,J,T} - \beta^0)/\beta^0$	$(\bar{\beta}^{spj}_{I,J,T} - \beta^0)/\beta^0$
10	-0.190	-0.028	-0.010
25	-0.078	-0.005	-0.002
50	-0.040	-0.001	-0.000
100	-0.020	-0.000	-0.000

B.3 Asymptotic Bias Corrections

For the following expressions we draw on the results of Fernández-Val and Weidner (2016), who have already derived the asymptotic distributions of the MLE estimators for structural parameters and APEs in classical two-way fixed effects models based on it-panels. As outlined in Cruz-Gonzalez, Fernández-Val, and Weidner (2017) the bias corrections of Fernández-Val and Weidner (2016) can easily be adjusted to two-way fixed effects models based on pseudo-panels with an ij-structure (i corresponds to importer and j to exporter), and importer and exporter fixed effects. We give an intuitive explanation. Since only J observations are informative per exporter fixed effects, we get a bias of order J for including exporter fixed effects, and vice versa a bias of order I for including importer fixed effects. Further, since there are no predetermined regressors in an ij-structure, we get two symmetric bias terms

$$\overline{\mathbf{B}}_{1,\infty} = \text{plim}_{I,J\to\infty}\left[-\frac{1}{2J}\sum_{j=1}^{J}\frac{\sum_{i=1}^{I}\mathbb{E}_\alpha[H_{ij}\partial_{\eta^2}F_{ij}(\mathbb{M}\,\mathbf{X})_{ij}]}{\sum_{i=1}^{I}\mathbb{E}_\alpha[\omega_{ij}]}\right], \tag{20}$$

$$\overline{\mathbf{B}}_{2,\infty} = \text{plim}_{I,J\to\infty}\left[-\frac{1}{2I}\sum_{i=1}^{I}\frac{\sum_{j=1}^{J}\mathbb{E}_\alpha[H_{ij}\partial_{\eta^2}F_{ij}(\mathbb{M}\,\mathbf{X})_{ij}]}{\sum_{j=1}^{J}\mathbb{E}_\alpha[\omega_{ij}]}\right], \tag{21}$$

where ω_{ij} is the ij-th diagonal entry of $\mathbf{\Omega}$, and $\mathbb{M} = \mathbf{I}_{IJ} - \mathbf{D}(\mathbf{D}'\mathbf{\Omega}\mathbf{D})^{-1}\mathbf{D}'\mathbf{\Omega}$. $\partial_{\iota^2}g(\cdot)$ denotes the second order partial derivative of an arbitrary function $g(\cdot)$ with respect to some parameter ι. The explicit expressions of H_{ijt} and $\partial_{\eta^2}F_{ijt}$ are reported in Table 1. Equations (20) and (21) are essentially $\overline{\mathbf{D}}_\infty$ from Fernández-Val and Weidner (2016) with adjusted indices. The same adjustment can be transferred to the APEs.

In the following we apply the same logic to derive the asymptotic bias terms in our two- and three-way error structure.

B.3.1 Two-way fixed effects

We get a bias of order J for including exporter-time fixed effects, since J observations are informative per exporter-time fixed effect. In the same way we get a bias of order I for including importer-time fixed effects. Similar to the case of the ij-structure of Cruz-Gonzalez, Fernández-Val, and Weidner (2017) we get two symmetric bias terms in the distributions of the structural parameters and the APEs, respectively, because including predetermined regressors does not violate the strict exogeneity assumption.

Asymptotic distribution of $\hat{\beta}$

$$\sqrt{IJ}(\widehat{\boldsymbol{\beta}}_{I,J,T} - \boldsymbol{\beta}^0) \to_d \overline{\mathbf{W}}_\infty^{-1}\mathcal{N}(\kappa\overline{\mathbf{B}}_{1,\infty} + \kappa^{-1}\overline{\mathbf{B}}_{2,\infty}, \overline{\mathbf{W}}_\infty), \quad \text{with} \tag{22}$$

$$\overline{\mathbf{B}}_{1,\infty} = \text{plim}_{I,J\to\infty}\left[-\frac{1}{2J}\sum_{t=1}^{T}\sum_{j=1}^{J}\frac{\sum_{i=1}^{I}\mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}]}{\sum_{i=1}^{I}\mathbb{E}_\alpha[\omega_{ijt}]}\right],$$

$$\overline{\mathbf{B}}_{2,\infty} = \text{plim}_{I,J\to\infty}\left[-\frac{1}{2I}\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{\sum_{j=1}^{J}\mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}]}{\sum_{j=1}^{J}\mathbb{E}_\alpha[\omega_{ijt}]}\right],$$

$$\overline{\mathbf{W}}_\infty = \text{plim}_{I,J\to\infty}\left[\frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\mathbb{E}_\alpha[\omega_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}(\mathbb{M}\,\mathbf{X})'_{ijt}]\right],$$

where $\sqrt{J/I} \to \kappa$ as $I, J \to \infty$.

Asymptotic distribution of $\hat{\delta}$

$$r(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta} - I^{-1}\overline{\mathbf{B}}^{\delta}_{1,\infty} - J^{-1}\overline{\mathbf{B}}^{\delta}_{2,\infty}) \to_d \mathcal{N}(0, \overline{\mathbf{V}}_{\infty}), \quad \text{with} \tag{23}$$

$$\overline{\mathbf{B}}^{\delta}_{1,\infty} = \text{plim}_{I,J\to\infty} \left[\frac{1}{2JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \frac{\sum_{i=1}^{I} - \mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2} F_{ijt}]\,\mathbb{E}_\alpha[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}] + \mathbb{E}_\alpha[\partial_{\eta^2}\boldsymbol{\Delta}_{ijt}]}{\sum_{i=1}^{I}\mathbb{E}_\alpha[\omega_{ijt}]} \right],$$

$$\overline{\mathbf{B}}^{\delta}_{2,\infty} = \text{plim}_{I,J\to\infty} \left[\frac{1}{2IT} \sum_{t=1}^{T} \sum_{i=1}^{I} \frac{\sum_{j=1}^{J} - \mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2} F_{ijt}]\,\mathbb{E}_\alpha[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}] + \mathbb{E}_\alpha[\partial_{\eta^2}\boldsymbol{\Delta}_{ijt}]}{\sum_{j=1}^{J}\mathbb{E}_\alpha[\omega_{ijt}]} \right],$$

$$\overline{\mathbf{V}}^{\delta}_{\infty} = \text{plim}_{I,J\to\infty} \frac{r^2}{I^2J^2T^2}\,\mathbb{E}_\alpha \left[\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \bar{\boldsymbol{\Delta}}_{ijt} \right) \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \bar{\boldsymbol{\Delta}}_{ijt} \right)' + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \boldsymbol{\Gamma}_{ijt}\boldsymbol{\Gamma}'_{ijt} \right],$$

where $\bar{\boldsymbol{\Delta}}_{ijt} = \boldsymbol{\Delta}_{ijt} - \boldsymbol{\delta}$, $\boldsymbol{\Delta}_{ijt} = [\Delta^1_{ijt}, \ldots, \Delta^m_{ijt}]'$, $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_m]'$, $\delta_k = \frac{1}{IJT}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\Delta^k_{ijt}$, $\boldsymbol{\Psi}_{ijt} = \partial_\eta \boldsymbol{\Delta}_{ijt}/\omega_{ijt}$, r is a convergence rate, and

$$\begin{aligned}\boldsymbol{\Gamma}_{ijt} =& \mathbb{E}_\alpha \left[(IJ)^{-1} \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \partial_\beta \boldsymbol{\Delta}_{ijt} - (\mathbb{P}\,\mathbf{X})_{ijt}\,\partial_\eta \boldsymbol{\Delta}_{ijt} \right]' \overline{\mathbf{W}}^{-1}_{\infty}\,\mathbb{E}_\alpha \left[(\mathbb{M}\,\mathbf{X})_{ijt}\,\omega_{ijt}\boldsymbol{\nu}_{ijt} \right] \\ &- \mathbb{E}_\alpha \left[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}\,\partial_\eta \ell_{ijt} \right].\end{aligned}$$

$\partial_\iota g(\cdot)$ denotes the first order partial derivative of an arbitrary function $g(\cdot)$ with respect to some parameter ι. The expression $\overline{\mathbf{V}}^{\delta}_{\infty}$ can be modified by assuming that $\{\lambda_{it}\}_{IT}$ and $\{\psi_{jt}\}_{JT}$ are independent sequences, and λ_{it} and ψ_{jt} are independent for all it, jt:

$$\begin{aligned}\overline{\mathbf{V}}^{\delta}_{\infty} = \text{plim}_{I,J\to\infty} \frac{r^2}{I^2J^2T^2}\,\mathbb{E}_\alpha \Bigg(&\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{r=1}^{J} \bar{\boldsymbol{\Delta}}_{ijt}\bar{\boldsymbol{\Delta}}'_{irt} + \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i\neq p}^{I} \bar{\boldsymbol{\Delta}}_{ijt}\bar{\boldsymbol{\Delta}}'_{pjt} \\ &+ \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \boldsymbol{\Gamma}_{ijt}\boldsymbol{\Gamma}'_{ijt} \Bigg).\end{aligned}$$

B.3.2 Three-way fixed effects

With the inclusion of pair fixed effects, we introduce an additional bias of order T, since only T observations are informative per pair fixed effect. Another difference that occurs in contrast to the two-way fixed effects case is that predetermined regressors lead to a violation of the strict exogeneity assumption. To deal with this issue we adapt the asymptotic bias terms $\overline{\mathbf{B}}_\infty$ and $\overline{\mathbf{B}}^\delta_\infty$ of Fernández-Val and Weidner (2016) to the new structure.

Conjectured asymptotic distribution of $\hat{\boldsymbol{\beta}}$

$$\sqrt{IJT}(\widehat{\boldsymbol{\beta}}_{I,J,T} - \boldsymbol{\beta}^0) \to_d \overline{\mathbf{W}}_\infty^{-1}\mathcal{N}(\kappa_1\overline{\mathbf{B}}_{1,\infty} + \kappa_2\overline{\mathbf{B}}_{2,\infty} + \kappa_3\overline{\mathbf{B}}_{3,\infty}, \overline{\mathbf{W}}_\infty), \quad \text{with}$$

$$\overline{\mathbf{B}}_{1,\infty} = \text{plim}_{I,J,T\to\infty}\left[-\frac{1}{2JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\frac{\sum_{i=1}^{I}\mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}]}{\sum_{i=1}^{I}\mathbb{E}_\alpha[\omega_{ijt}]}\right],$$

$$\overline{\mathbf{B}}_{2,\infty} = \text{plim}_{I,J,T\to\infty}\left[-\frac{1}{2IT}\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{\sum_{j=1}^{J}\mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}]}{\sum_{j=1}^{J}\mathbb{E}_\alpha[\omega_{ijt}]}\right],$$

$$\overline{\mathbf{B}}_{3,\infty} = \text{plim}_{I,J,T\to\infty}\left[-\frac{1}{2IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{t=1}^{T}\mathbb{E}_\alpha[\omega_{ijt}]\right)^{-1}\left(\sum_{t=1}^{T}\mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}]\right.\right.$$

$$\left.\left.+2\sum_{\tau=t+1}^{T}\mathbb{E}_\alpha[H_{ijt}(Y_{ijt} - F_{ijt})\omega_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}]\right)\right],$$

$$\overline{\mathbf{W}}_\infty = \text{plim}_{I,J,T\to\infty}\left[\frac{1}{IJT}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\mathbb{E}_\alpha[\omega_{ijt}(\mathbb{M}\,\mathbf{X})_{ijt}(\mathbb{M}\,\mathbf{X})'_{ijt}]\right].$$

where $\sqrt{(JT)/I} \to \kappa_1$, $\sqrt{(IT)/J} \to \kappa_2$, and $\sqrt{(IJ)/T} \to \kappa_3$ as $I, J, T \to \infty$. The second term in the numerator of $\overline{\mathbf{B}}_{3,\infty}$ is dropped if all regressors are assumed to be strictly exogenous.

Conjectured asymptotic distribution of $\hat{\delta}$

$$r(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta} - I^{-1}\overline{\mathbf{B}}^{\delta}_{1,\infty} - J^{-1}\overline{\mathbf{B}}^{\delta}_{2,\infty} - T^{-1}\overline{\mathbf{B}}^{\delta}_{3,\infty}) \to_d \mathcal{N}(0, \overline{\mathbf{V}}^{\delta}_{\infty}), \quad \text{with}$$

$$\overline{\mathbf{B}}^{\delta}_{1,\infty} = \text{plim}_{I,J,T\to\infty}\left[\frac{1}{2JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\frac{\sum_{i=1}^{I} - \mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}]\,\mathbb{E}_\alpha[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}] + \mathbb{E}_\alpha[\partial_{\eta^2}\boldsymbol{\Delta}_{ijt}]}{\sum_{i=1}^{I}\mathbb{E}_\alpha[\omega_{ijt}]}\right],$$

$$\overline{\mathbf{B}}^{\delta}_{2,\infty} = \text{plim}_{I,J,T\to\infty}\left[\frac{1}{2IT}\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{\sum_{j=1}^{J} - \mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}]\,\mathbb{E}_\alpha[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}] + \mathbb{E}_\alpha[\partial_{\eta^2}\boldsymbol{\Delta}_{ijt}]}{\sum_{j=1}^{J}\mathbb{E}_\alpha[\omega_{ijt}]}\right],$$

$$\overline{\mathbf{B}}^{\delta}_{3,\infty} = \text{plim}_{I,J,T\to\infty}\left[\frac{1}{2IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{t=1}^{T}\mathbb{E}_\alpha[\omega_{ijt}]\right)^{-1}\left(\sum_{t=1}^{T} - \mathbb{E}_\alpha[H_{ijt}\partial_{\eta^2}F_{ijt}]\,\mathbb{E}_\alpha[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}]\right.\right.$$

$$\left.\left. + \mathbb{E}_\alpha[\partial_{\eta^2}\boldsymbol{\Delta}_{ijt}] + 2\sum_{\tau=t+1}^{T}\mathbb{E}_\alpha[\partial_\eta \ell_{ijt-l}\omega_{ijt}\,(\mathbb{M}\,\boldsymbol{\Psi})_{ijt}]\right)\right].$$

$$\overline{\mathbf{V}}^{\delta}_{\infty} = \text{plim}_{I,J,T\to\infty}\frac{r^2}{I^2J^2T^2}\,\mathbb{E}_\alpha\left[\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\bar{\boldsymbol{\Delta}}_{ijt}\right)\left(\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\bar{\boldsymbol{\Delta}}_{ijt}\right)'\right.$$

$$\left. + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\boldsymbol{\Gamma}_{ijt}\boldsymbol{\Gamma}'_{ijt} + 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{s>t}^{T}\bar{\boldsymbol{\Delta}}_{ijt}\boldsymbol{\Gamma}'_{ijs}\right],$$

$$\boldsymbol{\Gamma}_{ijt} = \mathbb{E}_\alpha\left[(IJT)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\partial_\beta\boldsymbol{\Delta}_{ijt} - (\mathbb{P}\,\mathbf{X})_{ijt}\,\partial_\eta\boldsymbol{\Delta}_{ijt}\right]'\overline{\mathbf{W}}^{-1}_{\infty}\,\mathbb{E}_\alpha\left[(\mathbb{M}\,\mathbf{X})_{ijt}\,\omega_{ijt}\boldsymbol{\nu}_{ijt}\right]$$

$$- \mathbb{E}_\alpha\left[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}\,\partial_\eta\ell_{ijt}\right],$$

and r is a convergence rate. The last term in the numerator of $\overline{\mathbf{B}}_{3,\infty}$ and $\overline{\mathbf{V}}^{\delta}_{\infty}$ are dropped if all regressors are assumed to be strictly exogenous. The expression $\overline{\mathbf{V}}^{\delta}_{\infty}$ can be further modified by assuming that $\{\lambda_{it}\}_{IT}$, $\{\psi_{jt}\}_{JT}$ and $\{\mu_{ij}\}_{IJ}$ are independent sequences, and λ_{it}, ψ_{jt} and μ_{ij} are independent for all it, jt, ij:

$$\widehat{\mathbf{V}}^{\delta} = \text{plim}_{I,J,T\to\infty}\frac{r^2}{I^2J^2T^2}\,\mathbb{E}_\alpha\left(\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{r=1}^{J}\bar{\boldsymbol{\Delta}}_{ijt}\bar{\boldsymbol{\Delta}}'_{irt} + \sum_{j=1}^{J}\sum_{t=1}^{T}\sum_{i\neq p}^{I}\bar{\boldsymbol{\Delta}}_{ijt}\bar{\boldsymbol{\Delta}}'_{pjt}\right.$$

$$\left. + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{s\neq t}^{T}\bar{\boldsymbol{\Delta}}_{ijt}\bar{\boldsymbol{\Delta}}'_{ijs} + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}\boldsymbol{\Gamma}_{ijt}\boldsymbol{\Gamma}'_{ijt} + 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{s>t}^{T}\bar{\boldsymbol{\Delta}}_{ijt}\boldsymbol{\Gamma}'_{ijs}\right),$$

B.4 Bias-corrected Ordinary Least Squares

Consider the three-way fixed effects linear probability model

$$y_{ijt} = \lambda_{it} + \psi_{jt} + \mu_{ij} + \mathbf{x}'_{ijt}\boldsymbol{\beta} + \epsilon_{ijt} \ ,$$

which can also be rewritten in matrix notation:

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \ . \tag{24}$$

We first deal with the computational burden. Applying the three-way fixed effects residual projection $\mathbb{M} = \mathbf{I}_{IJT} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ to (24), leads to the following concentrated regression:

$$\mathbb{M}\mathbf{y} = \mathbb{M}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \ . \tag{25}$$

The demeaning can be efficiently carried out by using the method of alternating projections (see Gaure, 2013).

Hahn and Moon (2006) have derived the bias of dynamic linear models with individual and time fixed effects. They show that there is only a bias of order $1/T$ stemming from the inclusion of individual effects in combination with predetermined regressors. Transferring their result to our problem with the three-way error component suggests that the inclusion of pair fixed effects in combination with predetermined regressors leads to the same order of the bias. Thus, the linear probability model needs only to be bias-corrected if not all regressors are strictly exogenous. This is, for example, the case in a dynamic model, where we include $\mathbf{y}_{t-1}$ to our set of regressors.
An estimator of the bias is given by

$$\widehat{\mathbf{B}} = \left(\frac{1}{IJT}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}(\mathbb{M}\mathbf{X})_{ijt}(\mathbb{M}\mathbf{X})'_{ijt}\right)^{-1}\left(-\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{l=1}^{L}\frac{1}{T-l}\sum_{t=l+1}^{T}\mathbf{X}_{ijt}\hat{\boldsymbol{\epsilon}}_{ijt-l}\right) \ ,$$

where $\hat{\boldsymbol{\epsilon}}$ is the residual of (25) and L is a bandwidth parameter.[28] This yields the bias-corrected estimator

$$\hat{\boldsymbol{\beta}} - \frac{\widehat{\mathbf{B}}}{IJT} \ , \tag{26}$$

[28] The residuals of equation (24) and equation (25) are identical (see Gaure, 2013).

where $\hat{\boldsymbol{\beta}} = \left((\mathbb{M}\mathbf{X})'(\mathbb{M}\mathbf{X})\right)^{-1}(\mathbb{M}\mathbf{X})'\mathbb{M}\mathbf{y}$.

C Monte Carlo Results — Dynamic Model

C.1 Two-way fixed effects

Table 8: Dynamic: Two-way FEs – x, $N = 50$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
	N = 50; T = 10									
MLE	5	2	5	0.95	0.14	0	1	1	0.97	0.95
ABC	-0	2	2	0.99	0.95	-0	1	1	0.98	0.95
SPJ	-1	2	2	0.96	0.90	-0	1	1	0.96	0.95
LPM						-0	1	1	0.89	0.91
	N = 50; T = 20									
MLE	5	1	5	0.97	0.00	0	1	1	0.97	0.95
ABC	-0	1	1	1.01	0.95	-0	1	1	0.98	0.95
SPJ	-1	1	1	0.97	0.88	-0	1	1	0.96	0.94
LPM						-0	1	1	0.88	0.92
	N = 50; T = 30									
MLE	5	1	5	0.93	0.00	0	1	1	0.97	0.94
ABC	-0	1	1	0.97	0.94	-0	1	1	0.98	0.95
SPJ	-1	1	1	0.93	0.86	-0	1	1	0.96	0.94
LPM						-0	1	1	0.90	0.92
	N = 50; T = 40									
MLE	5	1	5	0.98	0.00	0	1	1	1.00	0.96
ABC	-0	1	1	1.03	0.95	-0	1	1	1.01	0.96
SPJ	-1	1	1	0.98	0.83	-0	1	1	0.98	0.94
LPM						-0	1	1	0.92	0.92
	N = 50; T = 50									
MLE	5	1	5	0.92	0.00	0	1	1	0.95	0.93
ABC	-0	1	1	0.96	0.94	-0	1	1	0.95	0.94
SPJ	-1	1	1	0.94	0.80	-0	1	1	0.93	0.92
LPM						-0	1	1	0.86	0.90

Table 9: Dynamic: Two-way FEs – x, $N = 100$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
	N = 100; T = 10									
MLE	2	1	3	0.97	0.12	0	1	1	0.95	0.94
ABC	-0	1	1	0.99	0.94	-0	1	1	0.95	0.94
SPJ	-0	1	1	0.97	0.94	-0	1	1	0.94	0.93
LPM						-0	1	1	0.79	0.87
	N = 100; T = 20									
MLE	2	1	2	0.96	0.01	0	1	1	0.90	0.92
ABC	-0	1	1	0.98	0.94	-0	1	1	0.90	0.92
SPJ	-0	1	1	0.96	0.93	-0	1	1	0.89	0.91
LPM						-0	1	1	0.73	0.82
	N = 100; T = 30									
MLE	2	0	2	0.97	0.00	0	0	0	0.92	0.93
ABC	-0	0	0	0.99	0.95	-0	0	0	0.92	0.93
SPJ	-0	0	0	0.98	0.93	-0	0	0	0.91	0.92
LPM						-0	0	1	0.75	0.83
	N = 100; T = 40									
MLE	2	0	2	0.97	0.00	0	0	0	0.89	0.92
ABC	-0	0	0	0.99	0.95	-0	0	0	0.89	0.92
SPJ	-0	0	0	0.99	0.92	-0	0	0	0.88	0.92
LPM						-0	0	0	0.73	0.81
	N = 100; T = 50									
MLE	2	0	2	0.99	0.00	0	0	0	0.92	0.93
ABC	-0	0	0	1.00	0.95	-0	0	0	0.92	0.94
SPJ	-0	0	0	0.99	0.93	-0	0	0	0.91	0.93
LPM						-0	0	0	0.74	0.83

Table 10: Dynamic: Two-way FEs – x, $N = 150$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
	N = 150; T = 10									
MLE	2	1	2	0.98	0.12	-0	1	1	0.91	0.92
ABC	-0	1	1	0.99	0.95	-0	1	1	0.91	0.93
SPJ	-0	1	1	0.99	0.94	-0	1	1	0.91	0.93
LPM						-0	1	1	0.67	0.80
	N = 150; T = 20									
MLE	2	0	2	0.99	0.01	0	0	0	0.91	0.92
ABC	-0	0	0	1.00	0.95	-0	0	0	0.90	0.93
SPJ	-0	0	0	0.98	0.93	-0	0	0	0.90	0.92
LPM						-0	0	0	0.67	0.76
	N = 150; T = 30									
MLE	2	0	2	1.01	0.00	0	0	0	0.86	0.91
ABC	-0	0	0	1.02	0.95	-0	0	0	0.86	0.90
SPJ	-0	0	0	1.01	0.95	-0	0	0	0.86	0.91
LPM						-0	0	0	0.63	0.73
	N = 150; T = 40									
MLE	2	0	2	0.99	0.00	0	0	0	0.88	0.91
ABC	0	0	0	1.00	0.95	0	0	0	0.88	0.91
SPJ	-0	0	0	0.98	0.94	0	0	0	0.88	0.91
LPM						-0	0	0	0.66	0.75
	N = 150; T = 50									
MLE	2	0	2	1.02	0.00	0	0	0	0.90	0.93
ABC	-0	0	0	1.03	0.96	-0	0	0	0.90	0.93
SPJ	-0	0	0	1.02	0.95	-0	0	0	0.90	0.93
LPM						-0	0	0	0.67	0.73

Table 11: Dynamic: Two-way FEs – y_{t-1}, $N = 50$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
	N = 50; T = 10									
MLE	5	4	7	0.99	0.81	0	4	4	0.99	0.95
ABC	-0	4	4	1.03	0.95	-0	4	4	1.01	0.95
SPJ	-1	4	4	1.00	0.94	-0	4	4	0.98	0.94
LPM						5	4	7	0.97	0.76
	N = 50; T = 20									
MLE	5	3	6	0.96	0.65	-0	3	3	0.96	0.94
ABC	-0	3	3	1.00	0.95	-0	3	3	0.97	0.95
SPJ	-1	3	3	0.97	0.93	-0	3	3	0.94	0.94
LPM						5	3	6	0.96	0.56
	N = 50; T = 30									
MLE	5	3	6	0.95	0.48	0	3	3	0.94	0.92
ABC	0	3	3	0.99	0.95	0	3	3	0.96	0.93
SPJ	-1	3	3	0.97	0.93	0	3	3	0.94	0.93
LPM						6	3	6	0.94	0.40
	N = 50; T = 40									
MLE	5	2	5	0.98	0.38	0	2	2	0.99	0.95
ABC	-0	2	2	1.02	0.95	-0	2	2	1.01	0.95
SPJ	-1	2	2	1.01	0.94	-0	2	2	0.99	0.95
LPM						6	2	6	0.97	0.27
	N = 50; T = 50									
MLE	5	2	5	0.92	0.31	0	2	2	0.93	0.93
ABC	-0	2	2	0.96	0.94	-0	2	2	0.95	0.93
SPJ	-1	2	2	0.94	0.92	-0	2	2	0.92	0.93
LPM						6	2	6	0.93	0.21

Table 12: Dynamic: Two-way FEs – y_{t-1}, $N = 100$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
	N = 100; T = 10									
MLE	2	2	3	0.96	0.80	0	2	2	0.94	0.94
ABC	0	2	2	0.98	0.94	0	2	2	0.95	0.94
SPJ	-0	2	2	0.97	0.94	0	2	2	0.95	0.94
LPM						5	2	6	0.91	0.30
	N = 100; T = 20									
MLE	2	2	3	0.99	0.63	0	2	2	0.99	0.94
ABC	-0	1	1	1.01	0.95	-0	2	2	1.00	0.94
SPJ	-0	2	2	0.99	0.94	-0	2	2	0.98	0.94
LPM						6	2	6	0.96	0.06
	N = 100; T = 30									
MLE	2	1	3	0.97	0.52	0	1	1	0.97	0.94
ABC	-0	1	1	0.99	0.94	-0	1	1	0.98	0.94
SPJ	-0	1	1	0.96	0.94	-0	1	1	0.96	0.93
LPM						6	1	6	0.94	0.01
	N = 100; T = 40									
MLE	2	1	3	0.99	0.42	0	1	1	0.97	0.94
ABC	-0	1	1	1.01	0.95	-0	1	1	0.98	0.94
SPJ	-0	1	1	0.99	0.94	-0	1	1	0.96	0.94
LPM						6	1	6	0.94	0.00
	N = 100; T = 50									
MLE	2	1	3	0.94	0.31	0	1	1	0.92	0.93
ABC	-0	1	1	0.96	0.93	-0	1	1	0.92	0.93
SPJ	-0	1	1	0.95	0.93	-0	1	1	0.91	0.92
LPM						6	1	6	0.90	0.00

Table 13: Dynamic: Two-way FEs – y_{t-1}, $N = 150$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
	N = 150; T = 10									
MLE	2	1	2	0.98	0.79	0	2	2	0.96	0.94
ABC	0	1	1	0.99	0.95	0	2	2	0.97	0.94
SPJ	-0	1	1	0.98	0.95	0	2	2	0.95	0.94
LPM						6	2	6	0.92	0.04
	N = 150; T = 20									
MLE	2	1	2	0.98	0.66	-0	1	1	1.00	0.95
ABC	-0	1	1	1.00	0.95	-0	1	1	1.00	0.95
SPJ	-0	1	1	0.99	0.95	-0	1	1	0.99	0.95
LPM						5	1	6	0.96	0.00
	N = 150; T = 30									
MLE	2	1	2	0.98	0.53	0	1	1	0.99	0.95
ABC	0	1	1	1.00	0.95	0	1	1	0.99	0.95
SPJ	-0	1	1	0.98	0.95	0	1	1	0.98	0.95
LPM						6	1	6	0.94	0.00
	N = 150; T = 40									
MLE	2	1	2	0.96	0.42	-0	1	1	0.96	0.94
ABC	-0	1	1	0.97	0.94	-0	1	1	0.96	0.94
SPJ	-0	1	1	0.96	0.94	-0	1	1	0.95	0.94
LPM						6	1	6	0.91	0.00
	N = 150; T = 50									
MLE	2	1	2	0.94	0.34	-0	1	1	0.93	0.93
ABC	-0	1	1	0.95	0.94	-0	1	1	0.94	0.94
SPJ	-0	1	1	0.94	0.94	-0	1	1	0.93	0.94
LPM						6	1	6	0.90	0.00

C.2 Three-way fixed effects

Table 14: Dynamic: Three-way FEs – x, $N = 50$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 50; T = 10					
MLE	29	3	29	0.82	0.00	4	2	4	1.01	0.33
ABC (1)	-1	2	2	1.02	0.94	-1	2	2	1.09	0.94
ABC (2)	-1	2	2	1.01	0.93	-1	2	2	1.08	0.93
SPJ	-14	3	14	0.62	0.00	4	2	5	0.87	0.32
LPM (1)						0	2	2	0.94	0.93
LPM (2)						-0	2	2	0.94	0.93
					N = 50; T = 20					
MLE	16	1	16	0.87	0.00	3	1	3	0.97	0.36
ABC (1)	-0	1	1	0.98	0.94	-0	1	1	1.00	0.95
ABC (2)	-0	1	1	0.97	0.93	-0	1	1	0.99	0.95
SPJ	-5	1	5	0.86	0.04	1	1	1	0.91	0.89
LPM (1)						-0	1	1	0.90	0.93
LPM (2)						-0	1	1	0.90	0.92
					N = 50; T = 30					
MLE	12	1	12	0.92	0.00	2	1	2	1.00	0.48
ABC (1)	-0	1	1	1.01	0.95	-0	1	1	1.01	0.95
ABC (2)	-0	1	1	1.01	0.95	-0	1	1	1.01	0.94
SPJ	-3	1	3	0.93	0.15	-0	1	1	0.96	0.95
LPM (1)						-0	1	1	0.89	0.92
LPM (2)						-0	1	1	0.89	0.90
					N = 50; T = 40					
MLE	10	1	10	0.89	0.00	1	1	2	0.97	0.53
ABC (1)	-0	1	1	0.97	0.94	-0	1	1	0.98	0.93
ABC (2)	-0	1	1	0.97	0.94	-0	1	1	0.97	0.93
SPJ	-2	1	2	0.88	0.27	-0	1	1	0.91	0.91
LPM (1)						-0	1	1	0.84	0.89
LPM (2)						-0	1	1	0.84	0.86
					N = 50; T = 50					
MLE	9	1	9	0.90	0.00	1	1	1	1.01	0.61
ABC (1)	-0	1	1	0.97	0.94	-0	1	1	1.01	0.96
ABC (2)	-0	1	1	0.97	0.93	-0	1	1	1.01	0.96
SPJ	-2	1	2	0.90	0.33	-0	1	1	0.94	0.94
LPM (1)						-0	1	1	0.86	0.88
LPM (2)						-0	1	1	0.86	0.87

Table 15: Dynamic: Three-way FEs – x, $N = 100$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 100; T = 10					
MLE	24	1	24	0.89	0.00	4	1	4	1.04	0.02
ABC (1)	0	1	1	1.05	0.95	-0	1	1	1.08	0.94
ABC (2)	0	1	1	1.05	0.96	-1	1	1	1.08	0.91
SPJ	-9	1	9	0.70	0.00	6	1	6	0.89	0.00
LPM (1)						0	1	1	0.88	0.91
LPM (2)						-0	1	1	0.87	0.91
					N = 100; T = 20					
MLE	13	1	13	0.89	0.00	2	1	2	0.96	0.02
ABC (1)	0	1	1	0.98	0.93	0	1	1	0.98	0.95
ABC (2)	0	1	1	0.98	0.94	-0	1	1	0.97	0.94
SPJ	-3	1	3	0.86	0.01	1	1	1	0.89	0.54
LPM (1)						-0	1	1	0.85	0.89
LPM (2)						-0	1	1	0.85	0.87
					N = 100; T = 30					
MLE	9	1	9	0.91	0.00	2	0	2	0.96	0.05
ABC (1)	0	0	0	0.97	0.95	0	0	0	0.96	0.94
ABC (2)	-0	0	0	0.97	0.94	-0	0	0	0.96	0.94
SPJ	-1	1	2	0.91	0.14	0	0	1	0.93	0.86
LPM (1)						-0	0	1	0.82	0.86
LPM (2)						-0	0	1	0.82	0.81
					N = 100; T = 40					
MLE	7	0	7	0.91	0.00	1	0	1	0.94	0.12
ABC (1)	0	0	0	0.96	0.94	0	0	0	0.94	0.93
ABC (2)	-0	0	0	0.96	0.94	-0	0	0	0.94	0.92
SPJ	-1	0	1	0.92	0.32	0	0	0	0.91	0.91
LPM (1)						-0	0	1	0.79	0.81
LPM (2)						-0	0	1	0.79	0.73
					N = 100; T = 50					
MLE	6	0	6	0.94	0.00	1	0	1	1.00	0.17
ABC (1)	0	0	0	0.99	0.94	0	0	0	1.00	0.95
ABC (2)	-0	0	0	0.98	0.94	-0	0	0	1.00	0.95
SPJ	-1	0	1	0.95	0.48	0	0	0	0.96	0.94
LPM (1)						-0	0	0	0.80	0.76
LPM (2)						-0	0	1	0.80	0.69

Table 16: Dynamic: Three-way FEs – x, $N = 150$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 150; T = 10					
MLE	23	1	23	0.86	0.00	3	1	4	1.06	0.00
ABC (1)	1	1	1	1.01	0.82	-0	1	1	1.09	0.94
ABC (2)	0	1	1	1.00	0.88	-0	1	1	1.07	0.90
SPJ	-7	1	7	0.67	0.00	6	1	6	0.89	0.00
LPM (1)						0	1	1	0.84	0.88
LPM (2)						-0	1	1	0.83	0.90
					N = 150; T = 20					
MLE	11	0	11	0.94	0.00	2	0	2	0.97	0.00
ABC (1)	0	0	0	1.02	0.89	0	0	0	0.97	0.94
ABC (2)	0	0	0	1.01	0.93	-0	0	0	0.97	0.94
SPJ	-2	0	2	0.89	0.00	1	0	1	0.90	0.16
LPM (1)						-0	0	0	0.81	0.88
LPM (2)						-0	0	0	0.81	0.81
					N = 150; T = 30					
MLE	8	0	8	0.92	0.00	2	0	2	0.96	0.00
ABC (1)	0	0	0	0.98	0.91	0	0	0	0.97	0.93
ABC (2)	0	0	0	0.98	0.95	-0	0	0	0.97	0.95
SPJ	-1	0	1	0.91	0.06	0	0	1	0.92	0.73
LPM (1)						-0	0	0	0.79	0.80
LPM (2)						-0	0	0	0.79	0.66
					N = 150; T = 40					
MLE	6	0	6	0.95	0.00	1	0	1	0.95	0.01
ABC (1)	0	0	0	1.00	0.94	0	0	0	0.95	0.92
ABC (2)	-0	0	0	1.00	0.95	-0	0	0	0.95	0.94
SPJ	-1	0	1	0.94	0.22	0	0	0	0.92	0.87
LPM (1)						-0	0	0	0.75	0.68
LPM (2)						-0	0	0	0.75	0.54
					N = 150; T = 50					
MLE	5	0	5	0.95	0.00	1	0	1	0.97	0.02
ABC (1)	0	0	0	0.99	0.93	0	0	0	0.97	0.93
ABC (2)	-0	0	0	0.99	0.94	-0	0	0	0.97	0.94
SPJ	-1	0	1	0.95	0.38	0	0	0	0.95	0.91
LPM (1)						-0	0	0	0.76	0.61
LPM (2)						-0	0	0	0.76	0.45

Table 17: Dynamic: Three-way FEs – y_{t-1}, $N = 50$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 50; T = 10					
MLE	-62	5	62	0.95	0.00	-70	4	71	1.02	0.00
ABC (1)	-6	4	7	1.14	0.81	-7	5	8	1.11	0.76
ABC (2)	-7	5	9	1.05	0.68	-8	5	10	1.02	0.62
SPJ	24	6	25	0.77	0.01	-11	6	12	0.94	0.48
LPM (1)						2	5	5	1.02	0.95
LPM (2)						3	5	6	0.94	0.89
					N = 50; T = 20					
MLE	-27	4	27	0.94	0.00	-36	3	37	0.95	0.00
ABC (1)	-3	3	4	1.05	0.87	-3	3	5	1.00	0.85
ABC (2)	-1	3	3	1.00	0.94	-1	3	4	0.96	0.93
SPJ	5	4	6	0.89	0.69	-2	4	4	0.89	0.89
LPM (1)						8	3	9	0.95	0.28
LPM (2)						11	4	12	0.91	0.09
					N = 50; T = 30					
MLE	-16	3	16	0.97	0.00	-25	3	25	0.97	0.00
ABC (1)	-2	3	3	1.06	0.88	-2	3	3	1.01	0.87
ABC (2)	-0	3	3	1.03	0.95	-0	3	3	0.98	0.95
SPJ	2	3	3	0.95	0.88	-1	3	3	0.92	0.93
LPM (1)						10	3	11	0.96	0.03
LPM (2)						13	3	13	0.94	0.00
					N = 50; T = 40					
MLE	-11	2	11	0.96	0.01	-19	2	19	0.95	0.00
ABC (1)	-2	2	3	1.03	0.86	-2	2	3	0.99	0.85
ABC (2)	-0	2	2	1.01	0.95	-0	2	2	0.97	0.95
SPJ	1	2	3	0.93	0.92	-0	3	3	0.90	0.92
LPM (1)						11	2	12	0.95	0.01
LPM (2)						13	2	13	0.93	0.00
					N = 50; T = 50					
MLE	-7	2	8	0.94	0.07	-15	2	15	0.92	0.00
ABC (1)	-2	2	3	1.01	0.89	-2	2	3	0.95	0.87
ABC (2)	-0	2	2	0.99	0.95	-0	2	2	0.93	0.93
SPJ	0	2	2	0.92	0.92	-0	2	2	0.87	0.90
LPM (1)						12	2	12	0.92	0.00
LPM (2)						14	2	14	0.91	0.00

Table 18: Dynamic: Three-way FEs – y_{t-1}, $N = 100$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 100; T = 10					
MLE	-63	3	63	0.98	0.00	-70	2	70	1.04	0.00
ABC (1)	-6	2	7	1.13	0.22	-8	2	8	1.10	0.09
ABC (2)	-8	2	8	1.04	0.08	-9	2	10	1.01	0.03
SPJ	21	3	21	0.80	0.00	-11	3	11	0.94	0.02
LPM (1)						2	2	3	1.00	0.84
LPM (2)						4	3	4	0.92	0.66
					N = 100; T = 20					
MLE	-29	2	29	0.96	0.00	-37	2	37	0.96	0.00
ABC (1)	-3	2	4	1.03	0.42	-4	2	4	0.99	0.37
ABC (2)	-1	2	2	0.99	0.86	-2	2	2	0.95	0.83
SPJ	4	2	5	0.91	0.26	-2	2	3	0.90	0.80
LPM (1)						8	2	9	0.95	0.00
LPM (2)						11	2	11	0.91	0.00
					N = 100; T = 30					
MLE	-18	1	18	0.97	0.00	-25	1	25	0.96	0.00
ABC (1)	-3	1	3	1.03	0.50	-3	1	3	0.98	0.49
ABC (2)	-1	1	1	1.00	0.93	-1	1	2	0.95	0.92
SPJ	2	1	2	0.94	0.72	-1	1	2	0.90	0.90
LPM (1)						10	1	10	0.95	0.00
LPM (2)						13	1	13	0.92	0.00
					N = 100; T = 40					
MLE	-13	1	13	1.01	0.00	-19	1	19	1.01	0.00
ABC (1)	-2	1	2	1.06	0.57	-2	1	2	1.04	0.56
ABC (2)	-0	1	1	1.04	0.94	-0	1	1	1.02	0.94
SPJ	1	1	2	0.98	0.86	-0	1	1	0.96	0.93
LPM (1)						11	1	11	0.98	0.00
LPM (2)						13	1	13	0.96	0.00
					N = 100; T = 50					
MLE	-10	1	10	0.98	0.00	-15	1	15	0.97	0.00
ABC (1)	-2	1	2	1.03	0.61	-2	1	2	0.99	0.62
ABC (2)	-0	1	1	1.01	0.95	-0	1	1	0.98	0.94
SPJ	1	1	1	0.98	0.91	-0	1	1	0.95	0.93
LPM (1)						12	1	12	0.94	0.00
LPM (2)						14	1	14	0.93	0.00

Table 19: Dynamic: Three-way FEs – y_{t-1}, $N = 150$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 150; T = 10					
MLE	-63	2	64	0.95	0.00	-70	1	70	1.02	0.00
ABC (1)	-7	1	7	1.09	0.01	-8	2	9	1.08	0.00
ABC (2)	-8	2	9	1.01	0.00	-10	2	10	1.00	0.00
SPJ	20	2	20	0.78	0.00	-11	2	11	0.92	0.00
LPM (1)						2	2	3	0.98	0.71
LPM (2)						3	2	4	0.90	0.42
					N = 150; T = 20					
MLE	-30	1	30	0.99	0.00	-37	1	37	1.00	0.00
ABC (1)	-4	1	4	1.07	0.05	-4	1	4	1.03	0.03
ABC (2)	-2	1	2	1.02	0.69	-2	1	2	0.99	0.61
SPJ	4	1	4	0.92	0.05	-2	1	2	0.90	0.61
LPM (1)						8	1	8	0.96	0.00
LPM (2)						11	1	11	0.92	0.00
					N = 150; T = 30					
MLE	-19	1	19	0.98	0.00	-25	1	25	0.97	0.00
ABC (1)	-3	1	3	1.04	0.15	-3	1	3	0.99	0.13
ABC (2)	-1	1	1	1.01	0.89	-1	1	1	0.97	0.87
SPJ	2	1	2	0.96	0.47	-0	1	1	0.92	0.90
LPM (1)						10	1	10	0.93	0.00
LPM (2)						13	1	13	0.91	0.00
					N = 150; T = 40					
MLE	-14	1	14	1.01	0.00	-19	1	19	0.99	0.00
ABC (1)	-2	1	2	1.06	0.20	-2	1	2	1.01	0.19
ABC (2)	-0	1	1	1.03	0.92	-0	1	1	0.99	0.90
SPJ	1	1	1	0.96	0.76	-0	1	1	0.93	0.92
LPM (1)						11	1	11	0.96	0.00
LPM (2)						13	1	13	0.94	0.00
					N = 150; T = 50					
MLE	-11	1	11	0.97	0.00	-15	1	15	0.95	0.00
ABC (1)	-2	1	2	1.01	0.30	-2	1	2	0.97	0.30
ABC (2)	-0	1	1	0.99	0.92	-0	1	1	0.95	0.91
SPJ	1	1	1	0.96	0.84	-0	1	1	0.92	0.92
LPM (1)						12	1	12	0.92	0.00
LPM (2)						14	1	14	0.90	0.00

D Further Monte Carlo Results — Static Model

Although the main focus of our article is on the dynamic two- and three-way fixed effects model, the static counterparts are also highly relevant for applied work. For this reason, we study the finite sample properties of MLE, ABC, SPJ and LPM for these model specifications, too. In the following we briefly sketch the designs. Let $i = 1, \ldots, N$, $j = 1, \ldots, N$, $t = 1, \ldots, T$, $\beta_y = 0.5$, $\beta = 1$.

Design - Two-way fixed effects

$$y_{ijt} = \mathbf{1}[\beta x_{ijt} + \lambda_{it} + \psi_{jt} \geq \epsilon_{ijt}] \ ,$$

where $\lambda_{it} \sim \text{iid.}\, \mathcal{N}(0, 1/16)$, $\psi_{jt} \sim \text{iid.}\, \mathcal{N}(0, 1/16)$, and $\epsilon_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 1)$. Further, $x_{ijt} = 0.5 x_{ijt-1} + \lambda_{it} + \psi_{jt} + \nu_{ijt}$, where $\nu_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 0.5)$, $x_{ij0} \sim \text{iid.}\, \mathcal{N}(0, 1)$.

Design - Three-way fixed effects

$$y_{ijt} = \mathbf{1}[\beta x_{ijt} + \lambda_{it} + \psi_{jt} + \mu_{ij} \geq \epsilon_{ijt}] \ ,$$

where $\lambda_{it} \sim \text{iid.}\, \mathcal{N}(0, 1/24)$, $\psi_{jt} \sim \text{iid.}\, \mathcal{N}(0, 1/24)$, $\mu_{ij} \sim \text{iid.}\, \mathcal{N}(0, 1/24)$, and $\epsilon_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 1)$. Further, $x_{ijt} = 0.5 x_{ijt-1} + \lambda_{it} + \psi_{jt} + \mu_{ij} + \nu_{ijt}$, where $\nu_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 0.5)$, $x_{ij0} \sim \text{iid.}\, \mathcal{N}(0, 1)$.

Note that, unlike in the dynamic three-way fixed effects model, the OLS estimator of the linear probability model (LPM) does not require a bias correction for the specifications considered in this section.

We now review the key results of the simulation experiments.

Results - Two-way fixed effects

Static (see Tables 20, 21, 22): although MLE shows a distortion in the structural parameter estimates, the bias does not carry over to the estimates of APEs. The bias corrections

ABC and SPJ work well. They reduce the biases of the structural parameters and APEs to 1 or zero percent, and bring the CPs close to the nominal level. Overall, ABC, SPJ and MLE work similarly well if APEs are of interest. In terms of structural parameters, ABC exhibits a lower bias and better CPs than SPJ in samples with smaller N. LPM shows no distortion of the APEs in all settings, but we observe that with increasing N, the standard errors are underestimated, resulting in too low CPs.

Note that MLE is consistent under fixed T asymptotics. This is also evident from the simulation results, where the properties of the estimator do not change with T.

Results - Three-way fixed effects

Static (see Tables 23, 24, 25): we find a considerable distortion in the MLE estimates of the structural parameters, which decreases with rising T, but is not negligibly small even at $T = 50$. ABC and SPJ both reduce this bias considerably, but ABC works better in samples with smaller T. While the CPs of ABC quickly converge to the nominal level, the CPs of SPJ are still far away from 95 percent even at $T = 50$. If we look at the APEs, we see that all estimators have either a very small bias of 1 percent or none at all. With increasing T, their CPs are also getting closer to 95 percent.

Table 20: Static: Two-way FEs – x, $N = 50$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 50; T = 10					
MLE	5	2	5	0.97	0.10	0	1	1	0.98	0.94
ABC	-0	1	1	1.01	0.94	-0	1	1	0.99	0.94
SPJ	-1	1	2	0.98	0.93	-0	1	1	0.96	0.93
LPM						0	1	1	0.96	0.93
					N = 50; T = 20					
MLE	5	1	5	0.99	0.01	0	1	1	1.06	0.96
ABC	-0	1	1	1.03	0.96	-0	1	1	1.07	0.97
SPJ	-1	1	1	0.98	0.91	-0	1	1	1.04	0.95
LPM						-0	1	1	1.05	0.96
					N = 50; T = 30					
MLE	5	1	5	0.98	0.00	0	1	1	1.01	0.95
ABC	-0	1	1	1.02	0.95	-0	1	1	1.03	0.95
SPJ	-1	1	1	1.00	0.89	-0	1	1	1.00	0.95
LPM						0	1	1	0.99	0.94
					N = 50; T = 40					
MLE	5	1	5	0.94	0.00	0	1	1	0.98	0.95
ABC	-0	1	1	0.97	0.94	-0	1	1	0.99	0.95
SPJ	-1	1	1	0.95	0.84	-0	1	1	0.97	0.94
LPM						-0	1	1	0.97	0.94
					N = 50; T = 50					
MLE	5	1	5	0.97	0.00	0	1	1	1.02	0.95
ABC	-0	1	1	1.01	0.96	-0	1	1	1.04	0.96
SPJ	-1	1	1	0.98	0.83	-0	1	1	1.00	0.95
LPM						0	1	1	1.00	0.95

Table 21: Static: Two-way FEs – x, $N = 100$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 100; T = 10					
MLE	2	1	2	0.95	0.13	0	1	1	0.96	0.94
ABC	-0	1	1	0.97	0.94	-0	1	1	0.96	0.93
SPJ	-0	1	1	0.95	0.93	-0	1	1	0.95	0.93
LPM						0	1	1	0.85	0.90
					N = 100; T = 20					
MLE	2	1	2	0.98	0.00	0	0	0	0.99	0.96
ABC	-0	1	1	1.00	0.95	-0	0	0	1.00	0.95
SPJ	-0	1	1	0.99	0.94	-0	0	0	0.99	0.95
LPM						-0	0	0	0.89	0.92
					N = 100; T = 30					
MLE	2	0	2	1.00	0.00	0	0	0	1.03	0.95
ABC	0	0	0	1.02	0.96	-0	0	0	1.03	0.95
SPJ	-0	0	0	1.00	0.95	-0	0	0	1.03	0.96
LPM						0	0	0	0.92	0.93
					N = 100; T = 40					
MLE	2	0	2	0.98	0.00	0	0	0	0.97	0.94
ABC	-0	0	0	1.00	0.94	-0	0	0	0.97	0.94
SPJ	-0	0	0	0.98	0.93	-0	0	0	0.96	0.94
LPM						-0	0	0	0.87	0.91
					N = 100; T = 50					
MLE	2	0	2	1.00	0.00	0	0	0	0.99	0.95
ABC	-0	0	0	1.02	0.96	-0	0	0	0.99	0.95
SPJ	-0	0	0	1.02	0.94	-0	0	0	0.99	0.95
LPM						-0	0	0	0.88	0.92

Table 22: Static: Two-way FEs – x, $N = 150$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 150; T = 10					
MLE	1	0	2	0.99	0.12	0	0	0	1.02	0.96
ABC	-0	0	0	1.01	0.96	-0	0	0	1.02	0.96
SPJ	-0	0	0	1.00	0.95	-0	0	0	1.01	0.95
LPM						-0	0	0	0.84	0.90
					N = 150; T = 20					
MLE	1	0	2	0.95	0.01	0	0	0	0.95	0.94
ABC	0	0	0	0.96	0.94	0	0	0	0.95	0.94
SPJ	-0	0	0	0.95	0.93	0	0	0	0.95	0.94
LPM						0	0	0	0.79	0.86
					N = 150; T = 30					
MLE	1	0	2	1.01	0.00	0	0	0	0.96	0.95
ABC	-0	0	0	1.03	0.95	-0	0	0	0.97	0.94
SPJ	-0	0	0	1.02	0.94	-0	0	0	0.96	0.95
LPM						-0	0	0	0.79	0.88
					N = 150; T = 40					
MLE	1	0	2	0.99	0.00	0	0	0	0.97	0.94
ABC	-0	0	0	1.00	0.95	-0	0	0	0.97	0.94
SPJ	-0	0	0	0.99	0.94	-0	0	0	0.96	0.94
LPM						-0	0	0	0.80	0.88
					N = 150; T = 50					
MLE	1	0	2	0.99	0.00	0	0	0	0.95	0.94
ABC	-0	0	0	1.00	0.94	-0	0	0	0.95	0.94
SPJ	-0	0	0	0.99	0.94	-0	0	0	0.95	0.94
LPM						0	0	0	0.78	0.88

Table 23: Static: Three-way FEs – x, $N = 50$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 50; T = 10					
MLE	22	2	22	0.85	0.00	1	1	2	1.00	0.89
ABC	-1	2	2	1.03	0.88	-1	1	2	1.07	0.86
SPJ	-12	2	12	0.72	0.00	-0	2	2	0.88	0.91
LPM						0	1	1	1.04	0.96
					N = 50; T = 20					
MLE	12	1	12	0.92	0.00	0	1	1	1.00	0.94
ABC	-1	1	1	1.02	0.92	-0	1	1	1.03	0.93
SPJ	-4	1	4	0.88	0.08	-1	1	1	0.92	0.89
LPM						-0	1	1	1.04	0.96
					N = 50; T = 30					
MLE	10	1	10	0.94	0.00	0	1	1	1.02	0.94
ABC	-0	1	1	1.02	0.94	-0	1	1	1.04	0.94
SPJ	-2	1	2	0.93	0.28	-0	1	1	0.96	0.89
LPM						0	1	1	1.01	0.95
					N = 50; T = 40					
MLE	8	1	8	0.93	0.00	0	1	1	1.02	0.95
ABC	-0	1	1	0.99	0.92	-0	1	1	1.03	0.94
SPJ	-2	1	2	0.94	0.40	-0	1	1	0.99	0.90
LPM						-0	1	1	0.98	0.94
					N = 50; T = 50					
MLE	8	1	8	0.96	0.00	0	1	1	1.04	0.94
ABC	-0	1	1	1.03	0.93	-0	1	1	1.06	0.95
SPJ	-1	1	2	0.95	0.46	-0	1	1	0.99	0.91
LPM						0	1	1	0.99	0.94

Table 24: Static: Three-way FEs – x, $N = 100$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 100; T = 10					
MLE	18	1	18	0.89	0.00	1	1	1	1.05	0.90
ABC	-1	1	1	1.05	0.80	-1	1	1	1.09	0.70
SPJ	-8	1	8	0.74	0.00	0	1	1	0.89	0.87
LPM						-0	1	1	1.04	0.96
					N = 100; T = 20					
MLE	9	1	9	0.93	0.00	0	0	1	1.01	0.92
ABC	-0	1	1	1.00	0.92	-0	0	1	1.03	0.92
SPJ	-2	1	2	0.94	0.01	-0	1	1	0.96	0.91
LPM						0	0	0	0.96	0.93
					N = 100; T = 30					
MLE	7	0	7	0.95	0.00	0	0	0	1.05	0.95
ABC	-0	0	0	1.01	0.93	-0	0	0	1.06	0.95
SPJ	-1	0	1	0.93	0.21	-0	0	0	0.98	0.92
LPM						-0	0	0	0.97	0.95
					N = 100; T = 40					
MLE	6	0	6	0.96	0.00	0	0	0	1.00	0.94
ABC	-0	0	0	1.00	0.94	-0	0	0	1.01	0.94
SPJ	-1	0	1	0.95	0.44	-0	0	0	0.95	0.92
LPM						-0	0	0	0.93	0.93
					N = 100; T = 50					
MLE	5	0	5	0.94	0.00	0	0	0	0.99	0.94
ABC	-0	0	0	0.98	0.94	-0	0	0	1.00	0.94
SPJ	-1	0	1	0.94	0.57	-0	0	0	0.97	0.92
LPM						-0	0	0	0.91	0.93

Table 25: Static: Three-way FEs – x, $N = 150$

	Coefficients					APE				
	Bias	SD	RMSE	SE/SD	CP .95	Bias	SD	RMSE	SE/SD	CP .95
					N = 150; T = 10					
MLE	16	1	16	0.87	0.00	0	0	1	1.04	0.87
ABC	-1	1	1	1.02	0.77	-1	0	1	1.07	0.51
SPJ	-7	1	7	0.76	0.00	1	1	1	0.91	0.73
LPM						-0	0	0	0.95	0.94
					N = 150; T = 20					
MLE	8	0	8	0.92	0.00	0	0	0	1.00	0.91
ABC	-0	0	0	0.99	0.91	-0	0	0	1.01	0.89
SPJ	-2	0	2	0.89	0.00	-0	0	0	0.93	0.91
LPM						-0	0	0	0.93	0.93
					N = 150; T = 30					
MLE	6	0	6	0.93	0.00	0	0	0	0.97	0.93
ABC	-0	0	0	0.97	0.93	-0	0	0	0.98	0.92
SPJ	-1	0	1	0.93	0.08	-0	0	0	0.92	0.90
LPM						-0	0	0	0.88	0.92
					N = 150; T = 40					
MLE	5	0	5	0.95	0.00	0	0	0	1.01	0.93
ABC	-0	0	0	0.99	0.94	-0	0	0	1.02	0.94
SPJ	-1	0	1	0.93	0.33	-0	0	0	0.98	0.93
LPM						-0	0	0	0.90	0.93
					N = 150; T = 50					
MLE	4	0	4	0.98	0.00	0	0	0	1.05	0.95
ABC	-0	0	0	1.01	0.94	-0	0	0	1.05	0.95
SPJ	-0	0	0	0.97	0.51	-0	0	0	1.00	0.94
LPM						-0	0	0	0.92	0.92

E Application

Figure 8: Fitted Probabilities

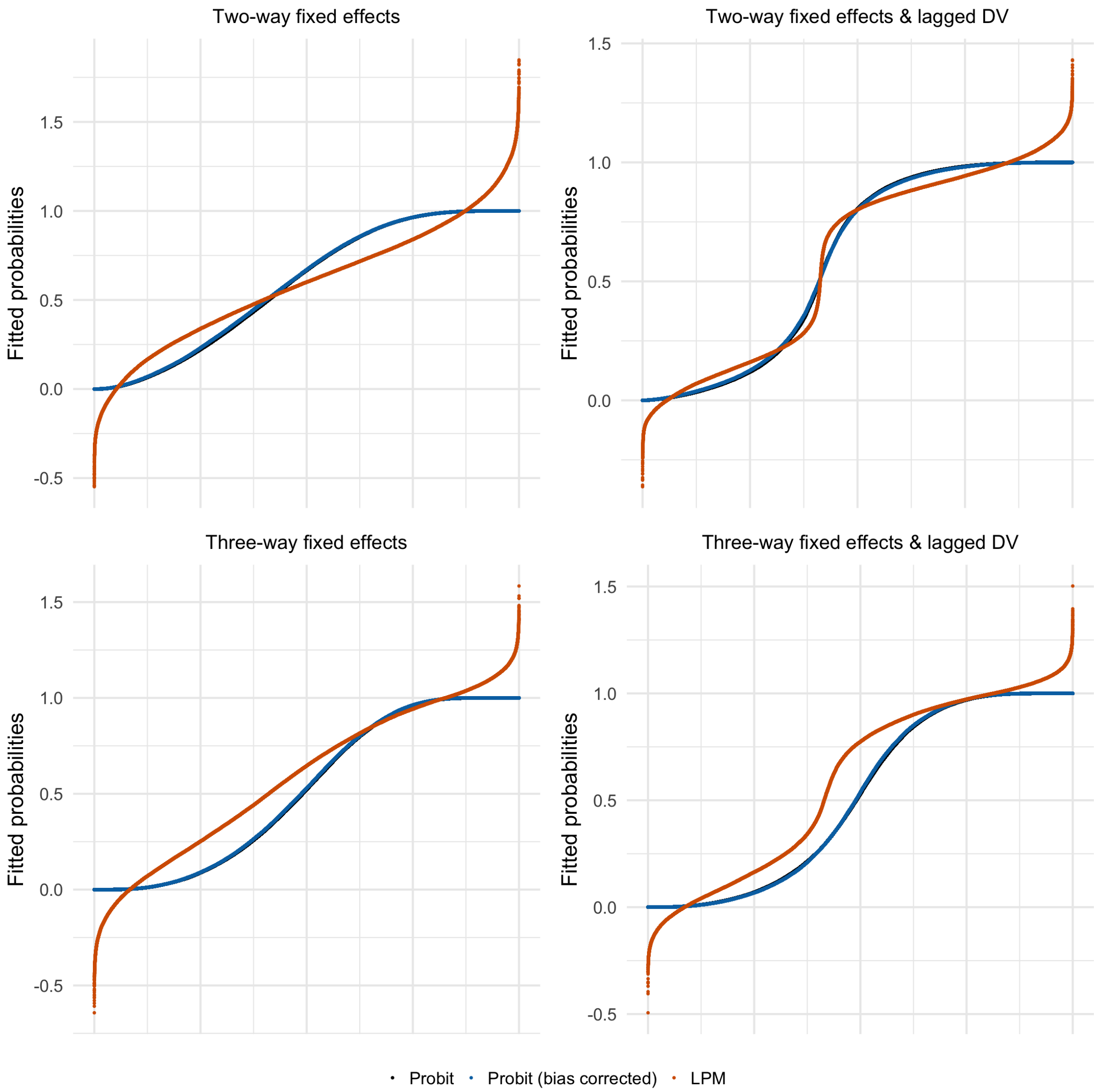


Table 26: Logit Estimation: Coefficients

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	-	-	2.869***	-	1.929***
	[-]	[-]	[2.985]	[-]	[1.798]
	(-)	(-)	(0.008)	(-)	(0.009)
log(Distance)	-	-1.454***	-0.980***	-	-
	[-1.181***]	[-1.494]	[-1.012]	[-]	[-]
	(0.005)	(0.006)	(0.007)	(-)	(-)
Land border	-	0.621***	0.231***	-	-
	[0.660***]	[0.643]	[0.244]	[-]	[-]
	(0.026)	(0.029)	(0.033)	(-)	(-)
Legal	-	0.262***	0.169***	-	-
	[0.172***]	[0.269]	[0.176]	[-]	[-]
	(0.007)	(0.008)	(0.009)	(-)	(-)
Language	-	0.737***	0.514***	-	-
	[0.663***]	[0.757]	[0.529]	[-]	[-]
	(0.009)	(0.01)	(0.012)	(-)	(-)
Colonial ties	-	1.345***	1.002***	-	-
	[0.342***]	[1.443]	[1.102]	[-]	[-]
	(0.036)	(0.061)	(0.07)	(-)	(-)
Currency Union	-	1.137***	0.775***	0.578***	0.421***
	[0.660***]	[1.173]	[0.807]	[0.64]	[0.497]
	(0.021)	(0.027)	(0.031)	(0.06)	(0.064)
FTA	-	1.059***	0.664***	0.130*	0.072
	[0.955***]	[1.077]	[0.674]	[0.123]	[0.054]
	(0.031)	(0.036)	(0.04)	(0.07)	(0.075)
WTO	-	0.228***	0.187***	0.095***	0.087***
	[0.462***]	[0.232]	[0.191]	[0.105]	[0.102]
	(0.009)	(0.014)	(0.016)	(0.028)	(0.031)
Fixed effects	i, j, t	it, jt	it, jt	it, jt, ij	it, jt, ij
Sample size	1204671	1204671	1171794	1204671	1171794
- perf. class.	12298	147760	141537	370617	374067
Deviance	8.857×10^5	6.976×10^5	5.2×10^5	4.728×10^5	4.184×10^5

Notes: Uncorrected coefficients in square brackets. Standard errors in parenthesis.

Table 27: Logit Estimation: APEs

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	-	-	0.331***	-	0.168***
	[-]	[-]	[0.332]	[-]	[0.13]
	(-)	(-)	(0.002)	(-)	(0.049)
log(Distance)	-	-0.138***	-0.067***	-	-
	[-0.140***]	[-0.137]	[-0.067]	[-]	[-]
	(0.005)	(0.005)	(0.001)	(-)	(-)
Land border	-	0.058***	0.016***	-	-
	[0.077***]	[0.059]	[0.016]	[-]	[-]
	(0.004)	(0.004)	(0.003)	(-)	(-)
Legal	-	0.025***	0.012***	-	-
	[0.020***]	[0.025]	[0.012]	[-]	[-]
	(0.001)	(0.001)	(0.001)	(-)	(-)
Language	-	0.069***	0.035***	-	-
	[0.078***]	[0.069]	[0.035]	[-]	[-]
	(0.003)	(0.001)	(0.001)	(-)	(-)
Colonial ties	-	0.122***	0.069***	-	-
	[0.040***]	[0.127]	[0.074]	[-]	[-]
	(0.004)	(0.006)	(0.006)	(-)	(-)
Currency Union	-	0.104***	0.053***	0.041***	0.027***
	[0.077***]	[0.104]	[0.054]	[0.04]	[0.028]
	(0.004)	(0.003)	(0.002)	(0.006)	(0.009)
FTA	-	0.098***	0.046***	0.009	0.004
	[0.110***]	[0.097]	[0.045]	[0.008]	[0.003]
	(0.005)	(0.004)	(0.003)	(0.006)	(0.006)
WTO	-	0.022***	0.013***	0.007**	0.005*
	[0.056***]	[0.021]	[0.013]	[0.006]	[0.006]
	(0.002)	(0.002)	(0.001)	(0.003)	(0.003)
Fixed effects	i, j, t	it, jt	it, jt	it, jt, ij	it, jt, ij
Sample size	1204671	1204671	1171794	1204671	1171794
- perf. class.	12298	147760	141537	370617	374067
Deviance	8.857×10^5	6.976×10^5	5.2×10^5	4.728×10^5	4.184×10^5

Notes: Uncorrected average partial effects in square brackets. Standard errors in parenthesis.

Table 28: Probit Estimation with Different Bandwidths: Coefficients

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	0.961***	1.112***	1.140***	1.154***	1.161***
	(0.036)	(0.037)	(0.039)	(0.04)	(0.04)
Currency Union	0.228***	0.217***	0.214***	0.214***	0.216***
	(0.05)	(0.048)	(0.048)	(0.048)	(0.047)
FTA	0.035	0.037	0.038	0.042	0.043
	(0.056)	(0.054)	(0.053)	(0.053)	(0.053)
WTO	0.041	0.039	0.039	0.040	0.042*
	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)
Trim	$L = 0$	$L = 1$	$L = 2$	$L = 3$	$L = 4$

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis.

Table 29: Probit Estimation with Different Bandwidths: Average Partial Effects

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	0.144***	0.173***	0.179***	0.182***	0.183***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Currency Union	0.026***	0.025***	0.024***	0.024***	0.025***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
FTA	0.004	0.004	0.004	0.005	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
WTO	0.005**	0.004**	0.004**	0.005**	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Trim	$L = 0$	$L = 1$	$L = 2$	$L = 3$	$L = 4$

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis.

Table 30: Logit Estimation with Different Bandwidths: Coefficients

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	1.606***	1.879***	1.929***	1.953***	1.965***
	(0.037)	(0.038)	(0.039)	(0.04)	(0.04)
Currency Union	0.448***	0.426***	0.421***	0.421***	0.425***
	(0.057)	(0.054)	(0.054)	(0.054)	(0.053)
FTA	0.065	0.069	0.072	0.077	0.080
	(0.063)	(0.061)	(0.06)	(0.06)	(0.06)
WTO	0.091***	0.087***	0.087***	0.088***	0.091***
	(0.028)	(0.027)	(0.027)	(0.027)	(0.027)
Trim	$L=0$	$L=1$	$L=2$	$L=3$	$L=4$

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis.

Table 31: Logit Estimation with Different Bandwidths: Average Partial Effects

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	0.133***	0.162***	0.168***	0.170***	0.172***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Currency Union	0.028***	0.027***	0.027***	0.027***	0.027***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
FTA	0.004	0.004	0.004	0.005	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
WTO	0.006***	0.005***	0.005***	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Trim	$L=0$	$L=1$	$L=2$	$L=3$	$L=4$

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis.

Table 32: Probit vs. OLS estimation: Average Partial Effects with Two-way Fixed Effects

	Dependent variable: y_{ijt}			
	(1)	(2)	(3)	(5)
lagged DV	-	-	0.599***	0.346***
	(-)	(-)	(0.001)	(0.003)
log(Distance)	-0.133***	-0.135***	-0.053***	-0.066***
	(0.001)	(0.005)	(0)	(0.001)
Land border	0.014***	0.035***	0.003*	0.015***
	(0.002)	(0.004)	(0.002)	(0.003)
Legal	0.008***	0.023***	0.002***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)
Language	0.098***	0.071***	0.040***	0.035***
	(0.001)	(0.001)	(0.001)	(0.001)
Colonial ties	0.021***	0.107***	0.008***	0.061***
	(0.003)	(0.007)	(0.002)	(0.005)
Currency Union	0.107***	0.103***	0.046***	0.053***
	(0.003)	(0.003)	(0.002)	(0.002)
FTA	-0.155***	0.090***	-0.063***	0.045***
	(0.002)	(0.004)	(0.002)	(0.003)
WTO	-0.010***	0.026***	-0.008***	0.013***
	(0.001)	(0.002)	(0.001)	(0.001)
Estimator	OLS	Probit	OLS	Probit
bias corrected	-	true	-	true
Sample size	1204671	1204671	1171794	1171794

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis.

Table 33: LPM Estimation with Different Bandwidths: Average Partial Effects

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	0.444***	0.466***	0.474***	0.480***	0.485***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Currency Union	0.008***	0.008**	0.008**	0.008**	0.008**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
FTA	-0.065***	-0.062***	-0.062***	-0.061***	-0.061***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
WTO	0.008***	0.008***	0.008***	0.008***	0.009***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Trim	$L=0$	$L=1$	$L=2$	$L=3$	$L=4$

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis.

References

Alessandria, George and Horag Choi (2007). “Do Sunk Costs of Exporting Matter for Net Export Dynamics?” In: *Quarterly Journal of Economics* 122.1, pp. 289–336.

Anderson, James E. and Eric van Wincoop (2003). “Gravity with Gravitas: A Solution to the Border Puzzle”. In: *American Economic Review* 93.1, pp. 170–192.

Baier, Scott L. and Jeffrey H. Bergstrand (2007). “Do Free Trade Agreements Actually Increase Members’ International Trade?” In: *Journal of International Economics* 71.1, pp. 72–95.

Balazsi, Laszlo, Laszlo Matyas, and Tom Wansbeek (2018). “The Estimation of Multidimensional Fixed Effects Panel Data Models”. In: *Econometric Reviews* 37.3, pp. 212–227.

Berman, Nicolas, Vincent Rebeyrol, and Vincent Vicard (2019). “Demand Learning and Firm Dynamics: Evidence from Exporters”. In: *Review of Economics and Statistics* 101.1, pp. 91–106.

Bernard, Andrew B., Esther Ann Bøler, Renzo Massari, Jose-Daniel Reyes, and Daria Taglioni (2017). “Exporter Dynamics and Partial-Year Effects”. In: *American Economic Review* 107.10, pp. 3211–3228.

Bernard, Andrew B. and J. Bradford Jensen (2004). “Why Some Firms Export”. In: *Review of Economics and Statistics* 86.2, pp. 561–569.

Correia, Sergio, Paulo Guimarães, and Thomas Zylkin (2019). “PPMLHDFE: Fast Poisson Estimation with High-Dimensional Fixed Effects”. In: *Unpublished Working Paper*.

Cruz-Gonzalez, Mario, Iván Fernández-Val, and Martin Weidner (2017). “Bias Corrections for Probit and Logit Models with Two-Way Fixed Effects”. In: *Stata Journal* 17.3, pp. 517–545.

Czarnowske, Daniel and Amrei Stammann (2019). “Binary Choice Models with High-Dimensional Individual and Time Fixed Effects”. In: *Unpublished Working Paper*.

Das, Sanghamitra, Mark J Roberts, and James R Tybout (2007). “Market Entry Costs, Producer Heterogeneity, and Export Dynamics”. In: *Econometrica* 75.3, pp. 837–873.

Dhaene, Geert and Koen Jochmans (2015). “Split-Panel jackknife Estimation of Fixed-Effect Models”. In: *Review of Economic Studies* 82.3, pp. 991–1030.

Eaton, Jonathan and Samuel Kortum (2002). “Technology, Geography, and Trade”. In: *Econometrica* 70.5, pp. 1741–1779.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz (2011). “An Anatomy of International Trade: Evidence From French Firms”. In: *Econometrica* 79.5, pp. 1453–1498.

Eaton, Jonathan, Samuel Kortum, and Sebastian Sotelo (2013). "International Trade: Linking Micro and Macro". In: *Advances in Economics and Econometrics: Tenth World Congress, Volume II: Applied Economics*. Ed. by Daron Acemoglu, Manuel Arellano, and Eddie Dekel. Cambridge University Press, pp. 329–370.

Egger, Peter and Mario Larch (2011). "An Assessment of the Europe Agreements' Effects on Bilateral Trade, GDP, and Welfare". In: *European Economic Review* 55.2, pp. 263–279.

Egger, Peter, Mario Larch, Kevin E Staub, and Rainer Winkelmann (2011). "The Trade Effects of Endogenous Preferential Trade Agreements". In: *American Economic Journal: Economic Policy* 3.3, pp. 113–143.

Feenstra, Robert C. (2004). *Advanced International Trade: Theory and Evidence*. Princeton, NJ: Princeton University Press.

Fernández-Val, Iván and Martin Weidner (2016). "Individual and time effects in nonlinear panel models with large N, T". In: *Journal of Econometrics* 192.1, pp. 291–312.

– (2018). "Fixed Effects Estimation of Large-T Panel Data Models". In: *Annual Review of Economics* 10.1, pp. 109–138.

Frisch, Ragnar and Frederick V. Waugh (1933). "Partial Time Regressions as Compared with Individual Trends". In: *Econometrica* 1.4, pp. 387–401.

Gaure, Simen (2013). "OLS with Multiple High Dimensional Category Variables". In: *Computational Statistics & Data Analysis* 66, pp. 8–18.

Guimarães, Paulo and Pedro Portugal (2010). "A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects". In: *Stata Journal* 10.4, pp. 628–649.

Hahn, Jinyong and Guido Kuersteiner (2007). "Bandwidth Choice for Bias Estimators in Dynamic Nonlinear Panel Models". In: *Unpublished Working Paper*.

Hahn, Jinyong and Hyungsik R. Moon (2006). "Reducing Bias of MLE in a Dynamic Panel Model". In: *Econometric Theory* 22.3, pp. 499–512.

Hahn, Jinyong and Whitney Newey (2004). "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models". In: *Econometrica* 72.4, pp. 1295–1319.

Halperin, Israel (1962). "The Product of Projection Operators". In: *Acta Sci. Math.(Szeged)* 23.1-2, pp. 96–99.

Head, Keith and Thierry Mayer (2014). "Gravity Equations: Workhorse, Toolkit, and Cookbook". In: *Handbook of International Economics*. Ed. by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff. 4th ed. Vol. 4. North Holland. Chap. 3, pp. 131–195.

Head, Keith, Thierry Mayer, and John Ries (2010). "The Erosion of Colonial Trade Linkages after Independence". In: *Journal of International Economics* 81.1, pp. 1–14.

Helpman, Elhanan, Marc J. Melitz, and Yona Rubinstein (2008). “Estimating Trade Flows: Trading Partners and Trading Volumes”. In: *Quarterly Journal of Economics* 123.2, pp. 441–487.

Hinz, Julian, Alexander Hudlet, and Joschka Wanner (2019). “Separating the Wheat from the Chaff: Fast Estimation of GLMs with High-Dimensional Fixed Effects”. In: *Unpublished Working Paper*.

Krugman, Paul (1980). “Scale Economies, Product Differentiation, and the Pattern of Trade”. In: *American Economic Review* 70.5, pp. 950–959.

Larch, Mario, Joschka Wanner, Yoto V. Yotov, and Thomas Zylkin (2019). “Currency Unions and Trade: A PPML Re-assessment with High-Dimensional Fixed Effects”. In: *Oxford Bulletin of Economics and Statistics* 81.3, pp. 487–510.

Lovell, Michael C. (1963). “Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis”. In: *Journal of the American Statistical Association* 58.304, pp. 993–1010.

Melitz, Marc J. (2003). “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”. In: *Econometrica* 71.6, pp. 1695–1725.

Neyman, J. and Elisabeth L. Scott (1948). “Consistent Estimates Based on Partially Consistent Observations”. In: *Econometrica* 16.1, pp. 1–32.

Nickell, Stephen (1981). “Biases in Dynamic Models with Fixed Effects”. In: *Econometrica* 49.6, pp. 1417–1426.

Piveteau, Paul (2019). “An Empirical Dynamic Model of Trade with Consumer Accumulation”. In: *Unpublished Working Paper*.

Redding, Stephen J. and Anthony J. Venables (2004). “Economic Geography and International Inequality”. In: *Journal of International Economics* 62.1, pp. 53–82.

Roberts, Mark J and James R Tybout (1997). “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs”. In: *American Economic Review* 87.4, pp. 545–564.

Ruhl, Kim J. and Jonathan L. Willis (2017). “New Exporter Dynamics”. In: *International Economic Review* 58.3, pp. 703–726.

Stammann, Amrei (2018). “Fast and Feasible Estimation of Generalized Linear Models with High-Dimensional k-way Fixed Effects”. In: *Unpublished Working Paper*.

Stammann, Amrei, Florian Heiß, and Daniel McFadden (2016). “Estimating Fixed Effects Logit Models with Large Panel Data”. In: *Unpublished Working Paper*.

Von Neumann, John (1950). *Functional Operators. Vol. II. The geometry of orthogonal spaces, volume 22 (reprint of 1933 notes) of Annals of Math.* Princeton University Press.

Weidner, Martin and Thomas Zylkin (2019). “Bias and Consistency in Three-way Gravity Models”. In: *Unpublished Working Paper*.