State Dependence and Unobserved Heterogeneity in the Extensive Margin of Trade*

JULIAN HINZ[†] AMREI STAMMANN[‡] JOSCHKA WANNER[§]

April 25, 2020

We study the role and drivers of persistence in the extensive margin of bilateral trade. Motivated by a stylized heterogeneous firms model of international trade with market entry costs, we propose new bias-corrected dynamic binary choice estimators with three sets of high-dimensional fixed effects. Monte Carlo simulations confirm their desirable statistical properties. A reassessment of the most commonly studied determinants of the extensive margin of trade demonstrates that both true state dependence and unobserved heterogeneity contribute strongly to trade persistence and that taking this persistence into account matters significantly in identifying the effects of trade policies on the extensive margin.

JEL Classification: C13, C23, C55, F14, F15
Keywords: Dynamic binary choice, extensive margin, high-dimensional fixed effects, incidental parameter bias correction, trade policy

*A previous version of this paper was circulated under the title "Persistent Zeros: The Extensive Margin of Trade". *Acknowledgements*: We thank Tanmay Belavadi, Daniel Czarnowske, Jonathan Eaton, Hartmut Egger, Miriam Frey, Mario Larch, Michael Pfaffermayr, Joel Stiebale, Yuta Watabe, Martin Weidner, and Thomas Zylkin for helpful comments. *Note*: An implementation of the estimators developed in this paper is available in the R-package *alpaca* on CRAN.

[†]*Affiliation:* Heinrich-Heine-University Düsseldorf, 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:* Kiel Institute for the World Economy. *Address:* Kiellinie 66, 24105 Kiel, Germany.

1 Introduction

What induces country pairs to trade? In 2015, still more than one quarter of potential bilateral trade relations reported zero trade flows.[1] Comparing these zero trade flows with trade relations in 2014, these zeros turn out to be extremely persistent: 84.6 percent of country pairs that did not trade in 2014 *did not* trade in 2015 either, as can be seen in the transition matrix depicted in Table 1. And similarly, 94.4 percent of pairs that *did* trade in 2014 continued to do so in the year after.[2]

Table 1: *Persistence in Bilateral Trade Relations (2014 – 2015)*

	Traded in 2015	
Traded in 2014	No	Yes
No	84.6 %	15.4 %
Yes	5.6 %	94.4 %

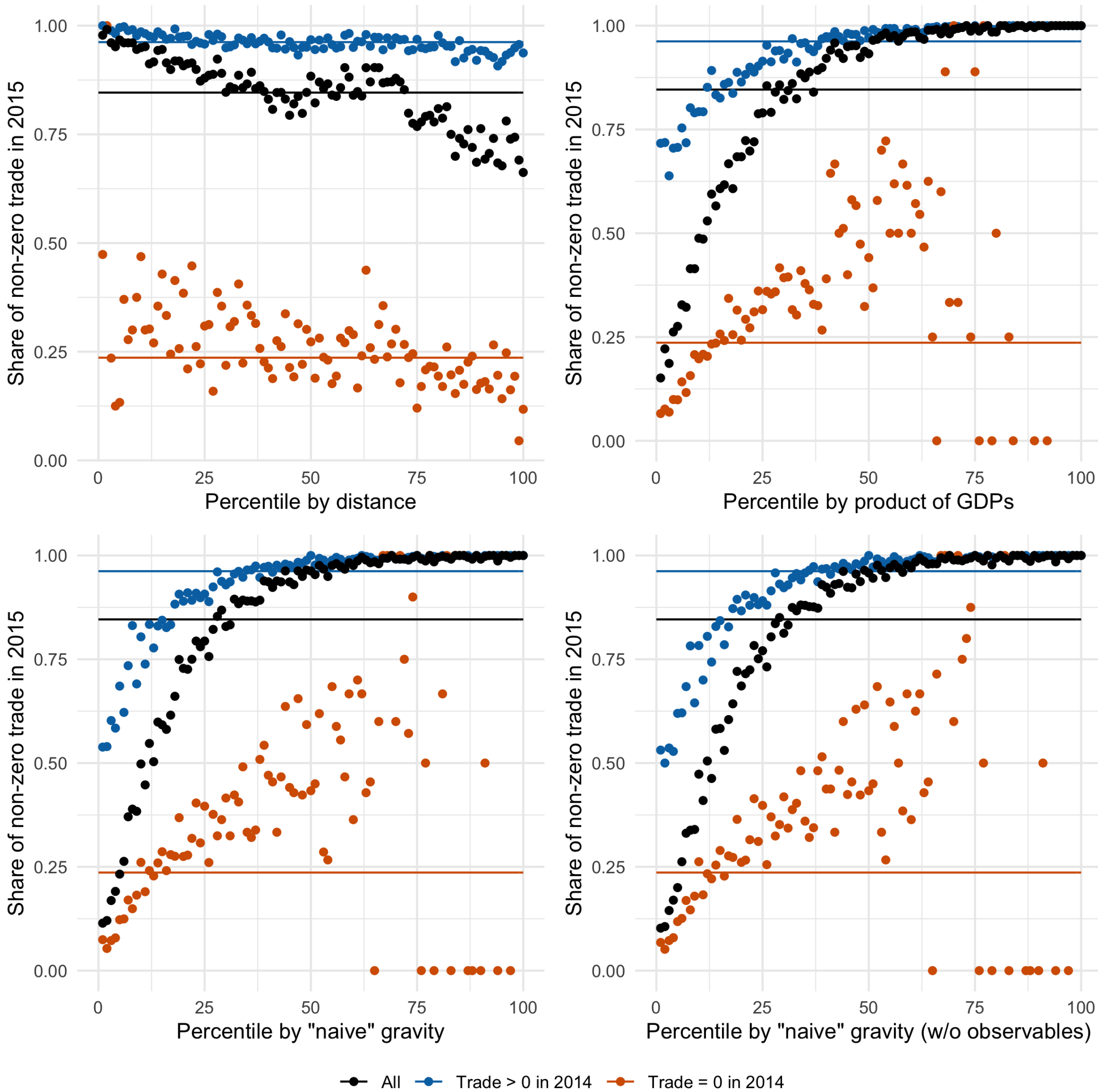
Both intuitively and based on existing theoretical and empirical insights, we would expect geographically close and economically large country pairs to have the greater bilateral trade potential and thus be more likely to engage in international trade. As distance is time-invariant and economic size does not change abruptly from one year to another, these gravity-like characteristics may explain (part of) the observed persistence. Figure 1 breaks down the share of nonzero trade flows in 2015 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 the countries' 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 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 the remarkable persistence of trade relations, even after controlling for differences in trade potential in terms of distance, size, and bilateral trade policy. More than 50 percent of those country pairs in the lowest percentile of trade potential trade again in 2015, provided they already did so in 2014. On the other hand, even comparably large and close pairs are likely not to trade in 2015 if they did not trade in 2014 either.[3]

[1] According to data from UN Comtrade.

[2] 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.

[3] A very similar pattern emerges for other points in time (see Figure A1 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 A2 in Appendix A for the years 1997–2006).

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



Note: Trade data for 2014 – 2015 come from UN Comtrade, GDP, distances and gravity variables are sourced from CEPII (Head and Mayer, 2014).

Two potential features of the extensive margin of trade that can generate the pattern documented by Table 1 and Figure 1 are what Heckman (1981) termed "true state dependence" — i.e. countries actually are more likely to trade *because* they did so in the previous period — and unobserved bilateral heterogeneity — i.e. persistence is due to unobservable factors continuously driving bilateral trade potential — denoted as "spurious state dependence" by Heckman (1981). In this paper we introduce estimators for the determinants of the extensive margin of international trade that explicitly take its persistence due to observable characteristics, true state dependence, and unobserved heterogeneity into account. We introduce features from the firm dynamics literature into a heterogeneous firms model of international trade with bounded productivity 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 computational 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 propose and implement new analytically and jackknife bias-corrected estimators for coefficients and average partial effects in three-way fixed effects specifications. Additionally, we provide an expression for long-run partial effects. Extensive simulation experiments demonstrate the desirable statistical properties of our proposed bias-corrected estimators. The empirical application provides 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 estimated effects of the most commonly studied potential determinants considerably: The impact of a common currency is reduced from about 10 percentage points to less than 4 percentage points, the impact of joint membership in the WTO decreases from 2.6 percentage points to 0.7 percentage points, and a common regional trade agreement loses statistical significance. Specifications with a lagged dependent variable and/or bilateral fixed effects further yield better predictions for which country pairs will trade than specifications that fail to account for state dependence appropriately.

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. Helpman, Melitz, and Rubinstein (2008) build a model of international trade with heterogeneous firms and bounded productivity in which a country only exports to a given destination if the most productive firm can afford to overcome the fixed costs of exporting. 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) embed the distinction between sunk costs and "period-by-period" fixed costs into general equilibrium.[4] We aim at reconciling the estimation of the aggregate extensive margin with the insight from the firm-level literature that dynamics feature prominently in the determination of the exporting decision by deriving an econometric specification that explicitly incorporates previous export experience at the country pair level.

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). Since Baier and Bergstrand (2007), it has furthermore become standard to include country pair fixed effects to tackle unobservable bilateral trade cost determinants. Additionally taking into account the multiplicative structure of the gravity equation following Santos Silva and Tenreyro (2006), nonlinear estimation with exporter-time, importer-time, and country pair fixed effects has become the gold standard for the intensive margin. Estimating the model introduced in this paper similarly calls for three sets of fixed effects, specific to exporters and importers in a given year, as well as to a given country pair over time. The binary nature of the decision whether to export to a destination market *at all*, also clearly asks for a nonlinear estimator. Therefore, in this paper, we put the estimation of the extensive margin on a par with the intensive margin gold standard by introducing a respective three-way fixed effects binary choice specification.

[4] A number of recent contributions also stress the dynamic character of firms' exporting behaviour and additionally provide alternative rationales for dynamic feedbacks beyond sunk costs of entry, such as "demand learning" or consumer accumulation (see e.g. Bernard, Bøler, Massari, Reyes, and Taglioni, 2017; Ruhl and Willis, 2017; Berman, Rebeyrol, and Vicard, 2019; Piveteau, 2019).

Third, the paper builds on and contributes to the literature on estimating nonlinear fixed effects models. Our proposed three-way fixed effects specification faces two difficulties. First, the large number of fixed effects poses a computational challenge for nonlinear estimators. Here we can rely on the recently suggested pseudo-demeaning algorithm by Stammann (2018) to overcome this issue. Second, however, as is known since Neyman and Scott (1948), the inclusion of fixed effects potentially introduces an incidental parameter problem, leading to inconsistent estimates. Recently, there have been a number of advances to deal with this problem, and a variety of approaches have been proposed (see Fernández-Val and Weidner (2018) for a recent overview). For estimating the extensive margin with binary choice models, there are so far two approaches, both of which focus exclusively on models for *cross-sectional* bilateral data with only importer and exporter fixed effects. Cruz-Gonzalez, Fernández-Val, and Weidner (2017) apply the bias correction of Fernández-Val and Weidner (2016), and Charbonneau (2017) proposes a conditional logit estimator. Our contribution is to develop suitable fixed effects binary choice estimators for bilateral cross-sectional data *over time*, i.e. network panel data. While it is sufficient to adapt the two-way bias corrections of Fernández-Val and Weidner (2016) for the theory-consistent estimation of our model including the fixed effects for exporter-time (it) and importer-time (jt), we develop a new bias correction for our preferred specification, which additionally includes a third, bilateral (ij), set of fixed effects.[5] Therefore, our article complements the work of Weidner and Zylkin (2020) on estimating the intensive margin of trade, who examine the incidental parameter problem in three-way pseudo-poisson (PPML) models under fixed T asymptotics and suggest appropriate bias corrections.

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 new bias-corrected three-way fixed effects estimator. We demonstrate its performance in Monte Carlo simulations in Section 4, before finally showing the estimator in action by estimating the model in Section 5. Section 6 concludes.

2 An Empirical Model of the Extensive Margin of Trade

We start by setting up a model of the extensive margin of trade that will later guide our econometric specification. We consider a stylized dynamic Melitz (2003)-type heterogeneous firms

[5] Similarly, it is possible to adapt the estimator of Charbonneau (2017) to the setting with exporter-time (it) and importer-time (jt) fixed effects. However, her approach is not suitable for our purposes for several reasons: 1. it does not allow for dynamics 2. it is limited to logit models, 3. it precludes the possibility to estimate average partial effects, 4. it is computationally infeasible in cases where the number of levels per fixed effects becomes large.

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.[6] Each period, a firm can decide to pay a fixed cost f^{prod}_{it} and start production of a differentiated variety using labour l as its only input, such that $l_t(\omega) = f^{prod}_{it} + q_t(\omega)/\varphi_t(\omega)$. 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

[6] 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.

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:

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

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 $\mathbb{E}_t[f_{ij(t+s)}^{exp}] = f_{ijt}^{exp}\ \forall s \in \mathbb{N}$.[7] 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

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

[7]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 $\bar{g} < \delta$.

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:

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

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. Empirically, (vii) induces true state dependence. As previous exporters do not have to incur entry costs, they are more likely to stay active in the destination market and the extensive margin becomes more persistent than would be implied merely by the persistence of productivity, market potential, and trade costs.

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 geographic distance, whether i and j are both WTO members, and 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.[8] Note, however, that the nature of the persistence captured by these fixed effects is different from the one that is due to the entry dynamics. This additional state dependence is "spurious" in the sense that countries are not actually more likely to export to a destination because of the prior experience, but because they keep incorporating the same unobserved factors over time. With the three sets of fixed effects and our parametrization for time-varying trade cost determinants, we arrive at the following econometric model:

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

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. As mentioned above, we capture the three sets of unobserved components by introducing according sets of fixed effects. A supposed alternative using random effects is actually not possible, at least for the it and jt effects, as they are implied by the theoretical model. We therefore cannot make the distributional assumptions required in a random effects setting. The theory is silent about the exact form of the bilateral heterogeneity. We decide for a third set of fixed effects as the most general option in order to avoid assumptions on its distribution or its correlation to observed factors.

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

[8] 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.

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 Dynamic Binary Choice Estimators with Three-Way Fixed Effects

Having set up the empirical framework, we now turn to the estimation procedure. As equation (3) demands 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 new bias corrections to address the induced incidental parameter problem. In a third step, we show how long-run average partial effects can be estimated.

3.1 Feasible Estimation with High-Dimensional Three-Way Fixed Effects

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 $\boldsymbol{\alpha}_{ijt} = [\lambda_{it}, \psi_{jt}, \mu_{ij}]'$.[9] Further, F_{ijt} is either the logistic or the standard normal cumulative distribution function. See Table 2 for the relevant expressions and derivatives.

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

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

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

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

Table 2: *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: $\eta_{ijt} = \mathbf{x}'_{ijt}\boldsymbol{\beta} + \lambda_{it} + \psi_{jt} + \mu_{ij}$ is the linear predictor.

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

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

The brute-force computation of equation (4) quickly becomes computationally demanding, if not impossible.[10] 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 negative 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.

[10] In a balanced data set ($I = J = N$) with three-way fixed effects, the number of parameters to be estimated is $\approx N(N-1) + 2NT$. In a trade panel data set with 200 countries and 50 years, the number of fixed effects in this case amounts to 59800 parameters.

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).[11] 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

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

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 $\widehat{\Delta}_{ijt}^{k} = \widehat{F}_{ijt}|_{x_{ijtk=1}} - \widehat{F}_{ijt}|_{x_{ijtk=0}}$ 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. In its standard form, though, 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 a related context with 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 three-way error structure.[12] An appropriate covariance estimator for the APEs of the three-way fixed effects model is

$$\begin{aligned} \widehat{\mathbf{V}}^{\delta} = \frac{1}{I^2 J^2 T^2} \Bigg(& \underbrace{\left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\widehat{\boldsymbol{\Delta}}}_{ijt} \right) \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\widehat{\boldsymbol{\Delta}}}_{ijt} \right)^{\prime}}_{v_1} + \underbrace{\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\boldsymbol{\Gamma}}_{ijt} \widehat{\boldsymbol{\Gamma}}_{ijt}^{\prime}}_{v_2} \\ & + \underbrace{2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{s>t} \widehat{\widehat{\boldsymbol{\Delta}}}_{ijt} \widehat{\boldsymbol{\Gamma}}_{ijs}^{\prime}}_{v_3} \Bigg) \, , \end{aligned} \tag{6}$$

[11] 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).

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

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

$$\widehat{\boldsymbol{\Gamma}}_{ijt} = \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \partial_{\beta} \widehat{\boldsymbol{\Delta}}_{ijt} - \left(\widehat{\mathbb{P}} \mathbf{X} \right)_{ijt} \partial_{\eta} \widehat{\boldsymbol{\Delta}}_{ijt} \right)' \widehat{\mathbf{A}}^{-1} \left(\widehat{\mathbb{M}} \mathbf{X} \right)_{ijt} \hat{\omega}_{ijt} \hat{\boldsymbol{\nu}}_{ijt} - \left(\widehat{\mathbb{P}} \widehat{\boldsymbol{\Psi}} \right)_{ijt} \partial_{\eta} \hat{\ell}_{ijt} \, ,$$

with $\widehat{\mathbf{A}} = (\widehat{\mathbb{M}}\mathbf{X})'\widehat{\boldsymbol{\Omega}}(\widehat{\mathbb{M}}\mathbf{X})$, $\widehat{\boldsymbol{\Psi}}_{ijt} = \partial_{\eta} \widehat{\boldsymbol{\Delta}}_{ijt} / \hat{\omega}_{ijt}$, and $\partial_{\eta} \hat{\ell}_{ijt}$ defined in Table 2. To clarify 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.[13] 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

[13] 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.

analytical and split-panel jackknife bias corrections for structural parameters and APEs in the context of nonlinear models with individual and time fixed effects.

We adapt and extend the bias corrections of Fernández-Val and Weidner (2016) to our three-way error component.[14,15] Fernández-Val and Weidner (2018) conjecture, based on their previously discussed formula, $bias \sim p/n$, that the bias of a three-way fixed effects estimator in an ijt-panel setting is of order $(IT+JT+IJ)/(IJT)$ and of the form $B_1/I+B_2/J+B_3/T$. In line with the bias structure in two-way error component models, the inclusion of importer-time and exporter-time fixed effects entails two bias terms of order $1/I$ and $1/J$, respectively. 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. Based on this conjecture we propose novel analytical and jackknife bias corrections for three-way fixed effects models. In Appendix B.2, we illustrate the statistical problem and the working of bias corrections with a version of the prominent Neyman and Scott (1948) variance example.

For the split-panel jackknife bias correction, this three-part bias structure implies that we need to split our panel across three dimensions, leading to the following estimator for the structural parameters:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}^{sp} &= 4\widehat{\boldsymbol{\beta}}_{I,J,T} - \widehat{\boldsymbol{\beta}}_{I/2,J,T} - \widehat{\boldsymbol{\beta}}_{I,J/2,T} - \widehat{\boldsymbol{\beta}}_{I,J,T/2}, \quad \text{with} \\
\widehat{\boldsymbol{\beta}}_{I/2,J,T} &= \frac{1}{2}\Big[\widehat{\boldsymbol{\beta}}_{\{i:i\leq\lfloor I/2\rfloor,J,T\}} + \widehat{\boldsymbol{\beta}}_{\{i:i\geq\lceil I/2+1\rceil,J,T\}}\Big], \\
\widehat{\boldsymbol{\beta}}_{I,J/2,T} &= \frac{1}{2}\Big[\widehat{\boldsymbol{\beta}}_{\{I,j:j\leq\lfloor J/2\rfloor,T\}} + \widehat{\boldsymbol{\beta}}_{\{I,j:j\geq\lceil J/2+1\rceil,T\}}\Big], \\
\widehat{\boldsymbol{\beta}}_{I,J,T/2} &= \frac{1}{2}\Big[\widehat{\boldsymbol{\beta}}_{\{I,J,t:t\leq\lfloor T/2\rfloor\}} + \widehat{\boldsymbol{\beta}}_{\{I,J,t:t\geq\lceil T/2+1\rceil\}}\Big].
\end{aligned}
\tag{7}
$$

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.

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

[14] In Appendix B.3, we also derive the bias corrections for a two-way fixed effects model in our ijt network panel structure. Previous two-way bias corrections considered either classical it panel structures or ij pseudo-panels.

[15] We do not elaborate on the leave-one-out jackknife bias correction because it requires all variables to be independent over time and thus rules out predetermined and serially-correlated regressors (Fernández-Val and Weidner, 2018).

conjecture for the asymptotic MLE distribution in the three-way setting (which we present in Appendix B.3) and propose the following analytical bias correction which subtracts estimates of the leading bias terms from the MLE:

$$\tilde{\boldsymbol{\beta}}^a = \hat{\boldsymbol{\beta}}_{I,J,T} - \frac{\widehat{\mathbf{B}}_1^\beta}{I} - \frac{\widehat{\mathbf{B}}_2^\beta}{J} - \frac{\widehat{\mathbf{B}}_3^\beta}{T}, \quad \text{with} \quad \widehat{\mathbf{B}}_1^\beta = \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{B}}_1, \widehat{\mathbf{B}}_2^\beta = \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{B}}_2, \widehat{\mathbf{B}}_3^\beta = \widehat{\mathbf{W}}^{-1}\widehat{\mathbf{B}}_3, \quad (8)$$

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

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

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

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

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 2.

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 $L \in \{1, 2, 3, 4\}$.

The first two correction terms in equation (8) are generalizations of the corresponding components in the ij-pseudo panel structure of Cruz-Gonzalez, Fernández-Val, and Weidner (2017) to our ijt structure (see Appendix B.3 for the derivation in a two-way it and jt fixed effects setting). The additional inclusion of a third set of (ij) fixed effects additionally leads to the third correction term that mimics the correction for individual fixed effects in an it-panel setting.

Moving to the APEs, the split-panel jackknife estimator is formed by replacing the estimators for the structural parameters with estimators for the APEs in formula (7). The analytically bias-corrected estimator, based on our conjecture for the asymptotic distribution provided in Appendix B.3, is given by

$$\tilde{\boldsymbol{\delta}}^a = \hat{\boldsymbol{\delta}} - \frac{\widehat{\mathbf{B}}_1^\delta}{I} - \frac{\widehat{\mathbf{B}}_2^\delta}{J} - \frac{\widehat{\mathbf{B}}_3^\delta}{T}, \quad \text{with} \tag{9}$$

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

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

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

The last part in the numerator of $\widehat{\mathbf{B}}_3^\delta$ is again dropped if all regressors are assumed to be strictly exogenous. Note that all quantities are evaluated at bias-corrected structural parameters and the corresponding estimates of the fixed effects.[16] Standard errors can still be obtained from equation (6).

3.3 Long-Run Average Partial Effects

In dynamic models, the simple average partial effect $\hat{\delta}_k$ does not provide the full picture of how the export probability is affected by a change in a regressor. Rather, there are additional feedback effects: In our context, the introduction of a permanent trade policy that increases the probability to export to a destination implies that in the next period, entry costs are more likely to have already been paid, and hence the impact becomes higher with increasing duration of the policy. To derive expressions for long-run effects, which additionally take these dynamic feedbacks into account, we make use of the long-run probability of $y_{ijt} = 1$ for a given set of regressors and fixed effects, also mentioned in Carro (2003) and Browning and Carro (2010):

$$\widetilde{F}_{ijt} = \frac{\widehat{F}_{ijt}|_{y_{ij(t-1)=0}}}{1 - \widehat{\Delta}_{ijt}^{y}}, \tag{10}$$

where $\widehat{\Delta}_{ijt}^{y} = \widehat{F}_{ijt}|_{y_{ij(t-1)=1}} - \widehat{F}_{ijt}|_{y_{ij(t-1)=0}}$. Long-run average partial effects are then given by

[16] For this purpose, we use a computationally efficient offset algorithm as in Czarnowske and Stammann (2019).

$$\hat{\delta}_k^{LR} = \frac{1}{IJT} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\Delta}_{ijt}^{k,LR}, \text{ with } \widehat{\Delta}_{ijt}^{k,LR} = \frac{\partial \widetilde{F}_{ijt}}{\partial x_{ijtk}} \text{ or } \widehat{\Delta}_{ijt}^{k,LR} = \widetilde{F}_{ijt}|_{x_{ijtk=1}} - \widetilde{F}_{ijt}|_{x_{ijtk=0}}, \quad (11)$$

in the case of continuous or binary regressors, respectively. We obtain bias-corrected estimates of δ_k^{LR} by substituting individual long-run partial effects $\widehat{\Delta}_{ijt}^{k,LR}$ for the simple individual partial effects in equation (9). Corresponding standard errors are calculated by applying the same substitution in the concentrated delta method given in equation (6).

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).[17] 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 with three-way fixed effects.[18] In line with our theoretical model, the simulations include unobserved components captured by fixed effects in the it, jt, and ij, as well as the lagged dependent variable. Specifically, we generate data according to

$$\begin{aligned} y_{ijt} &= \mathbf{1}[\beta_y y_{ijt-1} + \beta_x x_{ijt} + \lambda_{it} + \psi_{jt} + \mu_{ij} \geq \epsilon_{ijt}] \ , \\ y_{ij0} &= \mathbf{1}[\beta_x x_{ij0} + \lambda_{i0} + \psi_{j0} + \mu_{ij} \geq \epsilon_{ij0}] \ , \end{aligned}$$

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 $\epsilon_{ijt} \sim \text{iid.}\, \mathcal{N}(0, 1)$.[19] 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

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

[18]Further simulation experiments including dynamic panel models with two-way fixed effects and static panel models with three-way fixed effects are presented in Appendices C.2 and C.3. In an earlier version of this article we additionally report simulations results for static two-way fixed effects models.

[19]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.

$\nu_{ijt} \sim \text{iid.}\ \mathcal{N}(0, 0.5)$ and $x_{ij0} \sim \text{iid.}\ \mathcal{N}(0, 1)$. 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 A4 – A9 in Appendix C.1 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 focus on the biases and coverage probabilities for $N \in \{50, 150\}$ which are shown in Figures 2 and 3.

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 $N = 150$ and $T = 50$, 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 64 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 proceed with the APEs, where we also consider LPM as an alternative estimator. 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.[20] These results illustrate the superiority of

[20]A similar behaviour of LPM has been observed by Czarnowske and Stammann (2019) in the context of a

Figure 2: *Dynamic: Three-way Fixed Effects – Predetermined Regressor*

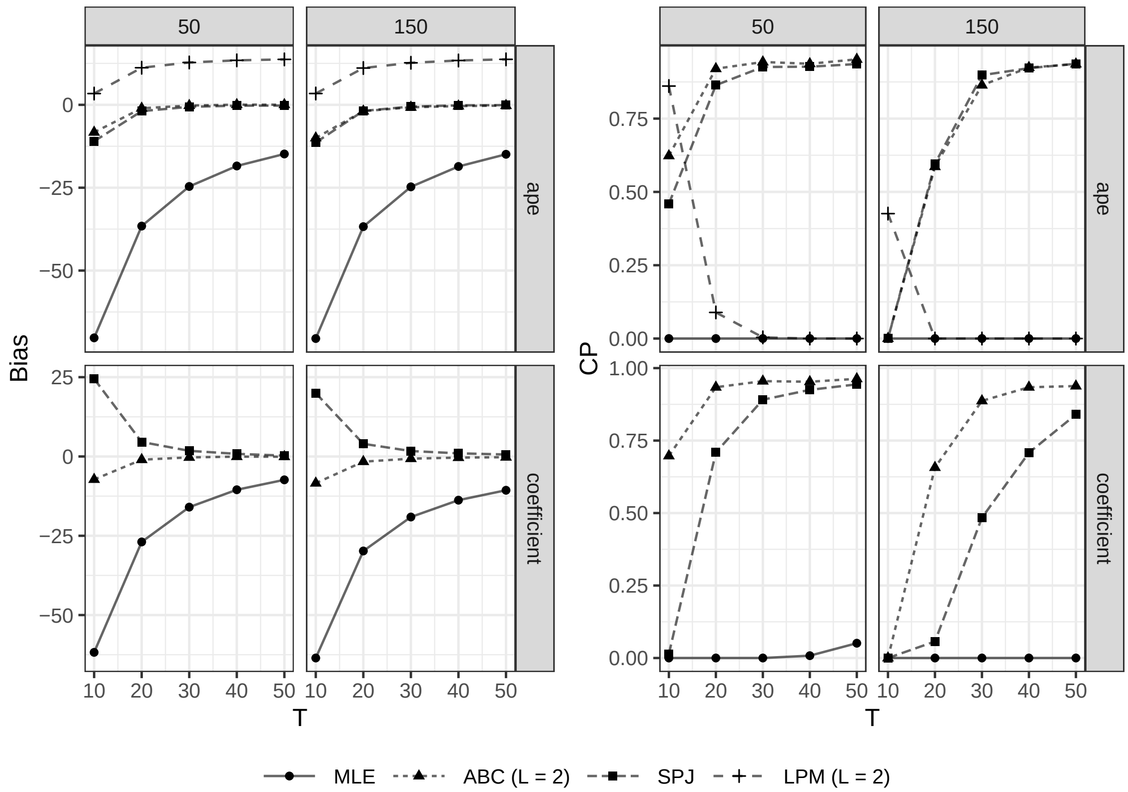
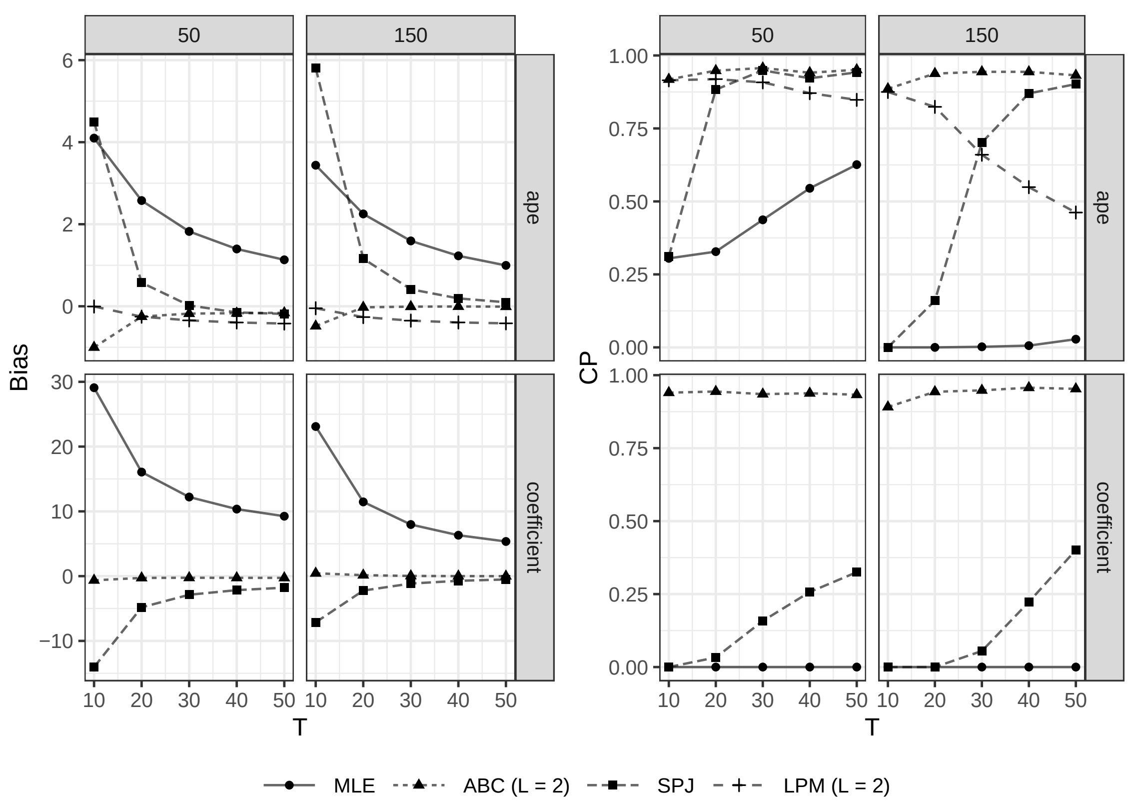


Figure 3: *Dynamic: Three-way Fixed Effects – Exogenous Regressor*



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 of the bias and lend support to our bias corrections. 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.

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, trade in the previous period, time-invariant unobservable trade barriers and bilateral trade policy variables:

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

This yields the following dynamic three-way fixed effects probit model:

$$\Pr(y_{ijt} = 1 | y_{ij(t-1)}, \mathbf{x}_{ijt}, \lambda_{it}, \psi_{jt}, \mu_{ij}) = F\left(\beta_y y_{ij(t-1)} + \mathbf{x}'_{ijt}\boldsymbol{\beta}_x + \lambda_{it} + \psi_{jt} + \mu_{ij}\right)\,, \qquad (12)$$

$y_{ij(t-1)}$ is the lagged dependent variable, $\mathbf{x}$ is a vector of observable bilateral variables,

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 $\mathbf{x}$, 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.

and 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. The effect of all time-invariant variables will only be identified in specifications in which we omit the bilateral fixed effects. In terms of data, we turn to the comprehensive gravity dataset provided alongside Head, Mayer, and Ries (2010), which encompasses annual information on bilateral trade flows and these variables of interest of 208 countries from 1948 – 2006.

5.1 Main results

Before turning to the regression results, we repeat the descriptive analysis about the persistence of the bilateral trade flows from Section 1, now considering the transition probabilities into export from period $t-1$ to t for the time horizon from 1948 – 2006. Table 3 confirms the high level of persistence: 86.8 percent of the countries which did not trade in the previous year did not trade in the following year and 92.3 percent of the countries which *did* trade in the previous year continued to trade in the following year. Thus, 79.1 percent of the *unconditional* probability to export in a given period can be explained by the export status in the previous period.[21] However, Table 3 does not reveal any information about the kind of persistence. In the following analysis we will investigate the importance of using dynamic model specifications which allow us to disentangle the observed persistence into two sources: (i) true state dependence and (ii) observed and unobserved heterogeneity.

Table 3: *Transition Probabilities*

	$y_{ijt}=0$	$y_{ijt}=1$
$y_{ij(t-1)}=0$	86.8 %	13.2 %
$y_{ij(t-1)}=1$	7.7 %	92.3 %

Table 4 reports average partial effects of several static and dynamic fixed effects probit specifications.[22] Bias-corrected estimates and their corresponding standard errors are printed in bold. For comparison, the uncorrected estimates are also shown. In column (1) we first mimic the static specification estimated by Helpman, Melitz, and Rubinstein (2008).[23] Their

[21]The number is computed as difference between the probability of exporting in period t conditional on exporting and not exporting in period $t-1$.

[22]Coefficient estimates are reported in Table A19 in the Appendix.

[23]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 estimates deviate somewhat from theirs, while remaining qualitatively similar.

specification includes exporter, importer, and time fixed effects.[24] All average partial effects have the expected sign, indicating a negative impact of distance on the probability to trade, while having a common border, the same origin of the legal system, a shared language, or a joint colonial history are all estimated to have a positive impact. Also 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. This specification can be considered a theory-consistent estimation of the model by HMR, and of our model if entry costs are zero and other bilateral trade cost determinants are fully observable. The average partial effects are qualitatively the same and quantitatively similar for most variables to those in column (1). However, e.g. the estimated effect of colonial ties more than doubles and the estimated WTO effect is roughly cut in half.

Specification (3) keeps the same fixed effects, but adds a lagged dependent variable and thus controls for one type of persistence. Assuming no unobservable bilateral heterogeneity, this specification correctly estimates the model set up in Section 2. As the partial effects of static and dynamic models are not directly comparable due to the feedbacks involved in the latter, we show two types of average partial effects for our dynamic specifications: the usual *direct* effects and the *long-run* effects described in Section 3.3. The first result to note for the third specification is the highly significant average partial effect 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 35 percentage points higher probability of nonzero trade, given the two countries were also engaged in trade in the previous year. This implies that 43.7 percent of the observed persistence are attributed to true state dependence in this first dynamic specification.[25] In terms of our model, this suggests a vast effect of market entry costs on the aggregate extensive margin. The second observation is that direct APEs are about 50 percent smaller than in column (2) across the board. However, once dynamic adjustments are taken into account, the average partial effects resulting from specifications (2) and (3) become very similar, suggesting that accounting for the market entry dynamics mainly matters for getting the *timing* of trade policy effects right, rather than for the overall magnitude of the effects.

[24]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.

[25]This value is calculated as the ratio of the estimated average partial effect and the unconditional exporting probability $(34.6/79.1)$.

Table 4: *Probit Estimation: Average partial effects*

	Dependent variable: y_{ijt}						
	(1)	(2)	(3)		(4)	(5)	
			direct	*long-run*		*direct*	*long-run*
lagged DV	-	-	**0.346*****	-	-	**0.179*****	-
	(-)	(-)	(**0.034**)	(-)	(-)	(**0.052**)	(-)
	-	-	0.344***	-	-	0.138***	-
	(-)	(-)	(0.035)	(-)	(-)	(0.048)	(-)
log(Distance)	-	**-0.135*****	**-0.066*****	**-0.133*****	-	-	-
	(-)	(**0.013**)	(**0.007**)	(**0.013**)	(-)	(-)	(-)
	-0.136***	-0.135***	-0.066***	-0.132***	-	-	-
	(0.001)	(0.014)	(0.007)	(0.013)	(-)	(-)	(-)
Land border	-	**0.035*****	**0.015*****	**0.030*****	-	-	-
	(-)	(**0.005**)	(**0.003**)	(**0.006**)	(-)	(-)	(-)
	0.054***	0.035***	0.015***	0.030***	-	-	-
	(0.004)	(0.005)	(0.003)	(0.006)	(-)	(-)	(-)
Legal	-	**0.023*****	**0.011*****	**0.022*****	-	-	-
	(-)	(**0.002**)	(**0.001**)	(**0.003**)	(-)	(-)	(-)
	0.019***	0.023***	0.011***	0.022***		-	-
	(0.001)	(0.002)	(0.001)	(0.003)	(-)	(-)	(-)
Language	-	**0.071*****	**0.035*****	**0.070*****	-	-	-
	(-)	(**0.007**)	(**0.004**)	(**0.007**)	(-)	(-)	(-)
	0.078***	0.071***	0.035***	0.069***	-	-	-
	(0.001)	(0.007)	(0.004)	(0.007)	(-)	(-)	(-)
Colonial ties	-	**0.107*****	**0.061*****	**0.117*****	-	-	-
	(-)	(**0.013**)	(**0.008**)	(**0.015**)	(-)	(-)	(-)
	0.039***	0.111***	0.066***	0.125***	-	-	-
	(0.004)	(0.013)	(0.008)	(0.015)	(-)	(-)	(-)
Currency union	-	**0.103*****	**0.053*****	**0.103*****	**0.038*****	**0.024*****	**0.038*****
	(-)	(**0.011**)	(**0.006**)	(**0.011**)	(**0.012**)	(**0.009**)	(**0.013**)
	0.078***	0.103***	0.054***	0.103***	0.037***	0.025**	0.036**
	(0.003)	(0.011)	(0.006)	(0.011)	(0.013)	(0.010)	(0.014)
FTA	-	**0.089*****	**0.045*****	**0.088*****	**0.009**	**0.004**	**0.007**
	(-)	(**0.010**)	(**0.005**)	(**0.010**)	(**0.007**)	(**0.006**)	(**0.009**)
	0.103***	0.088***	0.044***	0.086***	0.008	0.003	0.005
	(0.004)	(0.010)	(0.005)	(0.010)	(0.007)	(0.006)	(0.009)
WTO	-	**0.026*****	**0.013*****	**0.026*****	**0.006****	**0.004**	**0.007***
	(-)	(**0.003**)	(**0.002**)	(**0.004**)	(**0.003**)	(**0.003**)	(**0.004**)
	0.061***	0.026***	0.013***	0.025***	0.006*	0.005	0.007
	(0.001)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)	(0.004)
Fixed effects	i, j, t	it, jt	it, jt		it, jt, ij	it, jt, ij	
Sample size	1204671	1204671	1171794		1204671	1171794	
Deviance	8.891×10^5	7.019×10^5	5.183×10^5		4.76×10^5	4.189×10^5	

Notes: Column (1) uncorrected average partial effects, columns (2) - (5) bias-corrected average partial effects (bold font) and uncorrected average partial effects (normal font). Column (5) bias-corrected with L = 2. Standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Specification (4) 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 a second type of persistence — bilateral unobserved heterogeneity. This also 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. These results highlight the importance of controlling for unobserved country pair heterogeneity to avoid endogeneity problems associated with trade policy variables.

Finally, in the last two columns we present the results from our preferred specification (5), which implements equation (12). The estimation again includes the "full set" of fixed effects, i.e. exporter-time, importer-time and bilateral fixed effects, now combined with the lagged dependent variable, and therefore controls for both kinds of persistence simultaneously. Again, the average partial effect on the lagged dependent variable is highly significant. It now entails a partial effect of about 18 percentage points, i.e. roughly 22.6 percent of the observed persistence can be attributed to true state dependence and 77.4 percent to observed and unobserved factors. Failure to account for unobserved heterogeneity in specification (3) hence overestimated the importance of the lagged dependent variable (corresponding to entry costs in our model) roughly by a factor of two and therefore mislabelled a substantial part of spurious as true state dependence. Considering the effects of the time-varying trade policy variables, the only remaining statistically significant direct average partial effect in column (5) is estimated for a common currency at 2.4 percentage points. The direct impacts of a free trade agreement or joint membership of the WTO are statistically insignificant. Just as in the comparison between specifications (2) and (3), the average partial effects again become very similar when the static effects are compared to the long-run effects in the dynamic specification.

When comparing bias-corrected and uncorrected average partial effects throughout Table 4, it is noticeable that both differ only slightly for the exogenous regressors within the different specifications. The most significant impact is observed on the average partial effect for the predetermined variable, which in specification (5) differs by almost 24 percent. These results are in line with the theoretical properties of the estimators and the findings of our simulation study.[26] Despite the small biases for the exogenous variables, a bias correction is still necessary for our three-way fixed effects specifications because the biases are not negligible relative to the standard errors and thus inference of the uncorrected estimator is invalid. Note that for applications with a shorter time horizon, the biases will be more evident.

5.2 Predictive Analysis

After we have seen that the presented innovations matter significantly for the estimation of extensive margin determinants, we now consider the predictive performance of the different specifications discussed above. As an additional benchmark, we also look at a naive — purely descriptive — approach that predicts export decisions in period t solely based on the export decision in period $t-1$. For the resulting altogether six options, we evaluate the predictive power using the following measures: the total share of correctly predicted export decisions (accuracy), the share of correctly predicted decisions conditional on not exporting (true negative rate), and the share of correctly predicted decisions conditional on exporting (true positive rate). As becomes clear in Table 5, the naive approach already works very well for predictive purposes. Close to 90 percent of the exporting decisions in period t are correctly predicted by simply reproducing exporting decision from $t-1$, irrespective of the measure we are considering. As it turns out, both the static model with i, j, and t fixed effects (specification (1)) and the static model with it and jt fixed effects (specification (2)) produce poorer predictions than the naive approach — clearly so by about five percentage points in the former case. However, this changes once persistence is explicitly taken into account (specifications (3) to (5)). All of them at least slightly improve the predictions of the naive approach. Whether persistence is incorporated *only* via the lagged dependent variable *or* via pair fixed effects turns out to yield very similar predictive quality. Combining both in our preferred specification — the dynamic three-way fixed effects models — yields the highest predictive power. We are able to correctly predict 92.5 percent of the export decisions, 91.4 percent of the decisions conditional on not exporting, and 93.2 percent of the decisions conditional on exporting, implying an improvement compared to the naive approach by 2.4 – 2.5 percentage points. All in all, the predictive analysis underlines the importance of controlling for true state dependence and unobserved bilateral factors simultaneously.

[26] For the two-way models in columns (2) - (3) the order of the bias is $1/I + 1/J$, i.e. the bias only depends on the number of exporters/importers. For the three-way models in columns (4) - (5) the order of the bias is $1/I + 1/J + 1/T$, i.e. the bias additionally depends on the number of time periods. In our application the number of exporters/importers is relatively large but the number of time periods is substantially lower.

Table 5: *Predictive Analysis*

	Naive	(1)	(2)	(3)	(4)	(5)
Accuracy	90.0 %	83.0 %	86.9 %	91.1 %	91.3 %	92.5 %
True negative rate	88.9 %	78.9 %	83.6 %	89.6 %	90.0 %	91.4 %
True positive rate	90.8 %	86.0 %	89.2 %	92.2 %	92.2 %	93.2 %

Notes: Column (1) uncorrected probit model, columns (2) – (5) bias-corrected probit model.

5.3 Robustness Checks

We next consider two robustness checks for our main results from Section 5.1. First, while our preferred estimation of the extensive margin of trade with a dynamic three-way fixed effects binary choice estimator follows from the stylized facts and our theoretical model, the decision on *which* binary choice estimator to use hinges on the distributional assumption for the error term (and hence for the demand shock in our model). We followed Eaton, Kortum, and Kramarz (2011) and assumed log-normal shocks, leading us to using a probit for our main estimations. We now consider log-logistic shocks and the resulting logit estimator instead. Table 6 displays the results for the same specifications as in Table 4, but estimated using a logit. Reassuringly, the average partial effects are very similar to the probit case for all variables in all specifications. Introducing the different innovations step-by-step changes the estimated effects in the same way as for the probit. The insights discussed above therefore do not hinge on the specific distributional assumption made in the parametrization of the probability of success.

Second, in the estimation of our preferred specification (5) we have discretionary power in one dimension for the exact form of bias correction used. Specifically, the bandwidth parameter L used to estimate the spectral densities has to be chosen. We therefore follow the recommendation by Fernández-Val and Weidner (2016) and investigate the sensitivity of the results with respect to L. Table 7 depicts the direct and long-run average partial effects obtained with the bias-corrected dynamic three-way fixed effects probit estimator for $L \in \{1, 2, 3, 4\}$. Again, our results turn out to be very robust. There only appears to be a slight upward trend in the estimated persistence for larger L and the significance of the WTO effect is a little stronger for larger choices of L.

Table 6: *Logit Estimation: Average partial effects*

	Dependent variable: y_{ijt}						
	(1)	(2)	(3)		(4)	(5)	
			direct	*long-run*		*direct*	*long-run*
lagged DV	-	-	**0.331*****	-	-	**0.168*****	-
	(-)	(-)	(**0.033**)	(-)	(-)	(**0.049**)	(-)
	-	-	0.332***	-	-	0.130***	-
	(-)	(-)	(0.034)	(-)	(-)	(0.045)	(-)
log(Distance)	-	**-0.138*****	**-0.067*****	**-0.134*****	-	-	-
	(-)	(**0.014**)	(**0.007**)	(**0.013**)	(-)	(-)	(-)
	-0.140***	-0.137***	-0.067***	-0.133***	-	-	-
	(0.001)	(0.014)	(0.007)	(0.013)	(-)	(-)	(-)
Land border	-	**0.058*****	**0.016*****	**0.031*****	-	-	-
	(-)	(**0.007**)	(**0.003**)	(**0.006**)	(-)	(-)	(-)
	0.077***	0.059***	0.016***	0.032***	-	-	-
	(0.004)	(0.007)	(0.003)	(0.006)	(-)	(-)	(-)
Legal	-	**0.025*****	**0.012*****	**0.023*****	-	-	-
	(-)	(**0.003**)	(**0.001**)	(**0.003**)	(-)	(-)	(-)
	0.020***	0.025***	0.012***	0.023***	-	-	-
	(0.001)	(0.003)	(0.001)	(0.003)	(-)	(-)	(-)
Language	-	**0.069*****	**0.035*****	**0.069*****	-	-	-
	(-)	(**0.007**)	(**0.004**)	(**0.007**)	(-)	(-)	(-)
	0.078***	0.069***	0.035***	0.068***	-	-	-
	(0.001)	(0.007)	(0.004)	(0.007)	(-)	(-)	(-)
Colonial ties	-	**0.122*****	**0.069*****	**0.130*****	-	-	-
	(-)	(**0.014**)	(**0.009**)	(**0.016**)	(-)	(-)	(-)
	0.040***	0.127***	0.074***	0.136***	-	-	-
	(0.004)	(0.014)	(0.009)	(0.016)	(-)	(-)	(-)
Currency union	-	**0.104*****	**0.053*****	**0.102*****	**0.041*****	**0.027*****	**0.041*****
	(-)	(**0.011**)	(**0.006**)	(**0.011**)	(**0.013**)	(**0.009**)	(**0.014**)
	0.077***	0.104***	0.054***	0.102***	0.040***	0.028**	0.039**
	(0.003)	(0.011)	(0.006)	(0.011)	(0.014)	(0.011)	(0.015)
FTA	-	**0.098*****	**0.046*****	**0.088*****	**0.009**	**0.004**	**0.007**
	(-)	(**0.010**)	(**0.006**)	(**0.010**)	(**0.007**)	(**0.006**)	(**0.009**)
	0.110***	0.097***	0.045***	0.086***	0.008	0.003	0.004
	(0.004)	(0.011)	(0.005)	(0.010)	(0.007)	(0.006)	(0.008)
WTO	-	**0.022*****	**0.013*****	**0.026*****	**0.007****	**0.005***	**0.008****
	(-)	(**0.003**)	(**0.002**)	(**0.004**)	(**0.003**)	(**0.003**)	(**0.004**)
	0.056***	0.021***	0.013***	0.025***	0.006*	0.006*	0.008*
	(0.001)	(0.003)	(0.002)	(0.004)	(0.003)	(0.003)	(0.004)
Fixed effects	i, j, t	it, jt	it, jt		it, jt, ij	it, jt, ij	
Sample size	1204671	1204671	1171794		1204671	1171794	
Deviance	8.857×10^5	6.976×10^5	5.200×10^5		4.728×10^5	4.184×10^5	

Notes: Column (1) uncorrected average partial effects, columns (2) - (5) bias-corrected average partial effects (bold font) and uncorrected average partial effects (normal font). Column (5) bias-corrected with L = 2. Standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 7: *Probit Estimation with Different Bandwidths: Bias-Corrected Average Partial Effects*

	Dependent variable: y_{ijt}							
	$L=1$		$L=2$		$L=3$		$L=4$	
	direct	*long-run*	*direct*	*long-run*	*direct*	*long-run*	*direct*	*long-run*
lagged DV	0.174***	-	0.179***	-	0.182***	-	0.183***	-
	(0.051)	(-)	(0.052)	(-)	(0.053)	(-)	(0.053)	(-)
Currency union	0.025***	0.038***	0.024***	0.038***	0.024***	0.038***	0.025***	0.038***
	(0.009)	(0.013)	(0.009)	(0.013)	(0.009)	(0.013)	(0.009)	(0.013)
FTA	0.004	0.006	0.004	0.007	0.005	0.007	0.005	0.008
	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)	(0.009)
WTO	0.004	0.007*	0.004	0.007*	0.005*	0.007*	0.005*	0.008*
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

5.4 Comparison to Linear Probability Model

A common strategy to avoid the challenges associated with non-linear estimators in a binary choice setting is to rely on linear probability models instead. The Monte Carlo study presented in Section 4 demonstrated that this may be a far from innocent simplification. The previous subsection demonstrated that our probit estimates are very robust and therefore are a good benchmark to compare OLS estimates to in order to see whether a LPM is an appropriate alternative. Table 8 shows the LPM and probit estimates for both direct and long-run average partial effects in our preferred specification with three-way fixed effects and a lagged dependent variable. It is evident that OLS produces estimates that are far off the probit ones. OLS e.g. vastly overestimates the size of the true state dependence and ascribes a negative effect on the probability of trading for country pairs in a common FTA.

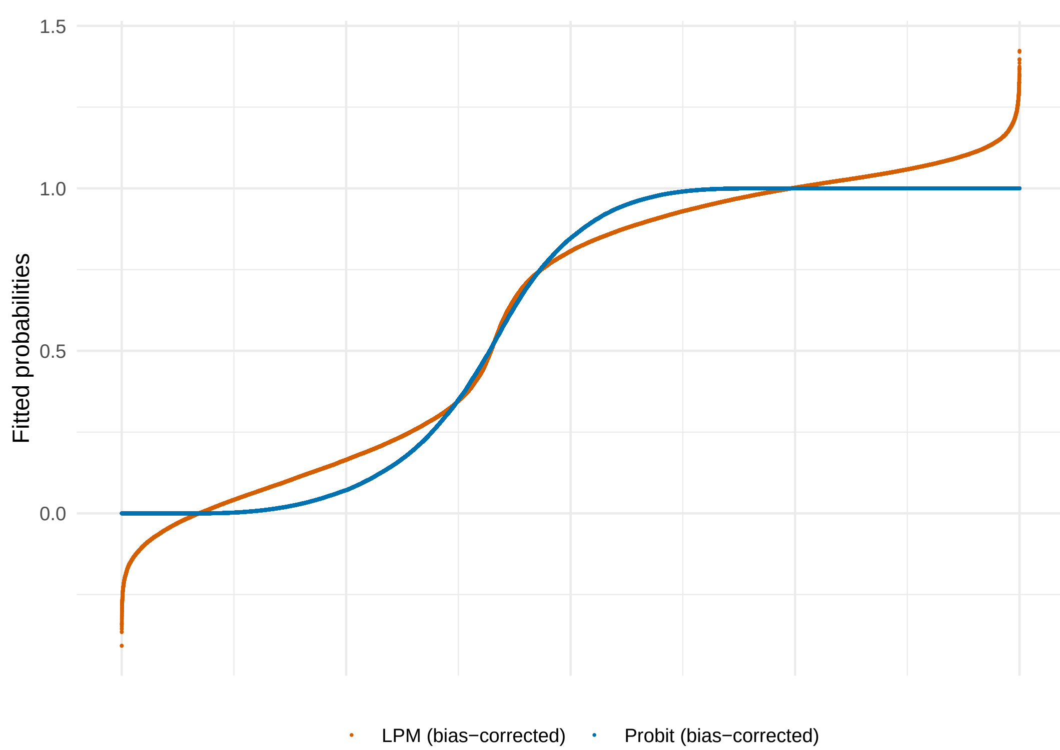
Figure 4 plots the fitted probabilities obtained with OLS and probit for our preferred specification in ascending order. It illustrates one reason for the poor performance of the LPM: in one third of the cases, it produces fitted probabilities outside of the unit interval. All partial effects depicted in Table 8 are obtained by subtracting fitted probabilities if the variable of interest is equal to zero from fitted probabilities if the variable of interest is equal to one and are hence unreasonable if fitted probabilities fall out of the unit interval. Note however, that even if the regressors of interest were continuous instead, estimated APEs based on the slope of the orange curve would not be meaningful either. As the probability of success approaches zero or one, this slope necessarily has to be close to zero — as it is in probit case, but not in the LPM case. Consider a country pair that essentially has a zero probability of trading: increasing the distance cannot make trade any more unlikely. An estimator that does not take this into account will not produce sensible average partial effect estimates. In line with our insights from the Monte Carlo analysis, the results from the application of LPM on real data lead us to a strong recommendation against its usage for the estimation of extensive margin determinants.

Table 8: *Probit vs. OLS Estimation: Average Partial Effects*

	Dependent variable: y_{ijt}			
	direct		*long-run*	
	OLS	Probit	OLS	Probit
lagged DV	0.474***	0.179***	-	-
	(0.001)	(0.052)	(-)	(-)
Currency union	0.008**	0.024***	0.015**	0.038***
	(0.003)	(0.009)	(0.006)	(0.013)
FTA	-0.062***	0.004	-0.117***	0.007
	(0.002)	(0.006)	(0.004)	(0.009)
WTO	0.008***	0.004	0.015***	0.007*
	(0.002)	(0.003)	(0.003)	(0.004)

Notes: Bias-corrected with L = 2. All columns include Origin × Year and Destination × Year fixed effects. Standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Figure 4: *Probit vs. OLS Estimation: Fitted Probabilities with Dynamic Three-way Fixed Effects*



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 dynamic 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. 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.

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.

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.

Browning, Martin and Jesus M. Carro (2010). "Heterogeneity in Dynamic Discrete Choice Models". In: *Econometrics Journal* 13.1, pp. 1–39.

Carro, Jesus M. (2003). "Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects". In: *CEMFI Working Paper* 0304.

Charbonneau, Karyne B. (Aug. 2017). "Multiple Fixed Effects in Binary Response Panel Data Models". In: *The Econometrics Journal* 20.3, S1–S13.

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.

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.

Heckman, James J. (1981). “Statistical Models for Discrete Panel Data”. In: *Structural Analysis of Discrete Data with Econometric Applications*. Ed. by C. F. Manski and Daniel McFadden. Cambridge, MA: MIT Press.

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.

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

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.

Santos Silva, J. M. C. and Silvana Tenreyro (2006). “The Log of Gravity”. In: *Review of Economics and Statistics* 88.4, pp. 641–658.

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 (2020). “Bias and Consistency in Three-way Gravity Models”. In: *Unpublished Working Paper*.

Appendix

A Stylized facts

Figure A1: *Determinants of the Extensive margin of Trade — Gravity and Persistence (1990–1991).*

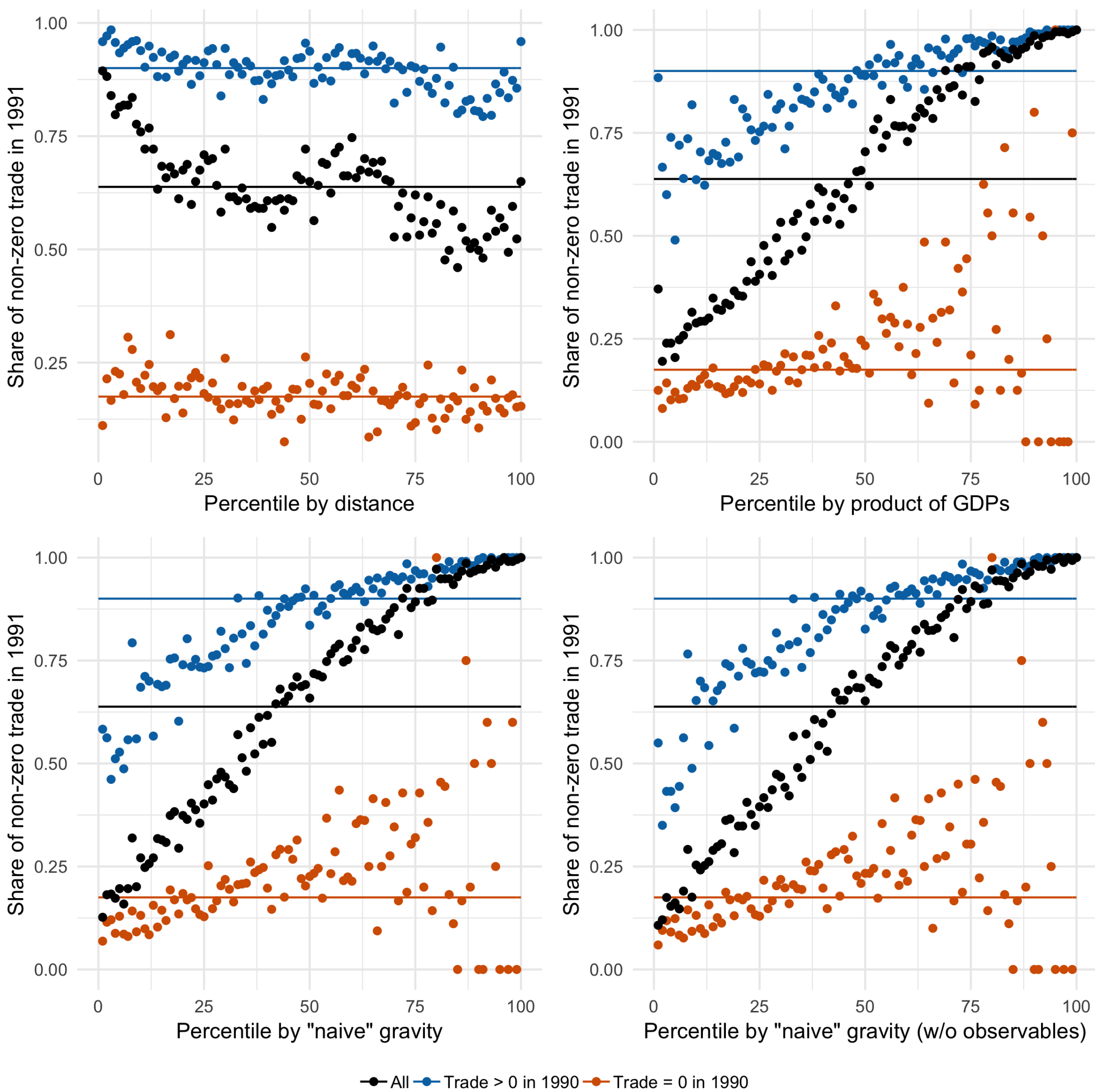
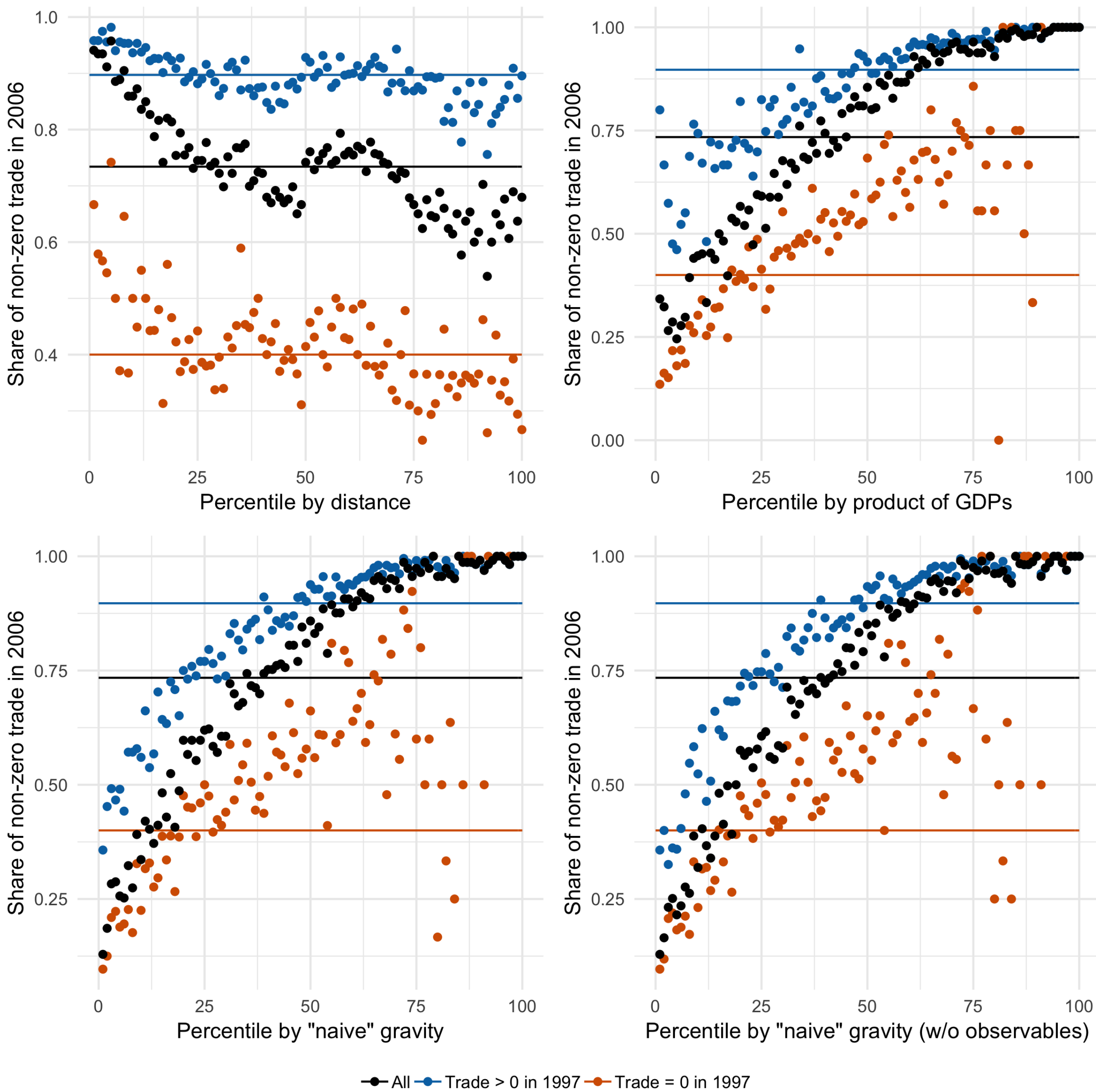


Figure A2: *Determinants of the Extensive Margin of Trade — Gravity and Persistence (1997–2006).*



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).[27] 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 (A1).

Table A1: *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.

Algorithm 1 MAP: Neumann-Halperin

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

[27] 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 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

$$y_{ijt} = \mathbf{x}'_{ijt}\boldsymbol{\beta} + \lambda_{it} + \psi_{jt} + \mu_{ij} + u_{ijt} \ . \tag{A1}$$

According to Balazsi, Matyas, and Wansbeek (2018), the appropriate within transformation corresponding to equation (A1) 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

$$\hat{\beta}_{I,J,T} = \frac{1}{IJT}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}(y_{ijt} - \bar{y}_{ij\cdot} - \bar{y}_{\cdot jt} - \bar{y}_{i\cdot t} + \bar{y}_{\cdot\cdot t} + \bar{y}_{\cdot j\cdot} + \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2 \tag{A2}$$

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 (A2) (conditional on the fixed effects) yields

$$\begin{aligned} \bar{\beta}_{I,J,T} = \mathbb{E}_{\alpha}[\hat{\beta}_{I,J,T}] &= \beta^0\left(\frac{(I-1)(J-1)(T-1)}{IJT}\right) \\ &= \beta^0\left(1 - \frac{1}{I} - \frac{1}{J} - \frac{1}{T} + \frac{1}{IT} + \frac{1}{JT} + \frac{1}{IJ} - \frac{1}{IJT}\right) \ , \end{aligned} \tag{A3}$$

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 (A3), we can form the analytically bias-corrected estimator

$$\tilde{\beta}^{a}_{I,J,T} = \hat{\beta}_{I,J,T} - \frac{\widehat{\mathbf{B}}^{\beta}_{1,I,J,T}}{I} - \frac{\widehat{\mathbf{B}}^{\beta}_{2,I,J,T}}{J} - \frac{\widehat{\mathbf{B}}^{\beta}_{3,I,J,T}}{T} \ , \tag{A4}$$

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 (A3) at costs of introducing higher order terms (see equation (A6)). Thus, we can rewrite the analytically bias-corrected estimator (A4)

$$\tilde{\beta}^{a}_{I,J,T} = \hat{\beta}_{I,J,T}\left(1 + \frac{1}{I} + \frac{1}{J} + \frac{1}{T}\right) \ . \tag{A5}$$

Taking the expectation of (A5) yields

$$
\bar{\beta}^a_{I,J,T} = \mathbb{E}_\alpha[\tilde{\beta}^a_{I,J,T}] = \beta^0 \left(1 - \frac{1}{I} - \frac{1}{J} - \frac{1}{T} + \frac{1}{IT} + \frac{1}{JT} + \frac{1}{IJ} - \frac{1}{IJT}\right)\left(1 + \frac{1}{I} + \frac{1}{J} + \frac{1}{T}\right) \tag{A6}
$$

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

Split-Panel Jackknife

As an alternative to equation (A5) 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 (A3) as we see from its expected value

$$
\bar{\beta}^{spj}_{I,J,T} = E_\phi[\hat{\beta}^{spj}_{I,J,T}] = 4\bar{\beta}_{I,J,T} - \bar{\beta}_{I/2,J,T} - \bar{\beta}_{I,J/2,T} - \bar{\beta}_{I,J,T/2} \tag{A7}
$$

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

Numerical Results

Table A2 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 A2: *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[28]

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

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

$$\begin{aligned} \bar{\beta}_{I,J,T} = \mathbb{E}_{\alpha}[\hat{\beta}_{I,J,T}] &= \beta^0 \left(\frac{(I-1)^2}{IJ} \right) \\ &= \beta^0 \left(1 - \frac{1}{I} - \frac{1}{J} + \frac{1}{IJ} \right) \; . \end{aligned} \tag{A8}$$

Analytical Bias Correction

Based on equation (A8) 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

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

[28] 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 $z_{ijt} - \bar{z}_{\cdot jt} - \bar{z}_{i \cdot t} + \bar{z}_{\cdot \cdot t}$.

Split-Panel Jackknife

A suitable split-panel jackknife estimator is

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

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 A3 demonstrate that the bias corrections work.

Table A3: *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

$$\overline{\mathbf{B}}_{1,\infty} = \text{plim}_{I,J\to\infty}\left[-\frac{1}{2J}\sum_{j=1}^{J}\frac{\sum_{i=1}^{I}\mathbb{E}_\alpha[H_{ij}\partial_{\eta^2}F_{ij}(\mathbb{M}\,\mathbf{X})_{ij}]}{\sum_{i=1}^{I}\mathbb{E}_\alpha[\omega_{ij}]}\right], \tag{A9}$$

$$\overline{\mathbf{B}}_{2,\infty} = \text{plim}_{I,J\to\infty}\left[-\frac{1}{2I}\sum_{i=1}^{I}\frac{\sum_{j=1}^{J}\mathbb{E}_\alpha[H_{ij}\partial_{\eta^2}F_{ij}(\mathbb{M}\,\mathbf{X})_{ij}]}{\sum_{j=1}^{J}\mathbb{E}_\alpha[\omega_{ij}]}\right], \tag{A10}$$

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 2. Equations (A9) and (A10) 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 $\hat{\beta}$

$$\sqrt{IJ}(\widehat{\boldsymbol{\beta}}_{I,J,T} - \boldsymbol{\beta}^0) \to_d \overline{\mathbf{W}}_\infty^{-1}\mathcal{N}(\kappa\overline{\mathbf{B}}_{1,\infty} + \kappa^{-1}\overline{\mathbf{B}}_{2,\infty}, \overline{\mathbf{W}}_\infty), \quad \text{with} \tag{A11}$$

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

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

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

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

Asymptotic distribution of $\hat{\delta}$

$$r(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta} - I^{-1}\overline{\mathbf{B}}^{\delta}_{1,\infty} - J^{-1}\overline{\mathbf{B}}^{\delta}_{2,\infty}) \to_d \mathcal{N}(0, \overline{\mathbf{V}}_{\infty}), \quad \text{with} \tag{A12}$$

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

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

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

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

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

Bias-corrected estimators

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:

$$\widehat{\boldsymbol{\beta}}^{sp} = 3\widehat{\boldsymbol{\beta}}_{I,J,T} - \widehat{\boldsymbol{\beta}}_{I/2,J,T} - \widehat{\boldsymbol{\beta}}_{I,J/2,T}, \quad \text{with} \tag{A13}$$

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

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

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 make use of the asymptotic distribution of the MLE, which we have described above. 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:

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

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

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

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

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 2.

The split-panel jackknife estimator works similarly with APEs as with structural parameters. We simply replace in formula (A13) 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:

$$
\begin{aligned}
\tilde{\boldsymbol{\delta}}^a &= \hat{\boldsymbol{\delta}} - \frac{\widehat{\mathbf{B}}_1^\delta}{I} - \frac{\widehat{\mathbf{B}}_2^\delta}{J}, \quad \text{with} \\
\widehat{\mathbf{B}}_1^\delta &= \frac{1}{2JT} \sum_{j=1}^{J} \sum_{t=1}^{T} \frac{\sum_{i=1}^{I} -\widehat{H}_{ijt} \partial_{\eta^2} \widehat{F}_{ijt} \left(\widehat{\mathbb{P}\boldsymbol{\Psi}} \right)_{ijt} + \partial_{\eta^2} \widehat{\boldsymbol{\Delta}}_{ijt}}{\sum_{i=1}^{I} \hat{\omega}_{ijt}}, \\
\widehat{\mathbf{B}}_2^\delta &= \frac{1}{2IT} \sum_{i=1}^{I} \sum_{t=1}^{T} \frac{\sum_{j=1}^{J} -\widehat{H}_{ijt} \partial_{\eta^2} \widehat{F}_{ijt} \left(\widehat{\mathbb{P}\boldsymbol{\Psi}} \right)_{ijt} + \partial_{\eta^2} \widehat{\boldsymbol{\Delta}}_{ijt}}{\sum_{j=1}^{J} \hat{\omega}_{ijt}}.
\end{aligned}
$$

The covariance can be estimated according to this simplified two-way fixed effects counterpart of equation (6) in the main text:

$$
\widehat{\mathbf{V}}^\delta = \frac{1}{I^2 J^2 T^2} \left(\left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\overline{\boldsymbol{\Delta}}}_{ijt} \right) \left(\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\overline{\boldsymbol{\Delta}}}_{ijt} \right)' + \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \widehat{\boldsymbol{\Gamma}}_{ijt} \widehat{\boldsymbol{\Gamma}}'_{ijt} \right). \tag{A14}
$$

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 $\overline{\mathbf{B}}_{\infty}^{\delta}$ of Fernández-Val and Weidner (2016) to the new structure.

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

$$
\begin{aligned}
\sqrt{IJT}(\widehat{\boldsymbol{\beta}}_{I,J,T} - \boldsymbol{\beta}^0) &\to_d \overline{\mathbf{W}}_{\infty}^{-1} \mathcal{N}(\kappa_1 \overline{\mathbf{B}}_{1,\infty} + \kappa_2 \overline{\mathbf{B}}_{2,\infty} + \kappa_3 \overline{\mathbf{B}}_{3,\infty}, \overline{\mathbf{W}}_{\infty}), \quad \text{with} \\
\overline{\mathbf{B}}_{1,\infty} &= \operatorname{plim}_{I,J,T\to\infty} \left[-\frac{1}{2JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \frac{\sum_{i=1}^{I} \mathbb{E}_{\alpha}[H_{ijt} \partial_{\eta^2} F_{ijt} (\mathbb{M}\,\mathbf{X})_{ijt}]}{\sum_{i=1}^{I} \mathbb{E}_{\alpha}[\omega_{ijt}]} \right], \\
\overline{\mathbf{B}}_{2,\infty} &= \operatorname{plim}_{I,J,T\to\infty} \left[-\frac{1}{2IT} \sum_{t=1}^{T} \sum_{i=1}^{I} \frac{\sum_{j=1}^{J} \mathbb{E}_{\alpha}[H_{ijt} \partial_{\eta^2} F_{ijt} (\mathbb{M}\,\mathbf{X})_{ijt}]}{\sum_{j=1}^{J} \mathbb{E}_{\alpha}[\omega_{ijt}]} \right], \\
\overline{\mathbf{B}}_{3,\infty} &= \operatorname{plim}_{I,J,T\to\infty} \Bigg[-\frac{1}{2IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\sum_{t=1}^{T} \mathbb{E}_{\alpha}[\omega_{ijt}] \right)^{-1} \Bigg(\sum_{t=1}^{T} \mathbb{E}_{\alpha}[H_{ijt} \partial_{\eta^2} F_{ijt} (\mathbb{M}\,\mathbf{X})_{ijt}] \\
&\qquad + 2 \sum_{\tau=t+1}^{T} \mathbb{E}_{\alpha}[H_{ijt}(Y_{ijt} - F_{ijt}) \omega_{ijt} (\mathbb{M}\,\mathbf{X})_{ijt}] \Bigg) \Bigg], \\
\overline{\mathbf{W}}_{\infty} &= \operatorname{plim}_{I,J,T\to\infty} \left[\frac{1}{IJT} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} \mathbb{E}_{\alpha}[\omega_{ijt} (\mathbb{M}\,\mathbf{X})_{ijt} (\mathbb{M}\,\mathbf{X})'_{ijt}] \right].
\end{aligned}
$$

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 $\hat{\delta}$

$$
\begin{aligned}
& 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. \\
&\qquad \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^2 J^2 T^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. \\
&\qquad \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] \\
&\quad - \mathbb{E}_\alpha \left[(\mathbb{P}\,\boldsymbol{\Psi})_{ijt}\, \partial_\eta \ell_{ijt} \right],
\end{aligned}
$$

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:

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

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:

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \,. \tag{A15}$$

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 (A15), leads to the following concentrated regression:

$$\mathbb{M}\mathbf{y} = \mathbb{M}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \,. \tag{A16}$$

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 $\hat{\epsilon}$ is the residual of (A16) and L is a bandwidth parameter.[29] This yields the bias-corrected estimator

$$\hat{\boldsymbol{\beta}} - \frac{\widehat{\mathbf{B}}}{IJT} \,, \tag{A17}$$

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

[29] The residuals of equation (A15) and equation (A16) are identical (see Gaure, 2013).

C Monte Carlo Results

C.1 Three-way Fixed Effects: Dynamic

This subsection provides the detailed results corresponding to the graphical representation and verbal discussion in Section 4 of the main text.

Table A4: *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	0.83	0.00	4	2	4	1.01	0.30
ABC (1)	-0	2	2	1.02	0.95	-1	2	2	1.07	0.93
ABC (2)	-1	2	2	1.01	0.94	-1	2	2	1.06	0.92
SPJ	-14	3	14	0.63	0.00	4	2	5	0.85	0.31
LPM (1)						0	2	2	0.93	0.93
LPM (2)						-0	2	2	0.92	0.92
					N = 50; T = 20					
MLE	16	1	16	0.90	0.00	3	1	3	0.96	0.33
ABC (1)	-0	1	1	1.02	0.95	-0	1	1	0.98	0.95
ABC (2)	-0	1	1	1.02	0.94	-0	1	1	0.98	0.95
SPJ	-5	1	5	0.84	0.03	1	1	1	0.87	0.88
LPM (1)						-0	1	1	0.91	0.92
LPM (2)						-0	1	1	0.90	0.92
					N = 50; T = 30					
MLE	12	1	12	0.92	0.00	2	1	2	1.04	0.44
ABC (1)	-0	1	1	1.01	0.95	-0	1	1	1.05	0.96
ABC (2)	-0	1	1	1.01	0.94	-0	1	1	1.05	0.96
SPJ	-3	1	3	0.93	0.16	0	1	1	0.98	0.95
LPM (1)						-0	1	1	0.94	0.94
LPM (2)						-0	1	1	0.94	0.91
					N = 50; T = 40					
MLE	10	1	10	0.92	0.00	1	1	2	1.01	0.54
ABC (1)	-0	1	1	1.00	0.94	-0	1	1	1.02	0.95
ABC (2)	-0	1	1	1.00	0.94	-0	1	1	1.01	0.94
SPJ	-2	1	2	0.93	0.26	-0	1	1	0.94	0.92
LPM (1)						-0	1	1	0.90	0.89
LPM (2)						-0	1	1	0.89	0.87
					N = 50; T = 50					
MLE	9	1	9	0.93	0.00	1	1	1	1.01	0.63
ABC (1)	-0	1	1	1.01	0.94	-0	1	1	1.01	0.95
ABC (2)	-0	1	1	1.00	0.93	-0	1	1	1.01	0.95
SPJ	-2	1	2	0.93	0.32	-0	1	1	0.97	0.94
LPM (1)						-0	1	1	0.87	0.88
LPM (2)						-0	1	1	0.87	0.85

Table A5: *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	25	1	25	0.84	0.00	4	1	4	0.99	0.02
ABC (1)	0	1	1	0.99	0.92	-0	1	1	1.02	0.92
ABC (2)	0	1	1	0.98	0.94	-1	1	1	1.01	0.90
SPJ	-8	1	9	0.66	0.00	6	1	6	0.82	0.00
LPM (1)						0	1	1	0.86	0.90
LPM (2)						-0	1	1	0.85	0.90
					N = 100; T = 20					
MLE	13	1	13	0.97	0.00	2	1	2	0.99	0.01
ABC (1)	0	1	1	1.06	0.95	0	1	1	1.01	0.96
ABC (2)	0	1	1	1.06	0.96	-0	1	1	1.01	0.96
SPJ	-3	1	3	0.91	0.01	1	1	1	0.92	0.51
LPM (1)						-0	1	1	0.89	0.92
LPM (2)						-0	1	1	0.89	0.89
					N = 100; T = 30					
MLE	9	1	9	0.94	0.00	2	0	2	0.96	0.06
ABC (1)	0	0	0	1.01	0.95	0	0	0	0.97	0.93
ABC (2)	-0	0	0	1.01	0.94	-0	0	0	0.97	0.94
SPJ	-1	1	2	0.93	0.14	0	0	1	0.93	0.86
LPM (1)						-0	0	1	0.84	0.88
LPM (2)						-0	0	1	0.84	0.79
					N = 100; T = 40					
MLE	7	0	7	0.99	0.00	1	0	1	0.97	0.10
ABC (1)	0	0	0	1.04	0.95	0	0	0	0.98	0.94
ABC (2)	-0	0	0	1.04	0.96	-0	0	0	0.98	0.95
SPJ	-1	0	1	0.94	0.34	0	0	0	0.93	0.92
LPM (1)						-0	0	0	0.81	0.82
LPM (2)						-0	0	1	0.81	0.74
					N = 100; T = 50					
MLE	6	0	6	0.94	0.00	1	0	1	0.95	0.17
ABC (1)	0	0	0	0.98	0.94	0	0	0	0.95	0.93
ABC (2)	-0	0	0	0.98	0.95	-0	0	0	0.95	0.94
SPJ	-1	0	1	0.94	0.49	0	0	0	0.93	0.94
LPM (1)						-0	0	0	0.78	0.76
LPM (2)						-0	0	1	0.78	0.68

Table A6: *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	0.85	0.00	3	1	3	1.02	0.00
ABC (1)	1	1	1	1.00	0.83	-0	1	1	1.04	0.92
ABC (2)	0	1	1	0.99	0.89	-0	1	1	1.03	0.89
SPJ	-7	1	7	0.71	0.00	6	1	6	0.88	0.00
LPM (1)						0	1	1	0.81	0.88
LPM (2)						-0	1	1	0.81	0.88
					N = 150; T = 20					
MLE	11	0	11	0.99	0.00	2	0	2	0.96	0.00
ABC (1)	0	0	0	1.07	0.90	0	0	0	0.96	0.93
ABC (2)	0	0	0	1.07	0.94	-0	0	0	0.96	0.94
SPJ	-2	0	2	0.92	0.00	1	0	1	0.87	0.16
LPM (1)						-0	0	0	0.82	0.88
LPM (2)						-0	0	0	0.82	0.82
					N = 150; T = 30					
MLE	8	0	8	0.95	0.00	2	0	2	0.95	0.00
ABC (1)	0	0	0	1.01	0.92	0	0	0	0.96	0.94
ABC (2)	0	0	0	1.00	0.95	-0	0	0	0.95	0.94
SPJ	-1	0	1	0.92	0.06	0	0	1	0.91	0.70
LPM (1)						-0	0	0	0.79	0.81
LPM (2)						-0	0	0	0.79	0.66
					N = 150; T = 40					
MLE	6	0	6	0.98	0.00	1	0	1	0.98	0.01
ABC (1)	0	0	0	1.02	0.94	0	0	0	0.98	0.94
ABC (2)	0	0	0	1.02	0.96	-0	0	0	0.98	0.94
SPJ	-1	0	1	0.96	0.22	0	0	0	0.95	0.87
LPM (1)						-0	0	0	0.78	0.70
LPM (2)						-0	0	0	0.78	0.55
					N = 150; T = 50					
MLE	5	0	5	0.97	0.00	1	0	1	0.93	0.03
ABC (1)	0	0	0	1.01	0.93	0	0	0	0.93	0.93
ABC (2)	-0	0	0	1.00	0.95	-0	0	0	0.93	0.93
SPJ	-1	0	1	0.96	0.40	0	0	0	0.90	0.90
LPM (1)						-0	0	0	0.73	0.62
LPM (2)						-0	0	0	0.73	0.46

Table A7: *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	6	62	0.94	0.00	-70	4	70	1.00	0.00
ABC (1)	-6	4	7	1.12	0.83	-7	5	8	1.09	0.76
ABC (2)	-7	5	9	1.03	0.70	-8	5	10	1.00	0.62
SPJ	24	7	25	0.75	0.01	-11	6	12	0.91	0.46
LPM (1)						2	5	5	1.00	0.92
LPM (2)						3	5	6	0.92	0.86
	N = 50; T = 20									
MLE	-27	4	27	0.94	0.00	-37	3	37	0.95	0.00
ABC (1)	-3	3	4	1.05	0.86	-3	3	5	1.00	0.83
ABC (2)	-1	3	3	1.00	0.93	-1	4	4	0.95	0.92
SPJ	5	4	6	0.88	0.71	-2	4	4	0.87	0.86
LPM (1)						8	3	9	0.96	0.28
LPM (2)						11	4	12	0.92	0.09
	N = 50; T = 30									
MLE	-16	3	16	0.99	0.00	-25	3	25	1.00	0.00
ABC (1)	-2	2	3	1.08	0.88	-2	3	3	1.04	0.87
ABC (2)	-0	3	3	1.05	0.95	-0	3	3	1.01	0.94
SPJ	2	3	3	0.97	0.89	-1	3	3	0.94	0.93
LPM (1)						10	3	11	0.99	0.02
LPM (2)						13	3	13	0.96	0.00
	N = 50; T = 40									
MLE	-10	2	11	0.96	0.01	-18	2	19	0.96	0.00
ABC (1)	-2	2	3	1.04	0.90	-2	2	3	1.00	0.88
ABC (2)	-0	2	2	1.02	0.95	0	2	2	0.97	0.94
SPJ	1	2	3	0.96	0.92	-0	3	3	0.92	0.93
LPM (1)						12	2	12	0.95	0.00
LPM (2)						13	2	14	0.93	0.00
	N = 50; T = 50									
MLE	-7	2	8	0.99	0.05	-15	2	15	1.00	0.00
ABC (1)	-2	2	2	1.06	0.90	-1	2	3	1.03	0.90
ABC (2)	-0	2	2	1.04	0.96	0	2	2	1.01	0.95
SPJ	0	2	2	0.98	0.94	-0	2	2	0.95	0.94
LPM (1)						12	2	12	0.95	0.00
LPM (2)						14	2	14	0.93	0.00

Table A8: *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	0.94	0.00	-70	2	70	1.00	0.00
ABC (1)	-6	2	7	1.09	0.23	-8	2	8	1.07	0.11
ABC (2)	-8	2	8	0.99	0.09	-9	3	10	0.97	0.05
SPJ	21	3	21	0.74	0.00	-11	3	12	0.90	0.02
LPM (1)						2	2	3	0.96	0.83
LPM (2)						4	3	5	0.88	0.65
	N = 100; T = 20									
MLE	-29	2	29	0.97	0.00	-37	2	37	0.97	0.00
ABC (1)	-3	2	4	1.05	0.43	-4	2	4	1.01	0.38
ABC (2)	-1	2	2	1.01	0.88	-2	2	2	0.97	0.84
SPJ	4	2	5	0.91	0.27	-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	0.97	0.00
ABC (1)	-3	1	3	1.03	0.53	-3	1	3	0.99	0.51
ABC (2)	-1	1	1	1.00	0.93	-1	1	2	0.97	0.92
SPJ	2	1	2	0.95	0.70	-0	1	2	0.91	0.92
LPM (1)						10	1	10	0.93	0.00
LPM (2)						13	1	13	0.91	0.00
	N = 100; T = 40									
MLE	-13	1	13	0.99	0.00	-19	1	19	0.98	0.00
ABC (1)	-2	1	2	1.04	0.57	-2	1	2	1.01	0.56
ABC (2)	-0	1	1	1.02	0.95	-0	1	1	0.99	0.94
SPJ	1	1	1	0.96	0.86	-0	1	1	0.94	0.93
LPM (1)						11	1	11	0.95	0.00
LPM (2)						13	1	13	0.93	0.00
	N = 100; T = 50									
MLE	-10	1	10	0.97	0.00	-15	1	15	0.97	0.00
ABC (1)	-2	1	2	1.01	0.64	-2	1	2	0.99	0.63
ABC (2)	-0	1	1	1.00	0.94	-0	1	1	0.97	0.94
SPJ	1	1	1	0.95	0.89	-0	1	1	0.93	0.93
LPM (1)						12	1	12	0.92	0.00
LPM (2)						14	1	14	0.91	0.00

Table A9: *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	-64	2	64	0.96	0.00	-71	1	71	1.02	0.00
ABC (1)	-7	1	7	1.11	0.00	-8	2	9	1.08	0.00
ABC (2)	-8	2	9	1.02	0.00	-10	2	10	0.99	0.00
SPJ	20	2	20	0.77	0.00	-11	2	12	0.91	0.00
LPM (1)						2	2	3	1.01	0.73
LPM (2)						3	2	4	0.93	0.43
	N = 150; T = 20									
MLE	-30	1	30	0.94	0.00	-37	1	37	0.94	0.00
ABC (1)	-4	1	4	1.01	0.07	-4	1	4	0.98	0.04
ABC (2)	-2	1	2	0.98	0.66	-2	1	2	0.94	0.59
SPJ	4	1	4	0.91	0.06	-2	1	2	0.89	0.60
LPM (1)						8	1	8	0.93	0.00
LPM (2)						11	1	11	0.89	0.00
	N = 150; T = 30									
MLE	-19	1	19	1.00	0.00	-25	1	25	0.99	0.00
ABC (1)	-3	1	3	1.06	0.12	-3	1	3	1.02	0.11
ABC (2)	-1	1	1	1.02	0.89	-1	1	1	0.99	0.86
SPJ	2	1	2	0.96	0.48	-1	1	1	0.93	0.90
LPM (1)						10	1	10	0.97	0.00
LPM (2)						13	1	13	0.94	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	0.23	-2	1	2	1.01	0.23
ABC (2)	-0	1	1	1.03	0.93	-0	1	1	0.99	0.92
SPJ	1	1	1	0.96	0.71	-0	1	1	0.93	0.92
LPM (1)						11	1	12	0.95	0.00
LPM (2)						13	1	13	0.93	0.00
	N = 150; T = 50									
MLE	-11	1	11	0.98	0.00	-15	1	15	0.98	0.00
ABC (1)	-2	1	2	1.01	0.30	-2	1	2	1.00	0.31
ABC (2)	-0	1	1	1.00	0.94	-0	1	1	0.99	0.94
SPJ	1	1	1	0.96	0.84	-0	1	1	0.95	0.94
LPM (1)						12	1	12	0.93	0.00
LPM (2)						14	1	14	0.92	0.00

C.2 Two-way fixed effects

The simulations in this section correspond to a theory-consistent estimation of the extensive margin outlined in Section 2 of the main text, 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

$$
\begin{aligned}
y_{ijt} &= \mathbf{1}[\beta_y y_{ijt-1} + \beta_x x_{ijt} + \lambda_{it} + \psi_{jt} \geq \epsilon_{ijt}] \ , \\
y_{ij0} &= \mathbf{1}[\beta_x x_{ij0} + \lambda_{i0} + \psi_{j0} \geq \epsilon_{ij0}] \ ,
\end{aligned}
$$

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 $\epsilon_{ijt} \sim \text{iid. } \mathcal{N}(0, 1)$.[30] 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 A10 – A15 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 A3 and A4 for better comprehensibility.

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

[30] 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 A3: *Dynamic: Two-way Fixed Effects – Predetermined Regressor*

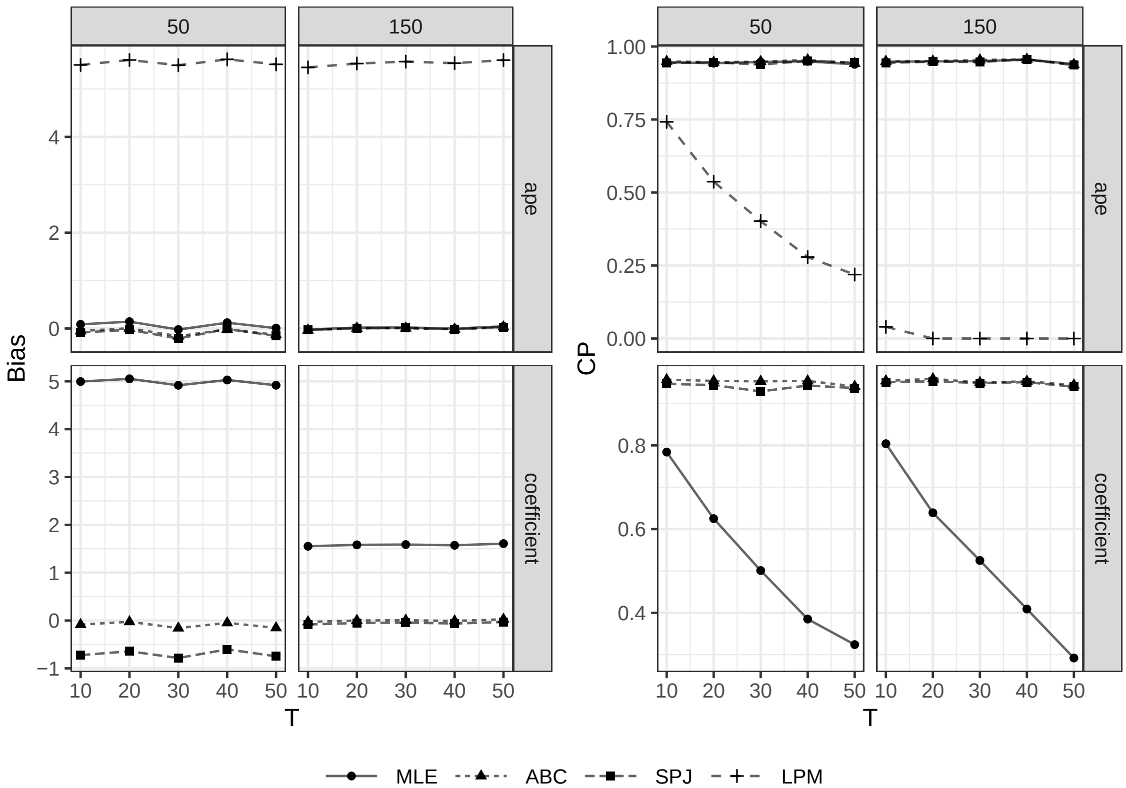
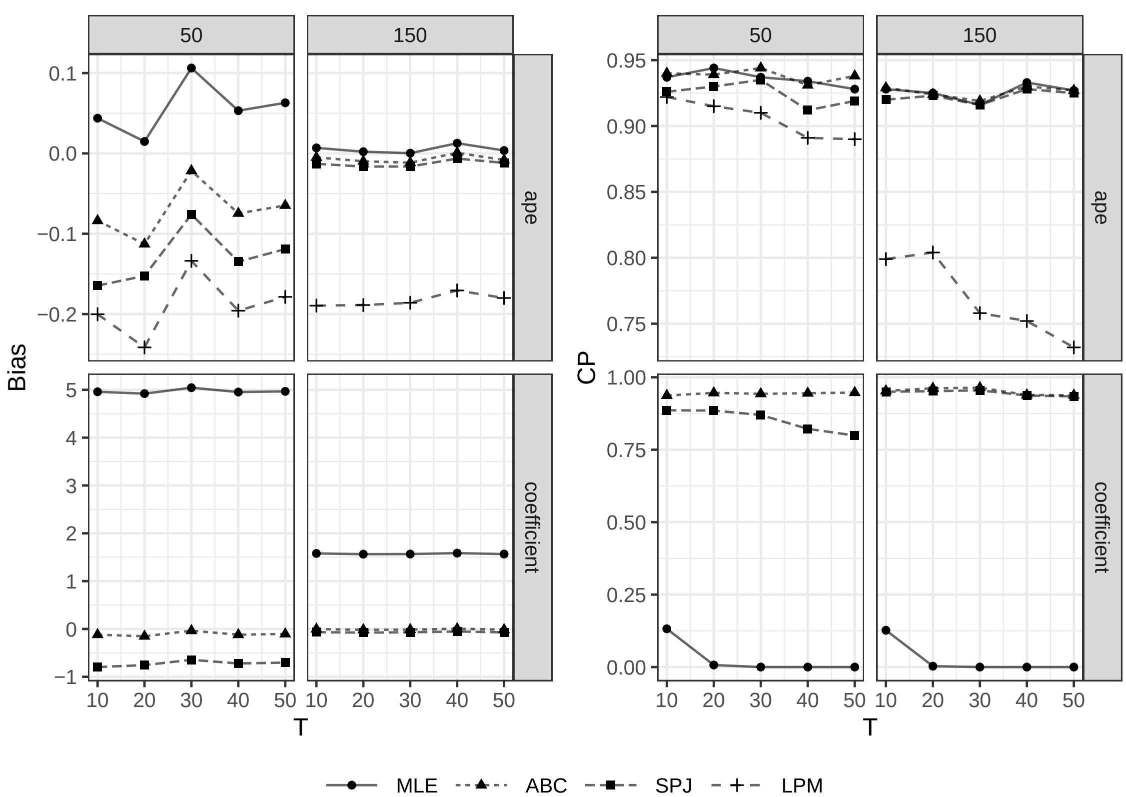


Figure A4: *Dynamic: Two-way Fixed Effects – Exogenous Regressor*



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.[31] 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.

[31] 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.

Table A10: *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	0.93	0.13	0	1	1	0.97	0.94
ABC	-0	2	2	0.97	0.94	-0	1	1	0.98	0.94
SPJ	-1	2	2	0.93	0.89	-0	1	1	0.95	0.93
LPM						-0	1	1	0.91	0.92
	N = 50; T = 20									
MLE	5	1	5	0.96	0.01	0	1	1	0.96	0.94
ABC	-0	1	1	1.00	0.95	-0	1	1	0.97	0.94
SPJ	-1	1	1	0.96	0.88	-0	1	1	0.94	0.93
LPM						-0	1	1	0.90	0.92
	N = 50; T = 30									
MLE	5	1	5	0.96	0.00	0	1	1	0.93	0.94
ABC	-0	1	1	1.00	0.94	-0	1	1	0.94	0.94
SPJ	-1	1	1	0.96	0.87	-0	1	1	0.92	0.94
LPM						-0	1	1	0.87	0.91
	N = 50; T = 40									
MLE	5	1	5	0.95	0.00	0	1	1	0.92	0.93
ABC	-0	1	1	0.99	0.94	-0	1	1	0.93	0.93
SPJ	-1	1	1	0.96	0.82	-0	1	1	0.90	0.91
LPM						-0	1	1	0.85	0.89
	N = 50; T = 50									
MLE	5	1	5	0.94	0.00	0	1	1	0.92	0.93
ABC	-0	1	1	0.98	0.95	-0	1	1	0.93	0.94
SPJ	-1	1	1	0.95	0.80	-0	1	1	0.90	0.92
LPM						-0	1	1	0.85	0.89

Table A11: *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	2	0.99	0.13	0	1	1	0.94	0.94
ABC	-0	1	1	1.01	0.95	-0	1	1	0.94	0.94
SPJ	-0	1	1	0.99	0.94	-0	1	1	0.92	0.93
LPM						-0	1	1	0.78	0.86
	N = 100; T = 20									
MLE	2	1	2	1.00	0.01	0	1	1	0.92	0.93
ABC	-0	1	1	1.02	0.95	-0	1	1	0.92	0.93
SPJ	-0	1	1	1.00	0.94	-0	1	1	0.92	0.92
LPM						-0	1	1	0.77	0.83
	N = 100; T = 30									
MLE	2	0	2	0.99	0.00	0	0	0	0.90	0.92
ABC	-0	0	0	1.01	0.96	-0	0	0	0.90	0.92
SPJ	-0	0	0	0.99	0.94	-0	0	0	0.90	0.93
LPM						-0	0	1	0.74	0.83
	N = 100; T = 40									
MLE	2	0	2	0.97	0.00	0	0	0	0.90	0.93
ABC	-0	0	0	0.99	0.94	-0	0	0	0.90	0.92
SPJ	-0	0	0	0.99	0.93	-0	0	0	0.90	0.92
LPM						-0	0	0	0.74	0.82
	N = 100; T = 50									
MLE	2	0	2	0.95	0.00	0	0	0	0.89	0.92
ABC	-0	0	0	0.97	0.94	-0	0	0	0.89	0.92
SPJ	-0	0	0	0.95	0.92	-0	0	0	0.88	0.92
LPM						-0	0	0	0.74	0.82

Table A12: *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	0	2	1.00	0.13	0	1	1	0.92	0.93
ABC	-0	0	0	1.02	0.95	-0	1	1	0.92	0.93
SPJ	-0	0	0	1.01	0.95	-0	1	1	0.92	0.92
LPM						-0	1	1	0.68	0.80
	N = 150; T = 20									
MLE	2	0	2	0.99	0.00	0	0	0	0.93	0.92
ABC	-0	0	0	1.01	0.96	-0	0	0	0.93	0.92
SPJ	-0	0	0	1.00	0.95	-0	0	0	0.92	0.92
LPM						-0	0	0	0.70	0.80
	N = 150; T = 30									
MLE	2	0	2	1.03	0.00	0	0	0	0.88	0.92
ABC	-0	0	0	1.05	0.96	-0	0	0	0.88	0.92
SPJ	-0	0	0	1.03	0.95	-0	0	0	0.88	0.92
LPM						-0	0	0	0.66	0.76
	N = 150; T = 40									
MLE	2	0	2	0.97	0.00	0	0	0	0.93	0.93
ABC	0	0	0	0.98	0.94	0	0	0	0.93	0.93
SPJ	-0	0	0	0.98	0.94	-0	0	0	0.93	0.93
LPM						-0	0	0	0.70	0.75
	N = 150; T = 50									
MLE	2	0	2	0.97	0.00	0	0	0	0.92	0.93
ABC	-0	0	0	0.99	0.94	-0	0	0	0.92	0.93
SPJ	-0	0	0	0.98	0.93	-0	0	0	0.92	0.92
LPM						-0	0	0	0.69	0.73

Table A13: *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	1.00	0.78	0	4	4	0.99	0.95
ABC	-0	4	4	1.05	0.96	-0	4	4	1.01	0.95
SPJ	-1	4	4	1.00	0.95	-0	5	5	0.97	0.94
LPM						6	4	7	0.97	0.74
	N = 50; T = 20									
MLE	5	3	6	1.00	0.62	0	3	3	0.99	0.94
ABC	-0	3	3	1.05	0.95	0	3	3	1.01	0.95
SPJ	-1	3	3	1.02	0.94	-0	3	3	0.98	0.95
LPM						6	3	6	0.98	0.54
	N = 50; T = 30									
MLE	5	3	6	0.96	0.50	-0	3	3	0.97	0.95
ABC	-0	3	3	1.00	0.95	-0	3	3	0.98	0.95
SPJ	-1	3	3	0.96	0.93	-0	3	3	0.95	0.94
LPM						5	3	6	0.94	0.40
	N = 50; T = 40									
MLE	5	2	6	0.98	0.38	0	2	2	0.99	0.95
ABC	-0	2	2	1.02	0.95	-0	2	2	1.00	0.95
SPJ	-1	2	2	1.00	0.94	-0	2	2	0.98	0.95
LPM						6	2	6	0.96	0.28
	N = 50; T = 50									
MLE	5	2	5	0.94	0.32	0	2	2	0.95	0.94
ABC	-0	2	2	0.98	0.94	-0	2	2	0.97	0.94
SPJ	-1	2	2	0.96	0.94	-0	2	2	0.94	0.94
LPM						6	2	6	0.94	0.22

Table A14: *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	0.98	0.81	-0	2	2	0.98	0.94
ABC	-0	2	2	1.00	0.95	-0	2	2	0.99	0.95
SPJ	-0	2	2	1.00	0.94	-0	2	2	0.98	0.95
LPM						5	2	6	0.95	0.29
	N = 100; T = 20									
MLE	2	2	3	0.99	0.65	-0	2	2	0.99	0.95
ABC	-0	2	2	1.01	0.95	-0	2	2	1.00	0.95
SPJ	-0	2	2	0.99	0.95	-0	2	2	0.99	0.94
LPM						5	2	6	0.96	0.06
	N = 100; T = 30									
MLE	2	1	3	0.96	0.52	-0	1	1	0.96	0.94
ABC	-0	1	1	0.98	0.95	-0	1	1	0.96	0.94
SPJ	-0	1	1	0.97	0.94	-0	1	1	0.95	0.93
LPM						6	1	6	0.92	0.02
	N = 100; T = 40									
MLE	2	1	3	1.02	0.40	0	1	1	1.00	0.95
ABC	-0	1	1	1.04	0.96	-0	1	1	1.01	0.96
SPJ	-0	1	1	1.02	0.95	-0	1	1	1.00	0.95
LPM						6	1	6	0.97	0.00
	N = 100; T = 50									
MLE	2	1	3	0.97	0.31	0	1	1	0.96	0.93
ABC	-0	1	1	0.99	0.94	-0	1	1	0.96	0.94
SPJ	-0	1	1	0.97	0.93	-0	1	1	0.95	0.93
LPM						6	1	6	0.93	0.00

Table A15: *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	0.99	0.80	-0	2	2	0.98	0.95
ABC	-0	1	1	1.00	0.95	-0	2	2	0.98	0.95
SPJ	-0	1	1	0.98	0.95	-0	2	2	0.97	0.94
LPM						5	2	6	0.93	0.04
	N = 150; T = 20									
MLE	2	1	2	1.02	0.64	0	1	1	1.00	0.95
ABC	-0	1	1	1.03	0.96	0	1	1	1.00	0.95
SPJ	-0	1	1	1.02	0.95	0	1	1	0.99	0.95
LPM						6	1	6	0.96	0.00
	N = 150; T = 30									
MLE	2	1	2	0.99	0.52	0	1	1	0.98	0.95
ABC	0	1	1	1.00	0.95	0	1	1	0.99	0.95
SPJ	-0	1	1	0.99	0.95	0	1	1	0.97	0.95
LPM						6	1	6	0.93	0.00
	N = 150; T = 40									
MLE	2	1	2	1.00	0.41	-0	1	1	0.99	0.95
ABC	-0	1	1	1.01	0.95	-0	1	1	0.99	0.96
SPJ	-0	1	1	1.00	0.95	-0	1	1	0.99	0.96
LPM						6	1	6	0.93	0.00
	N = 150; T = 50									
MLE	2	1	2	0.96	0.29	0	1	1	0.95	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

C.3 Static: Three-way Fixed Effects

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

where $\lambda_{it} \sim$ iid. $\mathcal{N}(0, 1/24)$, $\psi_{jt} \sim$ iid. $\mathcal{N}(0, 1/24)$, $\mu_{ij} \sim$ iid. $\mathcal{N}(0, 1/24)$, and $\epsilon_{ijt} \sim$ iid. $\mathcal{N}(0, 1)$. Further, $x_{ijt} = 0.5x_{ijt-1} + \lambda_{it} + \psi_{jt} + \mu_{ij} + \nu_{ijt}$, where $\nu_{ijt} \sim$ iid. $\mathcal{N}(0, 0.5)$, $x_{ij0} \sim$ 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 (see Tables A16, A17, A18). 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 A16: *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	21	2	22	0.88	0.00	1	1	2	1.03	0.91
ABC	-1	2	2	1.08	0.86	-1	1	2	1.11	0.86
SPJ	-12	2	12	0.73	0.00	-0	2	2	0.90	0.91
LPM						0	1	1	1.07	0.96
	N = 50; T = 20									
MLE	12	1	12	0.89	0.00	0	1	1	0.96	0.92
ABC	-1	1	1	0.98	0.91	-0	1	1	0.99	0.92
SPJ	-4	1	4	0.86	0.08	-1	1	1	0.88	0.85
LPM						-0	1	1	0.97	0.94
	N = 50; T = 30									
MLE	10	1	10	0.93	0.00	0	1	1	1.04	0.95
ABC	-0	1	1	1.01	0.92	-0	1	1	1.06	0.95
SPJ	-2	1	3	0.92	0.27	-0	1	1	0.97	0.90
LPM						-0	1	1	1.00	0.95
	N = 50; T = 40									
MLE	8	1	8	0.93	0.00	0	1	1	0.98	0.93
ABC	-0	1	1	0.99	0.94	-0	1	1	1.00	0.94
SPJ	-2	1	2	0.94	0.40	-0	1	1	0.95	0.90
LPM						0	1	1	0.95	0.94
	N = 50; T = 50									
MLE	8	1	8	0.91	0.00	0	1	1	1.02	0.95
ABC	-0	1	1	0.97	0.93	-0	1	1	1.04	0.95
SPJ	-1	1	2	0.90	0.48	-0	1	1	0.95	0.91
LPM						0	1	1	0.97	0.94

Table A17: *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	1.02	0.89
ABC	-1	1	1	1.04	0.84	-1	1	1	1.06	0.72
SPJ	-8	1	8	0.77	0.00	0	1	1	0.88	0.88
LPM						0	1	1	0.99	0.95
	N = 100; T = 20									
MLE	9	1	9	0.94	0.00	0	0	1	1.02	0.93
ABC	-0	1	1	1.02	0.92	-0	0	1	1.03	0.92
SPJ	-2	1	2	0.92	0.01	-0	1	1	0.96	0.92
LPM						-0	0	0	0.97	0.95
	N = 100; T = 30									
MLE	7	0	7	0.94	0.00	0	0	0	1.03	0.94
ABC	-0	0	0	1.00	0.94	-0	0	0	1.04	0.94
SPJ	-1	0	1	0.94	0.24	-0	0	0	0.97	0.92
LPM						0	0	0	0.98	0.94
	N = 100; T = 40									
MLE	6	0	6	0.95	0.00	0	0	0	1.02	0.94
ABC	-0	0	0	0.99	0.93	-0	0	0	1.03	0.95
SPJ	-1	0	1	0.94	0.46	-0	0	0	0.98	0.93
LPM						0	0	0	0.95	0.93
	N = 100; T = 50									
MLE	5	0	5	0.92	0.00	0	0	0	1.00	0.94
ABC	-0	0	0	0.96	0.92	-0	0	0	1.00	0.94
SPJ	-1	0	1	0.92	0.58	-0	0	0	0.98	0.94
LPM						0	0	0	0.91	0.92

Table A18: *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	0.90	0.00	0	0	1	1.05	0.88
ABC	-1	1	1	1.05	0.76	-1	0	1	1.08	0.50
SPJ	-7	1	7	0.78	0.00	1	1	1	0.92	0.75
LPM						-0	0	0	0.95	0.94
	N = 150; T = 20									
MLE	8	0	8	0.94	0.00	0	0	0	1.00	0.91
ABC	-0	0	0	1.00	0.90	-0	0	0	1.01	0.90
SPJ	-2	0	2	0.90	0.00	-0	0	0	0.95	0.92
LPM						0	0	0	0.93	0.92
	N = 150; T = 30									
MLE	6	0	6	0.98	0.00	0	0	0	1.03	0.94
ABC	-0	0	0	1.03	0.93	-0	0	0	1.04	0.94
SPJ	-1	0	1	0.96	0.09	-0	0	0	0.97	0.91
LPM						0	0	0	0.93	0.94
	N = 150; T = 40									
MLE	5	0	5	0.97	0.00	0	0	0	1.06	0.94
ABC	-0	0	0	1.01	0.94	-0	0	0	1.07	0.96
SPJ	-1	0	1	0.95	0.36	-0	0	0	0.99	0.93
LPM						0	0	0	0.94	0.93
	N = 150; T = 50									
MLE	4	0	4	0.99	0.00	0	0	0	0.99	0.95
ABC	-0	0	0	1.02	0.94	-0	0	0	1.00	0.94
SPJ	-0	0	0	0.98	0.51	-0	0	0	0.96	0.92
LPM						-0	0	0	0.89	0.92

D Application

Table A19: *Probit Estimation: Coefficients*

	Dependent variable: y_{ijt}				
	(1)	(2)	(3)	(4)	(5)
lagged DV	-	-	**1.664*****	-	**1.142*****
	(-)	(-)	(**0.004**)	(-)	(**0.005**)
	-	-	1.719***	-	1.057***
	(-)	(-)	(0.005)	(-)	(0.005)
log(Distance)	-	**-0.800*****	**-0.528*****	-	-
	(-)	(**0.003**)	(**0.004**)	(-)	(-)
	-0.656***	-0.821***	-0.546***	-	-
	(0.003)	(0.003)	(0.004)	(-)	(-)
Land border	-	**0.207*****	**0.118*****	-	-
	(-)	(**0.016**)	(**0.018**)	(-)	(-)
	0.260***	0.214***	0.124***	-	-
	(0.014)	(0.016)	(0.018)	(-)	(-)
Legal	-	**0.137*****	**0.089*****	-	-
	(-)	(**0.004**)	(**0.005**)	(-)	(-)
	0.090***	0.141***	0.093***	-	-
	(0.004)	(0.004)	(0.005)	(-)	(-)
Language	-	**0.426*****	**0.280*****	-	-
	(-)	(**0.006**)	(**0.007**)	(-)	(-)
	0.380***	0.436***	0.289***	-	-
	(0.005)	(0.006)	(0.007)	(-)	(-)
Colonial ties	-	**0.657*****	**0.487*****	-	-
	(-)	(**0.031**)	(**0.036**)	(-)	(-)
	0.190***	0.702***	0.542***	-	-
	(0.020)	(0.032)	(0.037)	(-)	(-)
Currency union	-	**0.631*****	**0.424*****	**0.303*****	**0.214*****
	(-)	(**0.015**)	(**0.017**)	(**0.032**)	(**0.034**)
	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*****	**0.074***	**0.038**
	(-)	(0.019)	(0.021)	(0.038)	(0.041)
	0.508***	0.552***	0.364***	0.072*	0.033
	(0.017)	(0.020)	(0.022)	(0.038)	(0.040)
WTO	-	**0.152*****	**0.101*****	**0.052*****	**0.039****
	(-)	(**0.008**)	(**0.009**)	(**0.016**)	(**0.017**)
	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
Deviance	8.891×10^5	7.019×10^5	5.183×10^5	4.76×10^5	4.189×10^5

Notes: Column (1) uncorrected coefficients, columns (2) - (5) bias-corrected coefficients (bold font) and uncorrected coefficients (standard font). Column (5) bias-corrected with L = 2. Standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A20: *Logit Estimation with Different Bandwidths: Bias-Corrected Average Partial Effects*

	Dependent variable: y_{ijt}							
	$L = 1$		$L = 2$		$L = 3$		$L = 4$	
	direct	*long-run*	*direct*	*long-run*	*direct*	*long-run*	*direct*	*long-run*
lagged DV	0.163***	-	0.168***	-	0.171***	-	0.172***	-
	(0.047)	(-)	(0.049)	(-)	(0.049)	(-)	(0.049)	(-)
Currency union	0.027***	0.041***	0.027***	0.041***	0.027***	0.041***	0.027***	0.041***
	(0.009)	(0.014)	(0.009)	(0.014)	(0.009)	(0.014)	(0.009)	(0.014)
FTA	0.004	0.007	0.004	0.007	0.005	0.008	0.005	0.008
	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)	(0.009)	(0.006)	(0.009)
WTO	0.005*	0.008**	0.005*	0.008**	0.006*	0.009**	0.006**	0.009**
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)

Notes: All columns include Origin × Year, Destination × Year and Origin × Destination fixed effects. Standard errors in parenthesis. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$