The projection approach for multi-way error components models with partially balanced panel data

Do Won Kwak[*], Kam Ki Tang[†] and Juyoung Cheong[‡]

Abstract

This paper presents a new estimation framework for partially balanced panel data models with multiple error components, which are the main source of the endogeneity problem in such models. When the dimension of error components is high, computational difficulty arises even if the data are partially balanced (i.e. balanced in some, but not all, dimensions). For linear models, our proposed projection approach can control for as many error components as the number of balanced dimensions plus one, and gives consistent estimates. For Poisson models, using correlated random effects obtained from multiple projections can also provide consistent estimates. Using Monte Carlo simulations, we confirm that the proposed estimators are consistent and produce correct inferences, with substantial power and no size bias. The proposed estimators are applied to improve the estimates of trade cost elasticity, through proper treatment of zero observations and unobserved factors. We obtain smaller trade cost elasticity estimates in both linear and Poisson models, compared to previous estimates.

Keywords: Pseudo Poisson ML; Unbalanced Panel Data; Within estimator.

[*]Asiatic Research Institute, Korea University, Seoul, Korea; e-mail: dwkwak@gmail.com
[†]School of Economics, University of Queensland, Australia; e-mail: kk.tang@uq.edu.au
[‡]Department of Economics, Kyung Hee University, Korea; e-mail: jcheong@khu.ac.kr

1 Introduction

This paper proposes two sets of computationally simple estimators for partially balanced panel data models with multi-way error components (MWEC) (i.e. unobserved factors), which are the main source of the endogeneity problem in such models. As a demonstration of the usefulness of these estimators, we apply the estimators to a gravity model of international trade, in which the data are typically unbalanced and proper control of the error components is critical.

The gravity model is the workhorse model in the empirical trade literature, and is typically formulated in either the log-linear or the exponential mean form. The log-linear specification is popular because it can easily control for MWEC – even with high dimensions – using either fixed effects (FEs), dummy variables, or a mixture of both. Country-level bilateral trade data have three dimensions: importer (i), exporter (j), and time (t). When applied to this type of data, MWEC models are often treated as having two dimensions: $p \times t$, where $p = i \times j$ is an importer-exporter pair. The p-dimensional error component can then be removed using within transformation, and the t-dimensional one can be accounted for using dummy variables. Because almost all bilateral trade data have a short time span, the small number of dummy variables used will not cause computational problems. However, this method cannot be directly extended to higher-dimensional disaggregated data (e.g. firm, industry or product-level data), as too many dummy variables would be required to account for unobserved factors. Furthermore, if the data are unbalanced – which is common with high-dimensional disaggregated data – bias arises from using multiple projections[1] to remove the error components (Wansbeek and Kapteyn, 1989; Davis, 2002; Baltagi, 2005; Balazsi et al., 2015). Conventional within transformation can only remove *one* error component without causing inconsistency/bias for unbalanced data (Balazsi et al., 2015).

Wansbeek and Kapteyn (1989) propose a generalized projection approach to MWEC models with unbalanced data. However, the proposed method requires the calculation of the generalized inverse of a matrix in which the number of both rows and columns is equal to a multiple of the number of dimensions. Such matrices are typically very large. For instance, for a moderate-sized dataset spanning over 50 importing countries, 50 exporting countries, 100 products and 5 years, the matrix would have over one million rows and over one million columns. The dataset used in the current paper is over 30 times larger in size still. The calculation of the generalized inverse of such large matrices would be computationally very challenging. Recently, Davis (2002) proposed a numerical procedure that does not require the calculation of the generalized inverse. However, this method still involves working with matrices with a large number of dimensions, and therefore remains computationally cumbersome, if not impossible. This may be the reason why this method, to our knowledge, has not been used in the empirical

[1]Projection is also referred to as demeaning or within transformation.

trade literature.

Although the within estimator for multiple error components models is biased and inconsistent for unbalanced data, it is always tempting for researchers to use it regardless, for the sake of computational convenience. Many empirical studies have also used high-dimensional FEs to account for multiple error components. Theoretically, one can apply FEs using dummy variables, but if the dimension of the FEs is high (e.g. 4 or above) and the computing capacity is ordinary, then this is not feasible in practice. In this paper, we derive a within estimator that is not only consistent, but also computationally simple to use.

High-dimensional data that are not entirely balanced, could nonetheless be partially balanced, in the sense that they are balanced within some but not all dimensions.[2] Consider the simplest example of a two-dimensional panel dataset, in which $i\,(=1,2,...m)$ denotes the individual and $t\,(=1,2,...T)$ the time period. The dataset is balanced in the i dimension if selection (i.e. observability) does not depend on i, i.e. at a given t, we either have complete observations for all i's or have no observation for any i. Similarly, the dataset is balanced in the t dimension if selection does not depend on t.

Our proposed estimators make use of the partial balancedness of data. Partial balancedness allows multiple projections to account for MWEC, without causing inconsistency/bias, under certain conditions. We show algebraically that if the data are partially balanced in h dimensions, then $h+1$ projections can remove $h+1$ error components, without causing any inconsistency, if the projections are chosen according to the balanced dimensions. Furthermore, we show, using Monte Carlo (MC) simulations, that the proposed estimators provide correct inferences with substantial power and no size bias.

Our proposed estimators have two desirable properties. Firstly, they are computationally simple and transparent in dealing with unobserved factors. This is important in pratice, because many empirical studies try to account for MWEC with FEs but cannot implement it using dummy variables or within transformation, with ordinary computing power. Secondly, and more importantly, consistent estimates can be obtained in the Pseudo Poisson Maximum Likelihood (PPML) model as well as in the log-linear model. The PPML model is one of the most common empirical formulations for the gravity model[3], after the log-linear one. For linear regression models and two-dimensional unbalanced panel data, the numerical equivalence of FEs and correlated random effects (CRE) estimators is shown in Wooldridge (2009) under strict exogeneity and in Abrevaya (2013) under sequential exogeneity. Using lower-dimensional projections of observed variables (i.e. explanatory variables) as a control function of MWEC, we can obtain the same results for the PPML model. Until now, there has been no known

[2]The concept of partial balancedness is applicable to data with at least three dimensions. Unbalancedness is not an issue for one dimensional data, while two-dimensional data are either balanced or unbalanced.

[3]For the gravity model of trade flows as the dependent variable, the Poisson specification is very common because it explicitly considers the effects of lots of zero observations.

estimator that could account for MWEC for partially balanced data in the non-linear models currently used in the empirical trade literature.

We apply the proposed estimators to estimate trade cost elasticity using both linear and Poisson models. This application makes important contributions in its own right. Ours is the first study that can comprehensively account for four distinct unobserved factors that are important in the estimation of trade cost elasticities and, thus, avoid the bias that has plagued previous studies.[4] It is also the first study that explicitly deals with zero observations for bilateral trade flows at the product level. Dealing with zero observations is an important issue when modeling product-level trade flows, because at this level of disaggregation, typically over 80 per cent of the observations are missing or recorded as zeros.[5] We also examine the consequence of not accounting for some of the unobserved factors at the product level. Empirical trade papers are increasingly interested in using more refined unobserved factors, such as firm-year and firm-buyer unobserved factors, to improve their estimates. However, as the dataset gets bigger, so too does the computational difficulty associated with handling unobserved factors. Our method is particularly useful with large datasets and multiple error components.

The rest of the paper is organized as follows. Section 2 describes our proposed estimators. Section 3 examines the finite sample properties of the estimators in MC simulation. Section 4 provides an application of the estimator to trade cost elasticity estimation. The last section concludes. The appendix provides a proof of the main result, more details of the data used in the application and further properties of the estimators in MC simulations.

2 The Multi-way Error Components Model

If the data are of d dimensions, then the maximum number of error components is d $(={}_d C_{d-1})$.[6] We can match each of these d error components with a projection with distinct dimensions. In balanced data, we can remove any lower-dimensional error components using projections as defined in the theorem by Frisch and Waugh (1933) and Lovell (1963) (the FWL theorem hereafter). But the FWL theorem cannot be extended to unbalanced data with MWEC (Wansbeek and Kapteyn, 1989; Antweiler, 2001; Baltagi et al., 2001; Davis, 2002; Baltagi, 2005; Baltagi and Song, 2006). The conventional projection method applied to unbalanced data can only remove one

[4]These unobserved factors include product-level multilateral resistance terms (MRTs), as emphasized in Anderson and Yotov (2011, 2012), and pair-product level unobserved fixed costs, which could arise from, for instance, information barriers, interest group lobbying, and government red tape, as modeled in Chaney (2008) and Krautheim (2012). Controlling for MRTs is important in practice, because elasticity estimate changes substantially with different error components and it is not possible to account for all these four error components using dummy variables or within transformation, with ordinary computing power.

[5]Anderson and van Wincoop (2004) and Broda and Weinstein (2006) illustrate that ignoring zero trade flows when estimating trade cost elasticities in the gravity equation using aggregate data can cause upward bias.

[6]For a typical panel data with cross-section and time, d is 2 and, for a typical trade data at the firm level (importer-firm-product-time) d is 4.

error component without causing bias (Matyas and Balazsi 2012; Kwak et al. 2012). However, if the data are partially balanced, our proposed projection approach can control for as many error components as the number of the balanced dimensions plus one, without causing bias. In what follows, we first discuss the approach in the context of linear models, and then in the context of Poisson models.

2.1 Linear Model

Consider MWEC as eq. (2) in the linear regression model:

$$y_{ijkt} = \mathbf{X}_{ijkt}\alpha + e_{ijkt} \tag{1}$$

$$e_{ijkt} = \omega_{ijk} + \eta_{ijt} + u_{ikt} + v_{jkt} + \varepsilon_{ijkt} \tag{2}$$

where $i\,(=1,2,...,m)$, $j\,(=1,2,...,n)$, $k\,(=1,2,...,K)$ and $t\,(=1,2,...,T)$ could, for instance, represent indexes for importers, exporters, products, and time, respectively. Besides the idiosyncratic error ε_{ijkt}, there are four distinct error components: ω_{ijk}, η_{ijt}, u_{ikt} and v_{jkt}, which are potentially the primary sources of bias.

2.1.1 Balanced data

We introduce a MWEC model using our explicit notations for selection indicators to describe the effects of missing observations in unbalanced data. We start with reviewing the within estimator for the MWEC model with balanced data.

Assumption 1. *(Strict Exogeneity of Covariates and Selection)*

$$E(\varepsilon_{ij}|\mathbf{X}_{ij}, \omega_{ijt}, \eta_{ijk}, u_{ikt}, v_{jkt}, \mathbf{S}_{ij}) = 0.$$

Here, $\mathbf{S}_{ij}$ is a vector of selection (or observability) indicators.

We denote $\varepsilon_{ij} = (\varepsilon_{ij11}, ..., \varepsilon_{ij1T}, \varepsilon_{ij21}, ..., \varepsilon_{ijk1}, ..., \varepsilon_{ijkT}, ..., \varepsilon_{ijK1}, ..., \varepsilon_{ijKT})$ and $\mathbf{S}_{ij} = (s_{ij11}, ..., s_{ij1T}, s_{ij21}, ..., s_{ijk1}, ..., s_{ijkT}, ..., s_{ijK1}, ..., s_{ijKT})$. We can similarly define $\mathbf{X}_{ij}$. For each ij pair, $\mathbf{S}_{ij}$ is allowed to be arbitrarily correlated with $\mathbf{X}_{ij}$ and unobserved factors $\omega_{ijk}, \eta_{ijt}, u_{ikt}$ and v_{jkt}, but errors ε_{ij} are uncorrelated with all elements in $\mathbf{X}_{ij}$ and $\mathbf{S}_{ij}$, conditional on the four unobserved factors.

Assumption 2. *(Full Rank)*

$$E(\tilde{\mathbf{X}}_{ijkt}^{\prime}\tilde{\mathbf{X}}_{ijkt}|\omega_{ijt},\eta_{ijk},u_{ikt},v_{jkt},s_{ijkt}=1) \textit{ is non-singular for all } i,j,k,t.$$

Under assumptions 1 and 2, it is well known that in the estimation of α in eq.(1) for balanced data, the conventional Least Squares Dummy Variables (LSDV) estimator and the correlated random effects (CRE) estimator[7] are equivalent to the pooled Least Squares (LS) estimator in the following within-transformed model:

$$\tilde{y}_{ijkt} = \tilde{\mathbf{X}}_{ijkt}\alpha + \tilde{\varepsilon}_{ijkt} \tag{3}$$

where multiple transformations from the FWL theorem, defined in eq.(4), remove all four unobserved factors in eq.(2):

$$\begin{aligned} \tilde{y}_{ijkt} \quad = \quad & y_{ijkt} - \bar{y}_{ijk\cdot} - \bar{y}_{ij\cdot t} - \bar{y}_{i\cdot kt} - \bar{y}_{\cdot jkt} \\ & + \bar{y}_{ij\cdot\cdot} + \bar{y}_{i\cdot k\cdot} + \bar{y}_{i\cdot\cdot t} + \bar{y}_{\cdot jk\cdot} + \bar{y}_{\cdot j\cdot t} + \bar{y}_{\cdot\cdot kt} \\ & - \bar{y}_{i\cdot\cdot\cdot} - \bar{y}_{\cdot j\cdot\cdot} - \bar{y}_{\cdot\cdot k\cdot} - \bar{y}_{\cdot\cdot\cdot t} + \bar{y}_{\cdot\cdot\cdot\cdot} \end{aligned} \tag{4}$$

In eq.(4), a dot in the subscript denotes the averaging dimension. For instance, $\bar{y}_{ijk\cdot} = \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}y_{ijkt}$, where s_{ijt} is an indicator for observability of both y_{ijkt} and $\mathbf{X}_{ijkt}$ (i.e. $s_{ijkt} = 1$ if both y_{ijkt} and $\mathbf{X}_{ijkt}$ are observed, and zero otherwise), and $T_{ijk} = \sum_{t=1}^{T} s_{ijkt}$ so that $T_{ijk} = T$ for balanced data and $T_{ijk} \leq T$ for unbalanced data.[8] We denote the averaging over other dimensions as follows: $\bar{y}_{i\cdot kt} = \frac{1}{n_{ikt}}\sum_{j=1}^{n} s_{ijt}y_{ijt}$, $n_{ikt} = \sum_{j=1}^{n} s_{ijkt}$, $\bar{y}_{\cdot jkt} = \frac{1}{m_{jkt}}\sum_{i=1}^{m} s_{ijkt}y_{ijkt}$ and $m_{jkt} = \sum_{i=1}^{m} s_{ijkt}$, $\bar{y}_{i\cdot k\cdot} = \frac{1}{T_{ik}}\sum_{t=1}^{T}\frac{1}{n_{ikt}}\sum_{j=1}^{n} s_{ijkt}y_{ijkt}$. We can similarly define the remaining averaging terms in eq.(4). For balanced panel data, the four error components, ω_{ijk}, η_{ijt}, u_{ikt} and v_{jkt}, can be removed by demeaning over the t, k, j and i dimensions, respectively, and the order of demeaning does not matter. Four successive demeanings applied to eq.(1) give $\tilde{y}_{ijkt}$ as the dependent variable in eq.(3), and $\tilde{\mathbf{X}}_{ijt}$ and $\tilde{\varepsilon}_{ijt}$ are obtained through a similar demeaning process. We call the estimator that uses these demeaned variables the four-way within estimator, which is consistent for balanced data.[9]

[7] The CRE estimator is also called the Mundlak estimator in Mundlak (1978).

[8] The selection indicator is very useful because it manifests the sources of bias for the conventional within estimators in unbalanced data.

[9] Under the assumption of strict exogeneity and sequential exogeneity, Wooldridge (2009) and Abrevaya (2013) both show that CRE estimators using the Mundlak and Chamberlain devices are consistent for two-dimensional unbalanced panel data. However, an extension of these estimators to higher-dimensional error components models is not straightforward, as it is not clear how to define strict exogeneity and sequential exogeneity in these cases.

2.1.2 Unbalanced data

In general, the within estimator is inconsistent when applied to unbalanced data (Wansbeek and Kapteyn, 1989; Antweiler, 2001; Baltagi et al., 2001; Davis, 2002; Baltagi, 2005; Baltagi and Song, 2006; Balazsi et al. (2015)). We reproduce this result using the observability indicator s_{ijkt} to explicitly describe the effect on the consistency of within estimator from the unbalancedness of data.

Proposition 1. *For unbalanced panel data with MWEC, only one error component can be removed by within transformation without causing any inconsistency.*[10]

Proof. We first show that one-dimensional within transformation can remove one error component completely when only using observed data. Demeaning over the t dimension can remove ω_{ijk} as follows:

$$s_{ijkt}\omega_{ijk} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}\omega_{ijk} = s_{ijkt}\omega_{ijk} - \omega_{ijk}\frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt} = s_{ijkt}\omega_{ijk} - \omega_{ijk}\frac{1}{T_{ijk}}T_{ijk} = s_{ijkt}\omega_{ijk} - \omega_{ijk} \tag{5}$$

where the first equality holds because ω_{ijk} does not depend on t. Thus, for observed values, $s_{ijkt} = 1$, and we have $s_{ijkt}\omega_{ijk} - \omega_{ijk} = 0$.

Similarly, we can remove v_{jkt} and η_{ijt} using i- and k-dimensional demeanings, respectively.

In general, MWEC can not be completely removed by multiple demeanings. Suppose we try to remove ω_{ijk} and u_{ikt} using t- and j-dimensional demeanings, respectively. Consider a demeaning sequence whereby demeanings are performed over the t dimension first, and the j dimension second. We denote the order of this sequence as 't-j'. For ω_{ijk}, the t-j demeaning order yields:

$$\begin{aligned} & s_{ijkt}\omega_{ijk} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}\omega_{ijk} - \frac{1}{n_{ikt}}\sum_{j=1}^{n}\Big(s_{ijkt}\omega_{ijk} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}\omega_{ijk}\Big) \\ = {} & s_{ijkt}\omega_{ijk} - \omega_{ijk} - \frac{1}{n_{ikt}}\sum_{j=1}^{n}(s_{ijkt}\omega_{ijk} - \omega_{ijk}) = 0 - \frac{1}{n_{ikt}}\sum_{j=1}^{n}(s_{ijkt}\omega_{ijk} - \omega_{ijk}) = 0 \qquad \text{for } s_{ijkt} = 1 \end{aligned}$$

And for u_{ikt}, it yields:

[10] Our result here is an algebraical result rather than a statistical one. As long as there are missing observations, regardless of this being random or not, the result holds. If the missing process is non-random, it will be another distinct source of bias. In this paper, we focus on dealing with the bais caused by unobserved error components.

$$
\begin{aligned}
&s_{ijkt}u_{ikt} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt} - \frac{1}{n_{ikt}}\sum_{j=1}^{n}\big(s_{ijkt}u_{ikt} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt}\big) \\
&= s_{ijkt}u_{ikt} - u_{ikt} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt} + \frac{1}{n_{ikt}}\sum_{j=1}^{n}\frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt} \\
&= 0 - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt} + \frac{1}{n_{ikt}}\sum_{j=1}^{n}\frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt} \qquad \text{for } s_{ijkt} = 1
\end{aligned}
$$

For observed data, $\frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt} \neq \frac{1}{n_{ikt}}\sum_{j=1}^{n}\frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijkt}u_{ikt}$. This is because once $s_{ijkt}u_{ikt}$ is demeaned over the t dimension, the values of $s_{ijkt}u_{ikt}$ will become $ijkt$-varying, so that the j-dimensional demeaning cannot remove u_{ikt} anymore.

We can similarly show that once we switch the order of demeanings to j-t, we can only remove u_{ikt} and not ω_{ijk}. More generally, we can show that only the *first* order demeaning (i.e. the one that comes first) can remove the corresponding error component completely when multiple demeanings are performed. □

Corollary 2. Bias due to an error component is mitigated more if the demeaning dimension that aims to remove the component is performed earlier.

Remark 3. Suppose η_{ijt} is the only unobserved factor correlated with the variable of interest. Let us consider four classes of demeaning orders over product k: $k\cdots$, $\cdot k\cdot\cdot$, $\cdot\cdot k\cdot$, and $\cdot\cdot\cdot k$, where $k\cdots$ indicates demeanings over the k dimension first (i.e. $kijt$, $kjti$, $ktji$ etc.), $\cdot k\cdot\cdot$ indicates demeanings over the k dimension second, and so forth. Within estimators from each of these demeaning orders have different magnitudes of bias due to the omitted variable η_{ijt}. We have already proven that class $k\cdots$ can completely remove the omitted variable bias (OVB) due to η_{ijt}. Simulation results in Appendix A show that the absolute magnitude of the remaining bias for the other classes increases as the demeaning order for k increases, i.e. $\cdot k\cdot\cdot$ removes the second largest portion of bias, followed by $\cdot\cdot k\cdot$, and then $\cdot\cdot\cdot k$.

Since within transformations can remove only one error component completely if it is demeaned first, and can remove other error components in subsequent demeanings only partially, the remainder of the latter error components can still be sources of inconsistency. We note, however, that the *absolute magnitude* of the bias is reduced, compared to estimates from pooled OLS.

2.1.3 Partially balanced data

Partially balancing unbalanced data is the main step we take to reduce the bias from error components. It is a useful strategy because there are practical cases where we can justify partially balancing the data even if we cannot balance the data entirely.

Definition 4. Data are partially balanced if the observability indicator depends on some but not all dimensions.

Corollary 5. *For data that are balanced in the k dimension, within transformations can remove* η_{ijt} *as well as one additional error component, using the k-dimensional demeaning. For instance, t-k demeanings can remove* ω_{ijk} *and* η_{ijt}*. Similarly, j-k and i-k demeanings can remove* u_{ikt} *and* η_{ijt}*, and* v_{jkt} *and* η_{ijt}*, respectively.*

Proof. Consider t-k demeanings:

$$\begin{aligned}
&s_{ijt}\omega_{ijk} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijt}\omega_{ijk} - \frac{1}{K_{ijt}}\sum_{k=1}^{K}(s_{ijt}\omega_{ijk} - \frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijt}\omega_{ijk})\\
&= s_{ijt}\omega_{ijk} - \omega_{ijk}\frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijt} - \frac{1}{K_{ijt}}\sum_{k=1}^{K}(s_{ijt}\omega_{ijk} - \omega_{ijk}\frac{1}{T_{ijk}}\sum_{t=1}^{T} s_{ijt})\\
&= s_{ijt}\omega_{ijk} - \omega_{ijk} - \frac{1}{K_{ijt}}\sum_{k=1}^{K}(s_{ijt}\omega_{ijk} - \omega_{ijk}) = 0 \qquad \text{for } s_{ijkt} = 1
\end{aligned}$$

And similarly, since s_{ijt} does not depend on k, and due to the fact that $\sum_{t=1}^{T} s_{ijt} = T_{ij}$, we have:

$$\begin{aligned}
&s_{ijt}\eta_{ijt} - \frac{1}{T_{ij}}\sum_{t=1}^{T} s_{ijt}\eta_{ijt} - \frac{1}{K_{ijt}}\sum_{k=1}^{K}(s_{ijt}\eta_{ijt} - \frac{1}{T_{ij}}\sum_{t=1}^{T} s_{ijt}\eta_{ijt})\\
&= s_{ijt}\eta_{ijt} - \frac{1}{T_{ij}}\sum_{t=1}^{T} s_{ijt}\eta_{ijt} - \eta_{ijt}\frac{1}{K_{ijt}}\sum_{k=1}^{K} s_{ijt} + \frac{1}{T_{ij}}\sum_{t=1}^{T}\eta_{ijt}\frac{1}{K_{ijt}}\sum_{k=1}^{K} s_{ijt}\\
&= s_{ijt}\eta_{ijt} - \frac{1}{T_{ij}}\sum_{t=1}^{T} s_{ijt}\eta_{ijt} - \eta_{ijt} + \frac{1}{T_{ij}}\sum_{t=1}^{T}\eta_{ijt} = 0 \qquad \text{for } s_{ijkt} = 1
\end{aligned}$$

□

Corollary 6. *For data that are balanced in the k and i dimensions, within transformations can remove* η_{ijt} *and* v_{jkt} *plus another error component. For instance, j-k-i demeanings can additionally remove* u_{ikt}*, and t-k-i demeanings* ω_{ijk}*. Switching the demeaning order between i and k does not change the results.*

Proof. We will show the results for j-k-i demeanings, and the insight can be extended to other orders of demeanings. Balancedness in k and i dimensions implies that $s_{ijkt} = s_{jt}$ as well as $\sum_{t=1}^{T} s_{jt} = T_j$ and $\sum_{j=1}^{n} s_{jt} = n_t$. Consider removing u_{ikt} first:

$$
\begin{aligned}
&s_{jt}u_{ikt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}u_{ikt} - \frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}u_{ikt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}u_{ikt}) \\
&-\frac{1}{n_{jt}}\sum_{i=1}^{n}[s_{jt}u_{ikt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}u_{ikt} - \frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}u_{ikt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}u_{ikt})] \\
&= s_{jt}u_{ikt} - u_{ikt} - \frac{1}{K_{jt}}\sum_{k=1}^{K} s_{jt}u_{ikt} + \frac{1}{K_{jt}}\sum_{k=1}^{K} u_{ikt}\frac{1}{n_t}\sum_{j=1}^{n} s_{jt} \\
&-\frac{1}{n_{jt}}\sum_{i=1}^{n} s_{jt}u_{ikt} + \frac{1}{n_{jt}}\sum_{i=1}^{n} u_{ikt}\frac{1}{n_t}\sum_{j=1}^{n} s_{jt} + \frac{1}{n_{jt}}\sum_{i=1}^{n}\frac{1}{K_{jt}}\sum_{k=1}^{K} s_{jt}u_{ikt} - \frac{1}{n_{jt}}\sum_{i=1}^{n}\frac{1}{K_{jt}}\sum_{k=1}^{K} u_{ikt} \\
&= s_{jt}u_{ikt} - u_{ikt} - \frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}u_{ikt} - u_{ikt}) - \frac{1}{n_{jt}}\sum_{i=1}^{n}(s_{jt}u_{ikt} - u_{ikt}) \\
&+\frac{1}{n_{jt}}\sum_{i=1}^{n}\frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}u_{ikt} - u_{ikt}) = 0 \qquad \text{for } s_{jt} = 1
\end{aligned}
$$

Similarly, we can remove η_{ijt} as follows:

$$
\begin{aligned}
&s_{jt}\eta_{ijt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}\eta_{ijt} - \frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}\eta_{ijt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}\eta_{ijt}) \\
&-\frac{1}{n_{jt}}\sum_{i=1}^{n}[s_{jt}\eta_{ijt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}\eta_{ijt} - \frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}\eta_{ijt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}\eta_{ijt})] \\
&= s_{jt}\eta_{ijt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}\eta_{ijt} - \eta_{ijt} + \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}\eta_{ijt} \\
&-\frac{1}{n_{jt}}\sum_{i=1}^{n} s_{jt}\eta_{ijt} + \frac{1}{n_{jt}}\frac{1}{n_t}\sum_{i=1}^{n}\sum_{j=1}^{n} s_{jt}\eta_{ijt} + \frac{1}{n_{jt}}\sum_{i=1}^{n}\eta_{ijt} - \frac{1}{n_{jt}}\frac{1}{n_t}\sum_{i=1}^{n}\sum_{j=1}^{n} s_{jt}\eta_{ijt} \\
&= s_{jt}\eta_{ijt} - \eta_{ijt} - \frac{1}{n_{jt}}\sum_{i=1}^{n}(s_{jt}\eta_{ijt} - \eta_{ijt}) = 0 \qquad \text{for } s_{jt} = 1
\end{aligned}
$$

And we can remove v_{jkt} as follows:

$$
\begin{aligned}
& s_{jt}v_{jkt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}v_{jkt} - \frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}v_{jkt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}v_{jkt}) \\
& -\frac{1}{n_{jt}}\sum_{i=1}^{n}[s_{jt}v_{jkt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}v_{jkt} - \frac{1}{K_{jt}}\sum_{k=1}^{K}(s_{jt}v_{jkt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}v_{jkt})] \\
& = s_{jt}v_{jkt} - \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}v_{jkt} - \frac{1}{K_{jt}}\sum_{k=1}^{K} s_{jt}v_{jkt} + \frac{1}{K_{jt}}\frac{1}{n_t}\sum_{k=1}^{K}\sum_{j=1}^{n} s_{jt}v_{jkt} \\
& -s_{jt}v_{jkt} + \frac{1}{n_t}\sum_{j=1}^{n} s_{jt}v_{jkt} + \frac{1}{K_{jt}}\sum_{k=1}^{K} s_{jt}v_{jkt} - \frac{1}{K_{jt}}\frac{1}{n_t}\sum_{k=1}^{K}\sum_{j=1}^{n} s_{jt}v_{jkt} = 0 \qquad \text{for } s_{jt} = 1.
\end{aligned}
$$

□

Proposition 7. *For partially balanced data, the number of error components that can be completely removed by within transformation (i.e. demeaning or projection)*[11] *is the number of balanced dimensions plus one. Which unobserved components can be removed without causing bias depends on which dimensions are balanced.*

Proof. See Appendix B. □

2.2 Exponential Conditional Mean Model

It is common to estimate gravity models with an exponential mean using the PPML method:

$$
y_{ijkt} = exp(\mathbf{X}_{ijkt}\alpha + \omega_{ijk} + \eta_{ijt} + u_{ikt} + v_{jkt}) \cdot \varepsilon_{ijkt} \tag{6}
$$

Our objective is to obtain a consistent estimate for α. The main problem of maximum likelihood estimation (MLE) is that each of the four error components is estimated as a parameter because they could potentially be correlated with the variable of interest in $\mathbf{X}_{ijkt}$. Therefore, it is computationally difficult to control for MWEC with dummy variables. The conditional Poisson model that uses conditioning arguments to account for MWEC in balanced data cannot be extended to unbalanced data if there is more than one error component. Appendix B provides evidence of finite sample bias based on simulation if there are multiple error components.

We suggest estimating eq.(6) using the correlated random effects (CRE) terms proposed in Wooldridge (2009). For two-dimensional panel data, Wooldridge (2009) shows the numerical equivalence of the CRE estimator to the conventional within estimator under a strict exogeneity assumption, and generalizes the results to non-linear

[11] We can always find a projection in matrix notation that is equivalent to the demeaning defined in scalar notation.

models.[12] We extend this generalization to multiple error component models with higher-dimensional, partially balanced data.[13]

2.2.1 Balanced data

Assuming the variable of interest in $\mathbf{X}_{ijkt}$ is a continuous variable, we consider a linear projection of $\mathbf{X}_{ijkt}$ to the vector space of ω_{ijk}, η_{ijt}, u_{ikt} and v_{jkt}. We denote the residuals in the null space of $\mathbf{X}_{ijkt} \cap \{\omega_{ijk}, \eta_{ijt}, u_{ikt}, v_{jkt}\}$ as $\widetilde{\mathbf{X}}_{ijkt}$. By definition of linear projection/demeaning, $\widetilde{\mathbf{X}}_{ijkt}$ is uncorrelated with all four error components, as can be seen in the following demeanings:

$$\begin{aligned}\mathbf{X}_{ijkt} = \quad & \widetilde{\mathbf{X}}_{ijkt} + \bar{\mathbf{X}}_{ijk\cdot} + \bar{\mathbf{X}}_{ij\cdot t} + \bar{\mathbf{X}}_{i\cdot kt} + \bar{\mathbf{X}}_{\cdot jkt} \\ & - \bar{\mathbf{X}}_{ij\cdot\cdot} - \bar{\mathbf{X}}_{i\cdot k\cdot} - \bar{\mathbf{X}}_{i\cdot\cdot t} - \bar{\mathbf{X}}_{\cdot jk\cdot} - \bar{\mathbf{X}}_{\cdot j\cdot t} - \bar{\mathbf{X}}_{\cdot\cdot kt} \\ & + \bar{\mathbf{X}}_{i\cdot\cdot\cdot} + \bar{\mathbf{X}}_{\cdot j\cdot\cdot} + \bar{\mathbf{X}}_{\cdot\cdot k\cdot} + \bar{\mathbf{X}}_{\cdot\cdot\cdot t} - \bar{\mathbf{X}}_{\cdot\cdot\cdot\cdot}\end{aligned} \tag{7}$$

We propose an estimator that uses all or some of the 15 mean terms in eq.(7) (i.e. all or some of the terms on the RHS except $\widetilde{\mathbf{X}}_{ijkt}$) as control variables to account for the four error components. Note that in conventional panel data with two-dimensional subscript indexes, only one mean term is available and it is known as the CRE term.

$$y_{ijkt} = exp(\mathbf{X}_{ijkt}\alpha + f(controls)) \cdot \varepsilon_{ijkt} \tag{8}$$

If the data are balanced, $f(controls)$, which includes all 15 mean terms, can account for all four error components.[14]

Proposition 8. *For balanced data, it can be shown, using the theorem by Frisch and Waugh (1933) and Lovell (1963), that the 15 mean terms in eq.(7) can be used as the basis to span the whole space of* $\mathbf{X}_{ijkt} \cap \{\omega_{ijk}, \eta_{ijt}, u_{ikt}, v_{jkt}\}$. *Thus, under a conditional exogeneity assumption, coefficient* α *estimated from the PPML method using* $E(y_{ijkt} | \mathbf{X}, f(controls)) = exp(\mathbf{X}_{ijkt}\alpha + f(controls))$ *as a conditional mean function is consistent, where* $f(controls)$ *is the linear projection of the 15 mean terms in eq.(7).*

[12] Abrevaya (2013) extends the equivalence of the CRE estimator based on the Chamberlain projection to the within estimator under a sequential exogeneity assumption.

[13] We can extend our approach using the CRE terms in partially balanced data to generally unbalanced data, but it is computationally more complicated because it requires high-dimensional interaction terms between dummy variables and the CRE terms. See Appendix C for an illustration of the estimator for generally unbalanced data.

[14] As in the linear case, each of the first four mean terms, $\bar{\mathbf{X}}_{ijk\cdot}, \bar{\mathbf{X}}_{ij\cdot t}, \bar{\mathbf{X}}_{i\cdot kt}$ and $\bar{\mathbf{X}}_{\cdot jkt}$ can account for only each of ω_{ijk}, η_{ijt}, u_{ikt} and v_{jkt}, respectively. To account for all four error components requires all 15 mean terms.

Proof. By definition of linear projection and decomposition shown for $\mathbf{X}_{ijkt}$, the residual, $\widetilde{\mathbf{X}}_{ijkt}$, is in the null space of $\mathbf{X}_{ijkt} \cap \{\omega_{ijk}, \eta_{ijt}, u_{ikt}, v_{jkt}\}$ as long as $f(controls)$ spans the same space. Thus, $\widetilde{\mathbf{X}}_{ijkt}$ is not just linearly independent of $f(controls)$, but also independent of the space of $\{\omega_{ijk}, \eta_{ijt}, u_{ikt}, v_{jkt}\}$. Therefore, control functions based on all 15 mean terms can span the space of $\mathbf{X}_{ijkt} \cap \{\omega_{ijk}, \eta_{ijt}, u_{ikt}, v_{jkt}\}$. This implies that:

$$cov(\mathbf{X}_{ijkt}, \omega_{ijk} + \eta_{ijt} + u_{ikt} + v_{jkt} | f(controls)) = 0 \tag{9}$$

So the PPML estimator for α applied to $E(y_{ijkt}|\mathbf{X}, f(controls)) = exp(\mathbf{X}_{ijkt}\alpha + f(controls))$ as the conditional mean is consistent. Our simulation results in Section 3 also confirm this. □

2.2.2 Unbalanced data

For unbalanced data, we show in the simulation that the control function using a subset or the full set of the 15 mean terms cannot provide a consistent estimator. The reason that Proposition 8 cannot be extended to unbalanced data is the same as in the linear case. Simulation results show that the CRE terms can only control one error component completely. Here, eq.(9) becomes:

$$cov(s_{ijkt}\mathbf{X}_{ijkt}, s_{ijkt}\omega_{ijk} + s_{ijkt}\eta_{ijt} + s_{ijkt}u_{ikt} + s_{ijkt}v_{jkt} | f(s_{ijkt} \cdot controls)) \neq 0,$$

or:

$$cov(\mathbf{X}_{ijkt}, \omega_{ijk} + \eta_{ijt} + u_{ikt} + v_{jkt} | f(controls), s_{ijkt} = 1) \neq 0.$$

2.2.3 Partially balanced data

Using a similar idea as in the linear model, simulations show that we can account for more than one error component in partially balanced data, without causing inconsistency/bias. Balancedness in the k and i dimensions implies that $s_{ijkt} = s_{jt}$, and demeanings over i and k do not affect s_{jt}. So we have:

$$cov(\mathbf{X}_{ijkt}, \eta_{ijt} + u_{ikt} + v_{jkt} | f(controls)) = cov(\mathbf{X}_{ijkt}, \eta_{ijt} + u_{ikt} + v_{jkt} | f(controls), s_{jt} = 1) = 0$$

Here the control function using the mean values over k, i and j (i.e. $\bar{\mathbf{X}}_{ij\cdot t}, \bar{\mathbf{X}}_{i\cdot kt}, \bar{\mathbf{X}}_{\cdot jkt}$, $\bar{\mathbf{X}}_{\cdot\cdot kt}, \bar{\mathbf{X}}_{\cdot j\cdot t}, \bar{\mathbf{X}}_{i\cdot\cdot t}$, and $\bar{\mathbf{X}}_{\cdots t}$) can remove partial correlation between $\mathbf{X}_{ijkt}$ and error components η_{ijt}, v_{jkt} and u_{ikt}. This implies conditional mean independence among $\{\mathbf{X}_{ijkt}, \eta_{ijt} + u_{ikt} + v_{jkt} | \ f(controls), s_{jt} = 1\}$.

Thus, for data that are balanced in dimensions i and k only, with a properly chosen $f(controls)$ we can have:

$$cov(\mathbf{X}_{ijkt}, \eta_{ijt} + u_{ikt} + v_{jkt} | f(controls), s_{jt} = 1) = 0$$

Using seven mean terms in the control function, we can obtain a consistent PPML estimator even when $\mathbf{X}_{ijkt}$ is correlated with the error components. The control function consists of the CRE terms that result from a linear projection of $\mathbf{X}_{ijkt}$ on the space of $\{u_{ikt}, v_{jkt}, \eta_{ijt}\}$, as follows:

$$\{\bar{\mathbf{X}}_{ij\cdot t}, \bar{\mathbf{X}}_{i\cdot kt}, \bar{\mathbf{X}}_{\cdot jkt}, \bar{\mathbf{X}}_{i\cdot\cdot t}, \bar{\mathbf{X}}_{\cdot j\cdot t}, \bar{\mathbf{X}}_{\cdot\cdot kt}, \bar{\mathbf{X}}_{\cdot\cdot\cdot t}\}.$$

Corollary 9. *For unbalanced data, we can control one error component using the CRE terms, as the conditional PPML can account for one error component using the sum of the dependent variable. (E.g. we can use the sum of the dependent variable over product dimension, k, to account for the error component, η_{ijt}.) For partially balanced data, using the CRE terms, we can account for as many error components as the number of partially balanced dimensions plus one.*

Therefore, for data that are balanced in, say, the i and k dimensions, linear projection of $\mathbf{X}_{ijkt}$ on the space of $\{\eta_{ijt}, u_{ikt}, v_{jkt}\}$ allows us to isolate a valid CRE term to account for endogeneity from η_{ijt}, u_{ikt} and v_{jkt}. Our simulation results show that the PPML estimator with the CRE terms in a proper control function for partially balanced data is consistent.

3 Monte Carlo Simulation

In this section, we examine the finite sample properties of the proposed within estimators. The insights from three-way error component (3WEC) and four-way error component (4WEC) models are essentially the same, but computational complexity increases very quickly with the number of dimensions and error components. Therefore, we focus on 3WEC models to confirm the consistency of the proposed estimators for partially balanced data. We also quantify finite sample bias in parameter estimates and inferences of these estimators for both unbalanced and partially balanced data.

3.1 Monte Carlo Design: 3WEC Model

We consider the following model for all simulations:

$$y_{ijt} = \alpha x_{1ijt} + u_{it} + v_{jt} + \omega_{ij} + \varepsilon_{ijt} \tag{10}$$

where x_{1ijt} is highly correlated with u_{it}, v_{jt} and ω_{ij}, and $\alpha = 1$. Our main interest is to obtain a consistent estimate of α .

Our data generating process (DGP) is based on:

$$x_{1ijt} = [a \cdot u_{it} + b \cdot v_{jt} + c \cdot \omega_{ij} + d \cdot N(0,1)] \tag{11}$$

where a, b, c can take a value of either 0 or 1, where a value of 1 causes correlation between x_1 and the corresponding error components; $u_{it} \sim iidN(0,1)$; $v_{jt} \sim iidN(0,1)$; $\omega_{ij} \sim iidN(0,1)$; idiosyncratic error $\varepsilon_{ijt} \sim iidN(0,1)$; d is chosen to determine the size of bias; and for the baseline design, the number of observations is 2,000 with $i = 1,...,20$, $j = 1,...,20$ and $t = 1,...,5$.

For the missing observation generation process, we first draw 400 observations for the pair $ij = 1,2,...,400$ from:

$$z_{ij} \sim uniform[0,1] \tag{12}$$

and drop y_{ijt} if $z_{ij} \leq 0.25$. Second, for $z_{ij} > 0.25$, we draw 1,500 observations from:

$$z_{2ijt} \sim uniform[0,1] \tag{13}$$

and drop y_{ijt} if $z_{2ijt} \leq 1/3$. Therefore, we end up with 1,000 observations, of which 50 per cent are missing. The missing observation generation process is completely random, but MWECs are correlated with x_1. The within estimators will still be inconsistent.

We define observability as $s_{ijt} = 1$ if $z_{ij} \geq 0.25$ and $z_{2ijt} > 1/3$, and $s_{ijt} = 0$ otherwise. We construct the data to be partially balanced in that it is balanced only in the t-dimension, by setting $s^p_{ijt} = 1$ if $z_{ij} \leq 0.25$ and $s^p_{ijt} = 0$ otherwise. Thus, for fixed t, we either have observations of y and x for all ij or have no observation for any ij.

We consider the following DGPs with seven sets of values for (a, b, c): (i) (1,0,0), (ii) (0,1,0), (iii) (0,0,1), (iv) (1,1,0), (v) (1,0,1), (vi) (0,1,1) and (vii) (1,1,1). For unbalanced panel data, we can obtain consistent estimators for (i), (ii) and (iii), where only one error component is the source of endogeneity. For (iv), (v) and (vi), the bias of within estimator is caused by two error components; while for (vii), the bias is caused by all three error components.

Given x_{1ijt}, u_{it}, v_{jt}, ω_{ij} and ε_{ijt}, and setting the true parameter value of α to be equal to 1, we can obtain y_{ijt} in eq.(10). In the simulation, we first quantify the finite sample bias of the within estimators. The bias can be expressed according to its three distinct error components sources in eq.(14). We show which dimensional projections/demeanings can remove the bias from each of these distinct error components sources and give consistent estimation and inference.

It should be pointed out that although we consider a model with only one observed covariate, x_{1ijt}, in the simulation, the analysis remains the same when there are other observed covariates. In the latter case, the model would be:

$$y_{ijt} = \alpha x_{1ijt} + \mathbf{X}_{2ijt}\beta + u_{it} + v_{jt} + \omega_{ij} + \varepsilon_{ijt}$$

We can then rewrite the pooled OLS estimator for α as:

$$plim\hat{\alpha} = \alpha + \underbrace{\frac{cov(x^*_{1ijt}, u^*_{it})}{var(x^*_{1ijt})}}_{\text{bias due to } u_{it}} + \underbrace{\frac{cov(x^*_{1ijt}, v^*_{jt})}{var(x^*_{1ijt})}}_{\text{bias due to } v_{jt}} + \underbrace{\frac{cov(x^*_{1ijt}, \omega^*_{ij})}{var(x^*_{1ijt})}}_{\text{bias due to } \omega_{ij}} \tag{14}$$

where x^*_{1ijt} is the residual from the regression of x_{1ijt} on all the other observed covariates, $\mathbf{X}_{2ijt}$, and, similarly, u^*_{it}, v^*_{jt}, and ω^*_{ij} are the residuals from the respective regressions of these variables on $\mathbf{X}_{2ijt}$.

3.2 Simulation Results: Linear Model

3.2.1 Finite sample mean bias

Table 1 reports the finite sample behavior of the within estimators. The results in the upper panel are based on unbalanced data, and those in the lower panel are based on data partially balanced in the t dimension. The "Sources of biases" column indicates the contribution of each error component to total bias, the rest are the remaining biases after each of the transformations. For instance, the first row of of the upper panel shows that u_{it} causes the entire 70 per cent bias for OLS estimation, but the bias is reduced to 18 per cent after the i-t-j transformation, and 0 per cent after the j-t-i transformation, and so forth. Table 2 repeats some of the simulations, but with different orders of transformation.

The following observations can be drawn from the two tables:

(i) As long as the directions of biases due to u_{it}, v_{jt} and ω_{ij} are the same, all within estimators reported in the two tables reduce the bias significantly compared to pooled OLS.[15]

(ii) For unbalanced data, it is only possible to completely eliminate a single-sourced bias if an appropriate within estimator is used. For instance, in the first row of the upper panel of Table 1, where the bias is entirely due

[15] If the directions of biases due to u_{it}, v_{jt} and ω_{ij} are not the same, biases of the OLS estimator caused by each of the error components could at least partially cancel each other out. So we cannot say that the within estimator always has a smaller bias than the pooled OLS estimator.

to u_{it}, performing the j-dimensional demeaning first can eliminate the bias completely.

(iii) When there are two or more sources of bias for unbalanced data, no demeaning sequence can eliminate the bias completely.

(iv) For data partially balanced in the t dimension, it is only possible to eliminate bias due to ω_{ij} and one more source of error components completely if an appropriate demeaning is used. For instance, in the fifth row of the lower panel of Table 1, where the bias is entirely due to u_{it} and ω_{ij}, the j-t-i demeaning can reduce the bias to zero.

(v) When there are three sources of bias, and the data are balanced in only one dimension, no demeaning sequence can eliminate the bias completely.

(vi) The magnitude of absolute bias depends upon the proportion of missing data: if the proportion is very small, then the remaining bias will be very small after multiple demeanings of any order. This implies that, in practice, for unbalanced data, a within estimator from multiple projections is inconsistent, but the magnitude of inconsistency could be very small as long as the missing proportion is small.

3.2.2 Inference: size bias and power

Table 3 reports the inference consistency based on the t-test under the null hypothesis that the within estimator is consistent. For the pooled OLS estimator, the bias due to each of the error components (i.e. unobserved heterogeneity) is provided in the "bias" column of Table 3. In the first row the bias is 0.66, as the mean of the OLS estimate for α is 1.66, while its true value is 1. The column headings "ba", "ub" and "pb" stand for balanced, unbalanced and partially balanced data, respectively, followed by the order of demeanings. For instance, the "ba: i-t-j" column reports the result for balanced data with the i-t-j demeanings. The mean value of the rejection probability for the t-test of $H_0 : \alpha = 1$ is reported in the third row, and its coverage values are reported in the fourth row.

The following observations can be drawn from the table:

(i) If a within estimator completely eliminates the bias due to unobserved heterogeneity for partially balanced data, the within estimator has no finite sample bias for a usual t-test with an appropriate degrees of freedom adjustment, so that this t-test provides a valid inference. In the "pb: i-t-j" and "pb: i-j-t" columns with non-zero v_{jt} (accounted for by the first-order i-dimensional demeaning) and ω_{ij} (accounted for by the non-first-order demeaning over the t dimension, which is balanced), the mean of the bias is zero. The mean rejection probability for a t-test is 0.054 for both cases, and its coverage contains true type-I error of 0.05. In the "pb: j-t-i" column with non-zero u_{it} and ω_{ij}, the mean rejection probability for a t-test is 0.053 and its coverage contains true type-I error of 0.05.

(ii) If all of the error components cannot be removed, the proposed estimator for α is biased/inconsistent but still has substantial power. For instance, in the "ub: i-t-j" and "ub: t-i-j" columns, the rejection probability is very close to 1 and the bias is only about 0.1 to 0.2. In the case of "ub: i-t-j", for a DGP of $(0, v_{jt}, \omega_{ij})$ the bias is as small as 0.065, but the power of a t-test is 0.319, while for a DGP of $(u_{it}, v_{jt}, 0)$ the bias increases slightly to 0.71 and the power also increases to 0.917.

Table 1: Linear regression: Tracing source biases- 50% random missing

	Source of bias				The order of demean-dimension for within estimators					
	$u_{it}+v_{jt}+\omega_{ij}$	u_{it}	v_{jt}	ω_{ij}	*i*	*i-t*	*i-t-j*	*j*	*j-t*	*j-t-i*
Unbalanced data for which all three dimensions are unbalanced										
$(u_{it},0,0)$	70%	70%	0%	0%	70.0%	70.1%	18%	0.0%	0.0%	0.0%
$(0,v_{jt},0)$	70%	0%	70%	0%	0.0%	0.0%	0.0%	70.8%	70.7%	18.1%
$(0,0,\omega_{ij})$	70%	0%	0%	70%	70.8%	2.4%	2.4%	70.6%	2.4%	2.4%
$(u_{it},v_{jt},0)$	66%	33%	33%	0%	40.0%	40.0%	7.1%	40.2%	40.1%	7.1%
$(u_{it},0,\omega_{ij})$	66%	33%	0%	33%	66.6%	40.4%	7.8%	80.0%	6.5%	6.5%
$(0,v_{jt},\omega_{ij})$	100%	0%	33%	67%	80.0%	6.5%	6.5%	100.1%	100.1%	29.8 %
$(u_{it},v_{jt},\omega_{ij})$	90%	30%	30%	30%	88.8%	80.0%	39.3%	88.9%	80.1%	39.4%
Partially balanced data for which the *t*-dimension is balanced										
$(u_{it},0,0)$	70%	70%	0%	0%	70.5%	70.5%	1.2%	0.0%	0.0%	0.0%
$(0,v_{jt},0)$	70%	0%	70%	0%	0.0%	0.0%	0.0%	70.7%	70.6%	1.2%
$(0,0,\omega_{ij})$	70%	0%	0%	70%	70.7%	0.0%	0.0%	70.6%	0.0%	0.0%
$(u_{it},v_{jt},0)$	66%	33%	33%	0%	40.0%	40.1%	0.5%	40.1%	40.1%	0.4%
$(u_{it},0,\omega_{ij})$	66%	33%	0%	33%	66.6%	39.9%	0.5%	80.0%	0.1%	0.0%
$(0,v_{jt},\omega_{ij})$	100%	0%	33%	67%	80.0%	0.1%	0.0%	100.0%	100.0%	1.6 %
$(u_{it},v_{jt},\omega_{ij})$	90%	30%	30%	30%	88.8%	79.8%	3.2%	88.9%	79.9%	3.2%

Note: The number of replications is 1,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_p, the second element is v_p and the third element is ω_p. u_{it}, v_{jt}, and ω_{ij} are sources of bias, as in equation (14).

Table 2: Linear regression: Tracing source biases- 50% random missing

	Source of bias				The order of demean-dimension for within estimators					
	$u_{it}+v_{jt}+\omega_{ij}$	u_{it}	v_{jt}	ω_{ij}	*t*	*t-i*	*t-i-j*	*i*	*i-j*	*i-j-t*
Unbalanced data for which all three dimensions are unbalanced										
$(u_{it},v_{jt},0)$	66%	33%	33%	0%	66.7%	43.9%	12.7%	50.0%	2.5%	2.5%
$(u_{it},0,\omega_{ij})$	66%	33%	0%	33%	40.0%	40.0%	7.1%	66.6%	50.6%	5.6%
$(0,v_{jt},\omega_{ij})$	100%	0%	33%	66%	99.9%	22.1%	22.1%	50.0%	50.1%	3.3 %
$(u_{it},v_{jt},\omega_{ij})$	90%	30%	30%	30%	66.7%	43.8%	12.6%	66.6%	40.9%	3.0%
Partially balanced data for which the *t*-dimension is balanced										
$(u_{it},v_{jt},0)$	66%	33%	33%	0%	66.7%	39.9%	0.4%	50.0%	0.9%	0.9%
$(u_{it},0,\omega_{ij})$	66%	33%	0%	33%	40.0%	40.0%	0.4%	66.6%	50.2%	0.9%
$(0,v_{jt},\omega_{ij})$	100%	0%	33%	67%	99.9%	0.1%	0.2%	50.0%	50.0%	0.0%
$(u_{it},v_{jt},\omega_{ij})$	90%	30%	30%	30%	66.6%	39.8%	0.3%	66.6%	40.3%	0.5%

Note: The number of replications is 1,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_p, the second element is v_p and the third element is ω_p. u_{it}, v_{jt}, and ω_{ij} are sources of bias, as in equation (14).

Table 3: Linear regression: Inference accuracy

	Bias	ba: i-t-j	ub: i-t-j	pb: i-t-j	ub: j-t-i	pb: j-t-i	ub: t-i-j	pb: t-i-j	ub: i-j-t	pb: i-j-t
(u_{it},v_{jt},0), mean of bias	.66	.00	.071	.005	.071	.004	.127	.004	.025	.009
(u_{it},v_{jt},0), SD	(.035)	(.013)	(.021)	(.015)	(.021)	(.015)	(.022)	(.015)	(.020)	(.015)
(u_{it},v_{jt},0), rejection% mean		.053	.917	.031	.913	.043	1	.044	.180	.060
(u_{it},v_{jt},0), rejection% coverage		(.039,.067)	(.900,.934)	(.020,.042)	(.896,.930)	(.030,.056)	(1,1)	(.031,.057)	(.156,.204)	(.045,.075)
(u_{it},0,ω_{ij}), mean of bias	.66	.00	.078	.005	.065	.000	.071	.004	.056	.009
(u_{it},0,ω_{ij}), SD	(.028)	(.013)	(.021)	(.015)	(.040)	(.030)	(.022)	(.015)	(.020)	(.015)
(u_{it},0,ω_{ij}), rejection% mean		.044	.960	.044	.410	.053	.907	.044	.774	.067
(u_{it},0,ω_{ij}), rejection% coverage		(.031,.057)	(.948,.972)	(.031,.057)	(.397,.441)	(.039,.067)	(.889,.925)	(.031,.057)	(.748,.800)	(.051,.083)
(0,v_{jt},ω_{ij}), mean of bias	1.00	.00	.065	.000	.298	.016	.220	.001	.065	.000
(0,v_{jt},ω_{ij}), SD	(.020)	(.013)	(.053)	(.030)	(.043)	(.031)	(.043)	(.029)	(.040)	(.030)
(0,v_{jt},ω_{ij}), rejection% mean		.052	.319	.054	1	.051	1	.038	.414	.054
(0,v_{jt},ω_{ij}), rejection% coverage		(.038,.066)	(.290,.348)	(.040,.068)	(1,1)	(.037,.065)	(1,1)	(.026,.050)	(.383,.445)	(.040,.068)

Note: The number of replications is 1,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_{it}, the second element is v_{jt} and the third element is ω_{ij}. The ba-, ub-, and pb- columns represent the estimates of the within estimators for balanced, unbalanced and partially balanced data, respectively.

3.3 Simulation Results: Poisson MLE Model

In the case of PPML models, the outcome variable takes exponential values:

$$y_{ijt} = exp(\alpha x_{1ijt} + u_{it} + v_{jt} + \omega_{ij}) \cdot exp(\varepsilon_{ijt}) \tag{15}$$

Apart from this respect, the rest of the simulation design, including the DGP, is the same as in the linear model case.

Results are reported in Tables 4 and 5, where the first column indicates the magnitude of the composite bias obtained from pooled PPML with no CRE term. We can draw the following observations:

(i) As long as the directions of biases due to u_{it}, v_{jt} and ω_{ij} are the same, including the CRE term reduces the bias significantly in both unbalanced and partially balanced data.

(ii) For unbalanced data, it is only possible to eliminate a single-sourced bias almost completely if an appropriate CRE term is used. For instance, in the second row of the upper panel of Table 4, where the bias is entirely due to v_{jt}, performing the i-dimensional demeaning first can reduce the bias down to a few percentage points (i.e. between 2.8 per cent and 3.1 per cent).

(iii) When there are two or more sources of bias for unbalanced data, no CRE terms can come close to eliminating the bias completely.

(iv) For data that are partially balanced in the t dimension, it is only possible to eliminate the bias due to ω_{ij} and one other source almost completely if an appropriate CRE term is used. For instance, in the second row of the lower panel of Table 5, where the bias is entirely due to u_{it} and ω_{ij}, the j-t-i demeaning can reduce the bias to just 1.4 per cent.

(v) When there are three or more sources of bias for data balanced in only one dimension, no demeaning sequence can eliminate the bias completely.

Table 6 provides inference consistency based on the t-test under the null hypothesis that the PPML estimator with a control function is consistent. The bias due to unobserved heterogeneity for the PPML estimator is provided in the first column and is about 0.94. The mean value of the rejection probability for the t-test of $H_0 : \alpha = 1$ is reported in the third row, its coverage values in the fourth row, and the results from using bootstrapped standard errors in the fifth and sixth rows.

The results show that, for partially balanced data, a PPML estimator for α with a proper CRE term has no finite sample bias, and a usual t-test with an appropriate degrees of freedom adjustment can provide a valid inference. We use bootstrapped standard errors to adjust the degrees of freedom. The estimated mean value of the rejection probability for the t-test is very close to 0.05 and its coverage always contains true type-I error of 0.05, as in the linear model case.

Table 4: PPML: Tracing source biases- 50% random missing

	Source of bias				The order of demean-dimension for within estimators					
	$u_{it}+v_{jt}+\omega_{ij}$	u_{it}	v_{jt}	ω_{ij}	t	t-i	t-i-j	i	i-j	i-j-t
Unbalanced data for which all three dimensions are unbalanced										
(u_{it},0,0)	70%	70%	0%	0%	65.0%	64.9%	7.9%	42.5%	2.8%	3.3%
(0,v_{jt},0)	70%	0%	70%	0%	64.7%	7.7%	7.8%	2.8%	2.4%	3.1%
(0,0,ω_{ij})	70%	0%	0%	70%	1.7%	1.5%	1.6%	44.2%	44.0%	0.6%
(u_{it},v_{jt},0)	97%	48%	49%	0%	94.8%	92.0%	57.3%	88.7%	55.8%	55.1%
(u_{it},0,ω_{ij})	97%	48%	0%	49%	90.5%	91.7%	72.4%	94.5%	92.3%	69.1%
(0,v_{jt},ω_{ij})	97%	0%	48%	49%	91.7%	71.8%	71.1%	89.5%	91.2%	70.4%
(u_{it},v_{jt},ω_{ij})	99%	33%	33%	33%	95.9%	94.6%	85.8%	96.2%	94.2%	86.7%
Partially balanced data for which the t-dimension is balanced										
(u_{it},0,0)	70%	70%	0%	0%	65.0%	64.9%	0.1%	42.1%	1.9%	1.7%
(0,v_{jt},0)	70%	0%	70%	0%	64.5%	1.0%	0.8%	2.4%	2.2%	2.4%
(0,0,ω_{ij})	70%	0%	0%	70%	0.1%	0.2%	0.2%	45.1%	44.6%	2.8%
(u_{it},v_{jt},0)	97%	48%	49%	0%	94.9%	89.6%	30.8%	88.8%	25.0%	27.0%
(u_{it},0,ω_{ij})	97%	48%	0%	49%	89.8%	90.6%	2.4%	93.8%	91.4%	4.2%
(0,v_{jt},ω_{ij})	97%	0%	48%	49%	91.3%	2.4%	1.5%	89.7%	91.3%	1.2 %
(u_{it},v_{jt},ω_{ij})	99%	33%	33%	33%	95.8%	93.9%	31.7%	96.2%	92.6%	32.3%

Note: The number of replications is 1,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x and are sources of bias. The first element is u_p, the second element is v_p and the third element is ω_p. u_{it}, v_{jt}, and ω_{ij} are sources of bias.

Table 5: PPML: Tracing source biases- 50% random missing

	Source of bias				The order of demean-dimension for within estimators					
	$u_{it}+v_{jt}+\omega_{ij}$	u_{it}	v_{jt}	ω_{ij}	i	i-t	i-t-j	j	j-t	j-t-i
Unbalanced data for which all three dimensions are unbalanced										
(u_{it},v_{jt},0)	97%	48%	49%	0%	90.6%	91.9%	54.7%	90.8%	91.6%	53.7%
(u_{it},0,ω_{ij})	97%	48%	0%	49%	94.6%	91.6%	71.2%	90.7%	70.2%	71.4%
(0,v_{jt},ω_{ij})	97%	0%	48%	49%	90.7%	68.2%	68.8%	93.4%	91.0%	69.9 %
(u_{it},v_{jt},ω_{ij})	99%	33%	33%	33%	96.1%	95.0%	85.6%	96.1%	94.2%	85.4%
Partially balanced data for which the t-dimension is balanced										
(u_{it},v_{jt},0)	97%	48%	49%	0%	89.5%	34.4%	34.4%	89.9%	91.1%	31.1%
(u_{it},0,ω_{ij})	97%	48%	0%	49%	93.9%	90.6%	1.2%	90.3%	91.8%	1.4%
(0,v_{jt},ω_{ij})	97%	0%	48%	49%	90.6%	5.2%	0.5%	94.3%	92.3%	0.6%
(u_{it},v_{jt},ω_{ij})	99%	33%	33%	33%	95.5%	93.0%	30.1%	95.8%	91.8%	30.1%

Note: The number of replications is 1,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_p, the second element is v_p and the third element is ω_p. u_{it}, v_{jt}, and ω_{ij} are sources of bias as in equation (14).

Table 6: PPML: Inference accuracy

	Pooled MLE bias	ba: i-t-j	ub: i-t-j	pb: i-t-j	ub: j-t-i	pb: j-t-i	ub: t-i-j	pb: t-i-j	ub: i-j-t	pb: i-j-t
$(u_{it}, 0, \omega_{ij})$, mean of bias	.94	.02	.71	.01	.71	.01	.73	.02	.70	.01
$(u_{it}, 0, \omega_{ij})$, SD	(.27)	(.84)	(.61)	(.95)	(.60)	(.94)	(.60)	(.93)	(.58)	(.90)
$(u_{it}, 0, \omega_{ij})$, rejection% mean		.096	.453	.142	.448	.135	.478	.129	.466	.144
$(u_{it}, 0, \omega_{ij})$, rejection% coverage		(.077,.114)	(.422,.484)	(.120,.164)	(.417,.479)	(.114,.156)	(.447,.509)	(.108,.150)	(.435,.497)	(.122,.166)
$(u_{it}, 0, \omega_{ij})$, rejection% mean*				.062		.063		.057		.063
$(u_{it}, 0, \omega_{ij})$, rejection% coverage*				(.046,.077)		(.048,.078)		(.043,.071)		(.048,.079)
$(0, v_{jt}, \omega_{ij})$, mean of bias	.94	.01	.69	.01	.70	.01	.71	.01	.69	.04
$(0, v_{jt}, \omega_{ij})$, SD	(.24)	(.93)	(.63)	(.98)	(.60)	(.95)	(.58)	(.94)	(.59)	(.96)
$(0, v_{jt}, \omega_{ij})$, rejection% mean		.146	.461	.145	.462	.139	.472	.143	.453	.161
$(0, v_{jt}, \omega_{ij})$, rejection% coverage		(.124,.168)	(.420,.492)	(.123,.167)	(.431,.493)	(.118,.160)	(.441,.503)	(.121,.165)	(.422,.484)	(.138,.184)
$(0, v_{jt}, \omega_{ij})$, rejection% mean*				.062		.054		.068		.074
$(0, v_{jt}, \omega_{ij})$, rejection% coverage*				(.046,.077)		(.039,.068)		(.051,.083)		(.058,.090)

Note: The number of replications for simulation is 1,000. Boostrapped SE are reported with * by using 100 replications. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_{it}, the second element is v_{jt} and the third element is ω_{ij}. Ba, ub, and pb columns represent the estimates of the within estimators for balanced, unbalanced and partially balanced data, respectively. Results with balanced data are only reported for the itj within estimator, but the estimated bias, SD, rejection mean and rejection coverage are all very close to partially balanced data results. Estimation results with balanced data are replicated with partially balanced data for all four distinct within transformations with distinct orders of demeanings/projections.

4 An Empirical Application

We apply the proposed estimators to estimate trade cost elasticity using highly disaggregated data and a gravity model with 4WEC.

4.1 Empirical Gravity Model

For the sake of completeness, in Appendix D we provide the micro-foundation for the gravity model for product (or sector) level estimation based on Helpman et al. (2008) (hereafter HMR).[16] The following estimable equation can be derived from the model:

$$
\begin{aligned}
lnM_{ijkt} \quad = \quad & (\sigma_k - 1)ln\alpha_k - (\sigma_k - 1)lnc_{jkt} + lnN_{jkt} + (\sigma_k - 1)lnP_{ikt} \\
& + lnE_{ikt} - (\sigma_k - 1)ln\tau_{ijkt} + lnV_{ijkt}
\end{aligned} \tag{16}
$$

where M_{ijkt} is the volume of imports of product k by country i from country j (at time t, which is the same for the other variables), α_k is equal to $(\sigma_k - 1)/\sigma_k$ and determines the elasticity of substitution across varieties of product k, c_{jkt} is a measure of the average product-specific productivity of firms in j, N_{jkt} is the number of firms producing k in j, P_{ikt} is the price index associated with varieties of k in i, E_{ikt} is the expenditure by i on k from all origins, τ_{ijkt} is the variable trade cost of exporting k from j to i, and V_{ijkt} is a monotonic function of the fraction of j's firms producing k that export to i. $-(\sigma_k - 1)$ is the trade elasticity with respect to the variable trade cost for product k (i.e. τ_{ijk}), and this is what we are interested in estimating.

P_{ikt}, and N_{jkt} and c_{jkt} are unobserved, but can be accounted for using exporter-product-specific factors and importer-product-specific factors. As shown in Appendix D, the above equation can be simplified to:

$$
lnM_{ijtk} \quad = -(\sigma_k - 1)ln\tau_{ijkt} + lnW_{ijkt} + u_{1ikt} + v_{1jkt} \tag{17}
$$

where W_{ijkt} is a monotonic function of V_{ijkt} and relates to the fraction of exporting firms, $u_{1ikt} = (\sigma_k - 1)lnP_{ikt} + lnE_{ikt} + (\sigma_k - 1)ln\alpha_k$, and $v_{1jkt} = lnN_{jkt} - (\sigma_k - 1)lnc_{jkt} + ln(\gamma a_{Lt}^{\gamma - \sigma_k + 1}) - ln(\gamma - \sigma_k + 1) - ln(a_{H_{kt}}^{\gamma} - a_{L_{kt}}^{\gamma})$.

Our departure from HMR is that the fraction of firms that export can vary across products and, more importantly, that in the estimation of eq.(17) we bypass the first stage estimation of the foreign market entry decision. This is possible because, using these 4WEC as the FEs, we can capture all variables that vary over ijk, ijt, ikt and

[16] In this model, the terms "product" and "sector" are interchangeable.

jkt. As long as foreign market entry decision depends solely on these variables, our approach is valid.

We assume lnW_{ijkt}, which is a function of factors that determine firm productivity and heterogeneity, consists of a number of additively separable factors:

$$lnW_{ijkt} = \omega_{2ijk} + \eta_{2ijt} + u_{2ikt} + v_{2jkt} + \varepsilon_{2ijkt} \tag{18}$$

where ω_{2ijk} and η_{2ijt} are pair-product and pair-time specific random shocks to fixed costs (i.e. they consist of the fixed costs in aggregate-level data in HMR), u_{2ikt} and v_{2jkt} are, respectively, importer-product-time specific and exporter-product-time specific random shocks to fixed costs, and ε_{2ijkt} consists of all remaining random elements that could affect firms' export market entry decisions, but are not correlated with $ln\tau_{ijkt}$. These five factors together determine the fraction of firms producing k that export from j to i at time t, as well as their volume of exports once they have entered the export market.

We assume that the variable trade cost is made up of tariff and non-tariff components, and that the latter consist of five additively separable factors, such that:

$$ln\tau_{ijkt} = ln(1 + tariff_{ijkt}) + \omega_{3ijk} + \eta_{3ijt} + u_{3ikt} + v_{3jkt} + \varepsilon_{3ijkt} \tag{19}$$

where $\omega_{3ijk} + \eta_{3ijt} + u_{3ikt} + v_{3jkt} + \varepsilon_{3ijkt}$ is the linear approximation of non-tariff-related variable trade costs.

Substituting eq.(18) and (19) into eq.(17), we obtain:

$$lnM_{ijtk} = -(\sigma_k - 1) \cdot ln(1 + tariff_{ijkt}) + \omega_{ijk} + \eta_{ijt} + u_{ikt} + v_{jkt} + \varepsilon_{ijkt} \tag{20}$$

where $\omega_{ijk} \equiv \omega_{2ijt} - (\sigma_k - 1)\omega_{3ijk}$, $\eta_{ijt} \equiv \eta_{2ijt} - (\sigma - 1)\eta_{3ijt}$, $u_{ikt} \equiv u_{1ikt} + u_{2ikt} - (\sigma_k - 1)u_{3ikt}$, $v_{jkt} \equiv v_{1jkt} + v_{2jkt} - (\sigma_k - 1)v_{3jkt}$, and $\varepsilon_{ijkt} \equiv \varepsilon_{2ijkt} - (\sigma_k - 1)\varepsilon_{3ijkt} + (\sigma - \sigma_k)\eta_{3ijt}$, with σ being the average of σ_k.

4.1.1 Unobserved heterogeneity

Although ω_{ijk}, η_{ijt}, u_{ikt} and v_{jkt} are unobservable, we can account for them using FEs with partially balanced data. In other words, $(\sigma_k - 1)$ can be consistently estimated as long as we have:

$$Cov(ln(1 + tariff_{ijkt}), \varepsilon_{ijkt} | \omega_{ijk}, \eta_{ijt}, u_{ikt}, v_{jkt}, s_{ijkt} = 1) = 0 \tag{21}$$

In other words, in the estimation of eq.(20), we assume that once we control for all ijt-, ijk-, ikt- and jkt-varying unobserved factors, the remaining trade costs factors, ε_{ijkt}, are uncorrelated with $tariff_{ijkt}$. Also, because we directly account for the determinants of export market entry decisions, eq.(21) is not conditional on $M_{ijkt} > 0$, unlike in the HMR method. This is important, because any estimation that is conditional on $M_{ijkt} > 0$ involves using highly unbalanced data, which will prohibit us from applying our proposed method to account for the error components.

4.1.2 Zero flows: self-selection of trade

A key insight of HMR is that zero trade flows commonly observed in bilateral trade data are the result of firms' self-selection into (or out of) a foreign market, and their decisions are not uniform across countries (as well as sectors in our case), due to heterogeneity in firms' productivity. HMR suggests a two-stage procedure to account for self-selection and firm heterogeneity for aggregate data. The first-stage estimation requires an exclusion restriction variables that affect firms' entrance into a foreign market, but not their trade volume. To directly extend the HMR approach to the estimation of product-level models, the first stage estimation would require an exclusion restriction variable that affects firms' entrance into individual product markets in a foreign country, but not their trade volumes in each of those markets. A commonly used exclusion restriction variable for aggregate trade data is religious similarity between countries, which varies only at the country level. Finding an exclusion restriction that varies at the product level is very difficult due to data limitations. Our estimator allows us to bypass the first stage of estimation by directly controlling for factors that affect both entry and export volume decisions. This is possible because multiple unobserved factors play the role of exclusion restrictions. As long as we can justify that four additively separable unobserved factors, ω_{ijk}, η_{ijt}, u_{ikt} and v_{jkt}, can approximate the determinants of a firm's self-selection process as well as firm heterogeneity, we can express the resulting gravity equation using a MWEC model as in eq. (20).

4.2 Data

The estimation of eq.(20) requires only import and tariff data, as all other factors are captured by the 4WEC. Import data are drawn from the UN COMTRADE database and cover 48 countries from 2004 to 2008, for about 5,000 products at the Harmonized System 6-digit (abbreviated as HS6 hereafter) product level.[17] HS6 is the most disaggregated international standard product level used by all countries. Bilateral tariff data are sourced from the TRAINS database of United Nations Conference on Trade and Development (UNCTAD) and the World Trade Organization (WTO). Details of the data are provided in Appendix E.

[17] The original data cover 72 countries, but the 25 member countries of the European Union are treated as one in our analysis, resulting in 48 importers and exports in our sample. Country and product coverage was largely determined by tariff data availability.

If the data were completely balanced, the total number of observations would have been 56.4 million ($= 48 \times 47 \times 5,000 \times 5$). However, only 91 per cent of these observations have tariff data and, more importantly, only 11 per cent have positive import figures.[18] Conventionally, there are two ad hoc methods for treating missing trade flows observations. The first method is to leave them as missing. However, this not only dramatically reduces the number of observations, but also leads to bias if the 'true' values of some missing figures are simply zeros. HMR shows that properly accounting for zero imports is important in the estimation of the gravity model. The second treatment method is to fill in all missing figures with zeros. However, this could also lead to bias if the 'true' values of some missing figures are not zero. Here, we take a middle way, by proposing a method to identify which missing trade flow figures should be left as missing and which should be replaced by zeros. We only impute a zero value if we can provide a justification, so that the measurement error from the zero imputation is bounded.

Our imputation decision is as follows. For a fixed country-pair at a given time (ijt), if we have no observations for all products, then we treat them as missing, so the observability indicator $s_{ijkt} = 0$ for all k. On the other hand, if we observe at least one product with positive flows for fixed ijt, then we impute all missing entries with zeros, so we have $s_{ijkt} = 1$. The rationale is that if country i is willing to report the imports of one or more products from j, it is reasonable to assume that it is also willing to do the same for other products from j, as long as there are positive imports to report, and therefore, any missing entries are likely to be zeros.

Table 7 shows the coverage of tariff data and import data for unbalanced ("ub") and partially balanced ("pb") data. We partially balanced the import data by imputing zeros for those units that have tariff data available. The "Import(pb: k)" column shows that balancing the data on the k dimension alone reduces the proportion of missing observations from 89 per cent ($= 100\% - 11\%$) down to 14 per cent ($= 100\% - 86\%$). We can apply the same logic to impute zeros for any fixed ikt. For instance, for fixed ikt, if we have no observations for all exporters, then we treat them as missing and so $s_{ijkt} = 0$ for all j, but if we have observations for one or more exporters, then we impute all missing entries with zero, and so $s_{ijkt} = 1$. The rationale is that, for fixed k and t, if country i is willing to report the imports from one exporting country, it is reasonable to assume that it is also willing to do the same for other exporters, as long as there are positive imports to report. Further balancing the data in the i dimension using the same logic provides an additional 4 per cent of observations. As we use all time periods, the import data is fully balanced in all four dimensions. However, tariff data are unbalanced but the proportion of missing is small. As such, the dataset used in the estimation is partially balanced and allows us to use the proposed estimator.

[18]The use of heavily unbalanced disaggregated data in most previous studies was inevitable, because most of the data sourced from the World Bank, the IMF, and the OECD report only positive trade flows. The degree of unbalancedness of data depends on the level of data disaggregation, increasing up to about 88.7 per cent for HS6 data in our sample.

Table 7: Basic Statistics

	Main dependent and independent variables				
	Tariff	Import(ub)	Import(pb: k)	Import(pb: k,j)	Import(pb: k,j,i)
Num of obs	51,239,196	6,384,381	48,289,677	50,417,005	50,901,091
Mean	7.926	4201.746	555.12	531.70	526.64
SD	(23.8)	(104802.1)	(38132.7)	(37319.7)	(37141.9)
% out of full sample	91%	11%	86%	89%	90%

Note: Standard deviations are provided in parentheses. A fully balanced dataset would have 56,400,000 (=48×47×5000×5) observations. Because certain pairs do not appear in the sample, we lose 1,342,436 (=56,400,000-55,057,564) or 2.4% of the observations. We lose a further 3,818,368 or 6.8% of the observations due to no reporting on tariff data for certain products. Thus, we lose about 9% of the observations in total.

The final, partially balanced, dataset is balanced in the i, j and k dimensions,[19] and has 75 per cent of entries imputed to be zero. For $392,105$ $(= 51,293,196 - 50,901,091)$ observations, we treated the import data as missing (i.e. $s_{ijkt} = 0$).

In summary, for the highly disaggregated trade data, the fraction of missing data is very high, and partially balancing the data makes a significant difference. For our dataset, if we were to estimate the trade cost elasticity with positive imports only, information on 75 per cent of the observations of tariffs and imports would not be used at all. Our simulation results show that the magnitude of bias increases substantially with the proportion of missing data for within estimators. This implies that we have made significant progress in the use of data by reducing the missing proportion from 89 per cent to 10 percent.

4.3 Estimation Results

Our estimation equation is obtained from eq.(20) in Section 2:

$$ln(M_{ijkt}) = \alpha_0 + \alpha \cdot ln(1 + tariff_{ijkt}) + \omega_{ijk} + \eta_{ijt} + u_{ikt} + v_{jkt} + \varepsilon_{ijkt} \quad (22)$$

Our primary objective is to obtain a consistent estimate of $\alpha = (1 - \sigma)$, the elasticity of trade costs.[20]

Table 8 shows OLS and FE estimation results for the log-linear model with positive import data only. Recall that this dataset has 90 per cent of the entries missing. Column (1) shows the OLS results. The estimated trade cost elasticity is -6.47 and it is statistically significant. Column (2) includes pair-time FEs, which subsume

[19] Note that the balancedness in the i, j and k dimensions implies that, for fixed t, either observations are available for all ijk or no observations are available for any ijk at all.

[20] In contrast to the theoretical model in Section 2, here we focus on the average trade elasticity across all products (i.e. $1 - \sigma_k = 1 - \sigma$), for the sake of simplicity of presentation. The estimate for $1 - \sigma_k$ using product-by-product estimations is available on request.

typical gravity equation covariates, such as size of importers and exporters, geographical and cultural distance between trading countries, and country-level multilateral resistance terms (MRTs). Compared to the OLS estimate, the estimated elasticity decreases by 37 per cent in absolute terms. Column (3) adds import-product-time and exporter-product-time FEs to account for country-product-level MRTs. The estimated elasticity is -3.97, which is very close to the estimate in Column (2). This implies that once we account for pair-time FEs, additional control of country-product-level MRTs has no significant effect on the trade cost elasticity estimates. Column (4) includes the most comprehensive set of FEs, by adding pair-product FEs, and it is our preferred specification. The estimated elasticity is -0.85, which is very small compared to the estimates in the previous literature.[21] However, we must be careful when interpreting the results in Columns (3) and (4) because, given that the data are highly unbalanced, potential bias due to multiple demeanings could be very large.

Table 9 reports the estimation results using the same set of model specifications, but with the data partially balanced in the i, j and k dimensions. Recall that for this partially balanced dataset, Proposition 6 and Corollary 11 imply that our proposed estimator can provide a consistent estimate in both linear and Poisson models. This table reports one of our main results. As shown in Columns (1) and (2), including zero flows dramatically decreases the elasticity estimates if country-product-level MRTs are not accounted for. This is sensible, because any given pair of countries typically trade only a small range of specific products (recall that 75 per cent of import entries are zeros). Therefore, even though tariffs are reduced a lot at the country-product level (e.g. due to recent trade liberalization that removed tariffs for all products), imports do not change much because they are zero throughout the sample period. Unlike the unbalanced data, including FEs in Column (3) to control country-product-level MRTs is critical to the elasticity estimate. This implies that firms' market entry decisions are crucially dependent on factors at the product level. Our preferred specification, which has the most comprehensive set of FEs, is shown in Column (4). The estimated elasticity is -2.69, which is about 33 per cent smaller in absolute terms than even the lower bound estimate of -4 in the literature,[22] as well as our estimate of -4.08 from the conventional method without product-level FEs.

The results in Table 9 have two implications. Firstly, particularly when there are lots of zero imports, properly accounting for multiple unobserved factors is critical. Comparison of the results between Columns (2) and (3) shows that ignoring country-product-level MRTs severely underestimates trade cost elasticity (-0.23 vs -6.37). On the other hand, ignoring pair-product unobserved factors severely overestimates trade cost elasticity (-6.37 vs

[21] Anderson and van Wincoop (2004) provide an excellent survey on trade cost elasticity and show that trade elasticity estimates in the literature range from -4 to -10. More recent empirical studies, such as Bergstrand et al. (2013) and Simonovska and Waugh (2014), also find similar estimates.

[22] There are a fair number of empirical studies trade cost elasticity. Anderson and van Wincoop (2004) have provided an excellent survey on this literature and show that trade cost elasticity estimates range from -4 to -10.

-2.39). Secondly, comparison of the results in Column (4) in Tables 8 and 9 shows that ignoring zero observations (or bias from within transformation due to unbalancedness) leads to severe underestimation of trade cost elasticity (-0.85 vs -2.69).

Silva and Tenreyro (2006) show that in the presence of heteroskedastic errors, the OLS estimates of log-linear models could be severely biased. They recommend using the PPML method instead. They show through simulation that the PPML estimator has some good properties. The gravity model can be adjusted for the PPML estimation using the following equation as the conditional mean specification:

$$E(M_{ijtk}|tariff_{ijkt}, \eta_{ijt}, \omega_{ijk}, u_{ikt}, v_{jkt}) = exp(\alpha_0 - (\sigma_k - 1) \cdot ln(1 + tariff_{ijkt}) + \eta_{ijt} + \omega_{ijk} + u_{ikt} + v_{jkt})$$

We can obtain a consistent estimate for α using the PPML model with the following CRE term for partially balanced data:

$$\begin{aligned} E(M_{ijtk}|tariff_{ijkt}, \omega_{ijk}, \eta_{ijt}, u_{ikt}, v_{jkt}) &= exp(\alpha_0 - \alpha \cdot ln(1 + tariff_{ijkt}) \\ &\quad + \beta_1 \overline{ltariff}_{ij\cdot t} + \beta_2 \overline{ltariff}_{i\cdot kt} + \beta_3 \overline{ltariff}_{\cdot jkt} + \beta_4 \overline{ltariff}_{i\cdot\cdot t} \\ &\quad + \beta_5 \overline{ltariff}_{\cdot j\cdot t} + \beta_6 \overline{ltariff}_{\cdot\cdot kt} + \beta_7 \overline{ltariff}_{\cdots t} + \ldots \beta_{15} \overline{ltariff}_{\cdots\cdot}) \end{aligned}$$

where $ln(1 + tariff) = ltariff$, and $\overline{ltariff}_{ijk\cdot} = \frac{1}{T_{ijk}} \sum_{t=1}^{T} s_{ijkt} ltariff_{ijkt}$, and so forth.

Table 10 shows the results. Exactly the same pattern is observed as in the log-linear model in terms of the changes in the elasticity estimates with respect to various FEs. The elasticity estimate in the preferred specification, shown in Column (4), is comparable to the log-linear estimate (-3.42 vs -2.69). Importantly, in both log-linear and PPML estimations with zero observations and the most comprehensive FEs, the trade cost elasticity estimates, -2.69 and -3.42, are smaller in absolute magnitude than the lower bound estimate in the literature.

Table 8: OLS and fixed effects estimates with positive import values only

	Elasticity estimates			
	(1)	(2)	(3)	(4)
Trade elasticity	-6.47***	-4.08***	-3.97***	-0.85***
	(0.60)	(0.24)	(0.74)	(0.30)
Fixed effects	No	ijt	ijt, ikt, jkt	ijt, ikt, jkt, ijk
Num of obs	6,384,381	6,384,381	6,384,381	6,384,381
Model	Log-linear Model			

Note: Cluster (pair) robust standard errors are provided in parentheses. Fixed effects are accounted for by the within transformations defined in Section 2.

Table 9: OLS and fixed effects estimates with zero imports and partially balanced data

	Elasticity estimates			
	(1)	(2)	(3)	(4)
Trade elasticity	-0.99***	-0.23***	-6.37***	-2.69***
	(0.14)	(0.04)	(1.11)	(0.63)
Fixed effects	No	ijt	ijt, ikt, jkt	ijt, ikt, jkt, ijk
Num of obs	50,901,091	50,901,091	50,901,091	50,901,091
Model	Log-linear Model			

Note: The mean estimate of $1-\sigma_k$ for 5,000 HS6 products is provided. Standard errors are provided in parentheses. We use the proposed within estimators with multiple demeanings for partially balanced data. Column (4) reports the result with the t-j-i-k demeanings. If data are partially balanced in three dimensions, the estimate in Column (4) from within estimators of any order of demeanings that remove all four error components should be the same. However, due to data limitations on tariffs, 9% of data are missing, so data are generally unbalanced even after partial balancing in three dimensions. Thus, as we change the demeaning orders of within estimators, trade elasticity estimates with FEs in Column (4) change, and they range from -2.69 to -2.84.

Table 10: Poisson estimates with zero imports and partially balanced data

	Elasticity estimates			
	(1)	(2)	(3)	(4)
Trade elasticity	-3.03***	-0.71***	-4.87***	-3.42***
	(0.43)	(0.12)	(1.14)	(0.61)
Fixed effects	No	ijt	ijt, ikt, jkt	ijt, ikt, jkt, ijk
Num of obs	50,901,091	50,901,091	50,901,091	50,901,091
Model	Poisson Model			

Note: The mean estimate of $1-\sigma_k$ for 5,000 HS6 products is provided. Standard errors are provided in parentheses. We use the proposed within estimators with multiple demeanings for partially balanced data.

5 Conclusion

This paper presents a new framework for estimating partially balanced panel data models with multi-way error components. For linear models, we show that the conventional projection method applied to unbalanced data can only remove one error component without causing bias. But for partially balanced data, our proposed projection approach can control for as many error components as the number of balanced dimensions plus one.

We extended our proposed method to the exponential conditional mean framework with the PPML model. We propose a PPML model with a control function obtained from multiple projections and achieve the same results as in the linear model case. We examine the finite sample properties of the proposed estimators using Monte Carlo simulations. The proposed estimators are found to be consistent and provide correct inferences with substantial power and no size bias.

The proposed estimators are particularly useful when both the size of the dataset and the dimension of unobserved factors are large. The dataset used in this paper has over 50 million observations, yet the proposed estimators can be implemented using a standard desktop computer.

The usefulness of the proposed estimators is illustrated through the estimation of trade cost elasticity using four-dimensional trade data. This application makes two important contributions to the empirical trade literature in its own right. First, it is the first study that can comprehensively account for four distinct unobserved factors that are considered important in the estimation of trade cost elasticities and, thus, avoid the bias that has plagued previous studies. Second, it is the first study that extends the treatment of zero bilateral trade flows from country-level data, as in Helpman et al. (2008), to product-level data. Dealing with zero observations is an important issue in modeling product-level trade data, because at this level of disaggregation, the majority of the observations are either missing or zeros. With proper treatment of zero observations and unobserved factors, we obtain smaller trade cost elasticity estimates in both linear and Poisson models, compared to estimates in the previous literature.

A Supplementary MC results

A.1 Degrees of freedom adjustment for the three-way within estimator

It is necessary to to adjust the standard errors when applying within estimators, because the incidental parameters problem causes them to be biased otherwise (Neyman and Scott, 1948; Lancaster, 2000). We lose one degree of freedom for each dimension from calculating the means. We suggest a standard error adjustment for three-dimensional data that is an extension of the adjustment for two-dimensional panel data, as follows.

If we run a simple OLS regression of $\widetilde{y}_{ijt}$ on $\widetilde{\mathbf{X}}_{ijt}$, certain statistical packages may incorrectly estimate the standard error, σ_ε^2, as:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{nmT-k} \tag{23}$$

instead of the correct one:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{(n-1)(m-1)(T-1)-k} \tag{24}$$

where we lose one degree of freedom for each dimension, because we use the demeaned variables, and $\widehat{\widetilde{\varepsilon}}_{ijt}$ are the residuals from the OLS regression. Without the degrees of freedom adjustment, the standard error will be underestimated, so that inferences will tend to over-reject the null hypothesis.

For instance, the degrees of freedom can be adjusted as follows for a simple t-test in a univariate case where the null is $H_0 : \alpha_1 = 0$. The statistic that is generated by certain statistical packages for inference after the OLS regression of $\widetilde{y}_{ijt}$ on $\widetilde{\mathbf{X}}_{ijt}$ is:

$$t^* = \frac{\hat{\alpha}_{1,ols} - 0}{\sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{nmT-k}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})_{11}}} \tag{25}$$

However, what we should use instead is:

$$t^{**} = \frac{\hat{\alpha}_{1,ols} - 0}{\sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{(n-1)(m-1)(T-1)-k}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})_{11}}} = t^*\sqrt{\frac{(n-1)(m-1)(T-1)-k}{nmT-k}} \tag{26}$$

where $(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})_{11}$ is the (1,1) element of $(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})$.

For unbalanced data where the number of missing observations is a_1 and the proportion of missing observations is $1-a_2$, certain statistical packages may incorrectly estimate the standard error σ_ε^2 as:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widehat{\varepsilon}}_{ijt}}{(nmT - a_1 - k)}. \tag{27}$$

The one we should use instead, however, is:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widehat{\varepsilon}}_{ijt}}{a_2 \cdot [(n-1)(m-1)(T-1)] - k}. \tag{28}$$

A.2 Results : 3WEC model for unbalanced data and the magnitude of bias in linear model

Table 11 reports the finite sample behaviour of the within estimators for α. The following observations can be drawn:

(i) As long as the directions of bias due to u_{it}, v_{jt} and ω_{ij} are the same, all six within estimators reduce the bias significantly. Even in the worst case (in the fifth row), the t-i-j and t-j-i transformations reduce the bias by four-fifths, from 264 per cent to 53.2 per cent.

(ii) The extent of bias reduction varies amongst the six estimators. For instance, in the first row, the reduction ranges from 0 per cent to 30.5 per cent.

(iii) It is possible to eliminate a single-sourced bias completely if an appropriate estimator is used. For instance, in the first row, where the bias is entirely due to u_{it}, performing the j-dimensional demeaning first can reduce the bias to zero, while doing it last will reduce the bias least.

(iv) When there are two or more sources of bias, no demeaning sequence can eliminate the bias completely.

(v) Whenever the bias due to ω_{ij} does not dominate those due to either u_{it} or v_{jt}, performing the t-dimensional demeaning later will lead to a larger bias reduction. For instance, except for the third row, in which the only source of bias is ω_{ij}, the bias reduction by the i-j-t demeanings tends to be larger than that resulting from the i-t-j demeanings, which in turn tends to be larger than for t-i-j demeanings.

(vi) A corollary of (v) is that if biases caused by u_{it}, v_{jt} and ω_{ij} are of comparable magnitudes, as in the seventh row in Table 11, performing the t-dimensional demeaning last is more effective in reducing bias.

Table 12 shows the same results for the data when 40 per cent of observations are missing. As the proportion of missing observations increases, the remaining bias after within transformations also increases, regardless of the order of demeanings. However, the relationship between the relative sizes of biases due to u_{it}, v_{jt} and ω_{ij} and the order of demeanings remains the same as for the data in Table 11, where 25 per cent is missing. Bias caused by different sources of unobserved heterogeneity can cancel each other out if the sign of the bias is not the same.

Table 11: Tracing source biases- 25% random missing

	Source of bias				The order of demean-dimension for within estimators					
	$u_{it}+v_{jt}+\omega_{ij}$	u_{it}	v_{jt}	ω_{ij}	tij	tji	jti	itj	ijt	jit
$(u_{it},0,0)$	198%	198%	0%	0%	30.5%	25.8%	0%	30.5%	5.5%	0%
$(0,v_{jt},0)$	197%	0%	197%	0%	25.7%	30.4%	30.6%	0%	0%	5.6%
$(0,0,\omega_{ij})$	200%	0%	0%	200%	0%	0%	5.9%	5.9%	11.6%	11.6%
$(u_{it},0,\omega_{ij})$	265%	131%	0%	134%	30.9%	26.2%	6.0%	36.0%	17.1%	11.5%
$(u_{it},v_{jt},0)$	264%	132%	132%	0%	53.2%	53.2%	31.2%	31.2%	6.1%	6.1%
$(0,v_{jt},\omega_{ij})$	265%	0%	131%	134%	26.2%	30.9%	36.0%	6.0%	11.6%	17.0%
$(u_{it},v_{jt},\omega_{ij})$	299%	99%	99%	101%	52.9%	52.9%	35.9%	35.9%	17.0%	17.0%
$(u_{it},\frac{v_{jt}}{2},\frac{\omega_{ij}}{4})$	300%	170%	86%	44%	42.8%	40.7%	17.9%	32.2%	8.7%	5.9%
$(u_{it},\frac{v_{jt}}{4},\frac{\omega_{ij}}{2})$	300%	170%	43%	87%	36.9%	33.5%	11.3%	33.4%	11.6%	7.3%
$(\frac{u_{it}}{2},v_{jt},\frac{\omega_{ij}}{4})$	300%	86%	170%	44%	40.7%	42.8%	32.3%	17.9%	5.9%	8.7%
$(\frac{u_{it}}{2},\frac{v_{jt}}{4},\omega_{ij})$	300%	85%	43%	173%	23.0%	21.7%	14.1%	21.9%	14.4%	12.9%
$(\frac{u_{it}}{4},v_{jt},\frac{\omega_{ij}}{2})$	301%	43%	170%	88%	33.6%	36.9%	33.6%	11.3%	7.4%	11.6%
$(\frac{u_{it}}{4},\frac{v_{jt}}{2},\omega_{ij})$	301%	43%	85%	173%	21.8%	23.1%	22.1%	14.2%	13.0%	14.4%

Note: The number of replications is 2,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_p, the second element is v_p and the third element is ω_p. u_{it}, v_{jt}, and ω_{ij} are sources of bias.

Table 12: Tracing source biases- 40% random missing

	Source of bias				The order of demean-dimension for within estimators					
	$u_{it}+v_{jt}+\omega_{ij}$	u_{it}	v_{jt}	ω_{ij}	tij	tji	jti	itj	ijt	jit
$(u_{it},0,0)$	197%	197%	0%	0%	49.5%	42.5%	0%	49.4%	9.8%	0%
$(0,v_{jt},0)$	197%	0%	197%	0%	42.7%	49.7%	49.7%	0%	0%	9.8%
$(0,0,\omega_{ij})$	200%	0%	0%	200%	0%	0%	10.2%	10.2%	19.5%	19.5%
$(u_{it},0,\omega_{ij})$	265%	131%	0%	134%	49.7%	42.8%	10.2%	57.6%	28.4%	19.5%
$(u_{it},v_{jt},0)$	264%	133%	132%	0%	82.8%	82.8%	49.7%	49.8%	9.9%	9.9%
$(0,v_{jt},\omega_{ij})$	265%	0%	131%	135%	42.7%	49.7%	57.6%	10.2%	19.5%	28.4%
$(u_{it},v_{jt},\omega_{ij})$	299%	99%	101%	100%	82.8%	82.8%	57.6%	57.6%	28.4%	28.4%
$(u_{it},\frac{v_{jt}}{2},\frac{\omega_{ij}}{4})$	300%	170%	86%	44%	68.2%	65.3%	29.5%	51.2%	14.2%	9.7%
$(u_{it},\frac{v_{jt}}{4},\frac{\omega_{ij}}{2})$	301%	170%	43%	87%	59.8%	55.0%	19.0%	53.9%	19.3%	12.3%
$(\frac{u_{it}}{2},v_{jt},\frac{\omega_{ij}}{4})$	301%	85%	171%	44%	66.0%	68.9%	52.0%	29.9%	9.9%	14.7%
$(\frac{u_{it}}{2},\frac{v_{jt}}{4},\omega_{ij})$	301%	85%	43%	173%	38.7%	36.7%	23.7%	36.4%	23.8%	21.6%
$(\frac{u_{it}}{4},v_{jt},\frac{\omega_{ij}}{2})$	301%	43%	171%	87%	55.0%	59.8%	54.1%	19.0%	12.4%	19.4%
$(\frac{u_{it}}{4},\frac{v_{jt}}{2},\omega_{ij})$	301%	42%	85%	173%	36.7%	38.7%	36.6%	23.6%	21.5%	23.9%

Note: The number of replications is 2,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_p, the second element is v_p and the third element is ω_p. u_{it}, v_{jt}, and ω_{ij} are sources of bias.

Table 13: Tracing source biases- 40% of random missing

	Source of bias				The order of demean-dimension for within estimators					
	$u_{it}+v_{jt}+\omega_{ij}$	u_{it}	v_{jt}	ω_{ij}	tij	tji	jti	itj	ijt	jit
(u_{it},0,-ω_{ij})	-1%#	44%	0%	-45%	36.0%	32.2%	-9.6%	27.5%	-7.9%	-16.9%
(-u_{it},0,ω_{ij})	1%#	-44%	0%	45%	-36.1%	-32.3%	9.5%	-27.7%	7.9%	17.1%
(u_{it},-v_{jt},0)	0%#	44%	-44%	0%	4.4%	-4.4%	-36.1%	35.9%	9.0%	-9.3%
(-u_{it},v_{jt},0)	0%#	-44%	44%	0%	-4.6%	4.1%	35.8%	-36.1%	-9.2%	9.0%
(0,v_{jt},-ω_{ij})	-1%#	0%	44%	-45%	32.1%	36.0%	27.5%	-9.5%	-17.1%	-8.0%
(0,-v_{jt},ω_{ij})	1%#	0%	-44%	45%	-32.2%	-36.0%	-27.6%	9.4%	17.0%	8.0%
(u_{it},v_{jt},-ω_{ij})	30%	30%	30%	-31%	50.9%	50.9%	27.6%	27.6%	-7.9%	-7.9%
(u_{it},-v_{jt},-ω_{ij})	-31%	30%	-30%	-31%	4.4%	-4.3%	-40.0%	27.6%	-7.9%	-23.3%
(u_{it},-v_{jt},ω_{ij})	31%	30%	-30%	31%	4.4%	-4.3%	-27.6%	40.1%	23.4%	8.0%

Note: Bias due to u and v, ω and v, and ω and u exactly cancel each other out. The number of replications is 2,000. $(\cdot,\cdot,\cdot)$ represent common elements in unobserved heterogeneity and endogenous x. The first element is u_p, the second element is v_p and the third element is ω_p. u_{it}, v_{jt}, and ω_{ij} are sources of bias, as in equation (3.11).

B Proof of Proposition 7

Proposition 7. For partially balanced data, the number of error components that can be completely removed by within transformation (i.e. demeaning or projection) is the number of balanced dimensions plus one. The unobserved components that can be removed without causing bias depends on which some dimensions are balanced.

Proof. For tractability of notation, we stick to our baseline model with four indexes (i, j, k, t) denoting four data dimensions. In this proof, we focus on explaining how we can manipulate demeaning orders to remove terms through demeanings.

Suppose the dataset is balanced in the i,j and k dimensions, but unbalanced in the t dimension. For this partially balanced dataset, $s_{ijkt} = s_t$. This implies that for fixed ijk, either $s_t = 1$ or $s_t = 0$. As the data are balanced in three dimensions, we have to show that we can remove all four error components, ω_{ijk}, η_{ijt}, u_{ikt} and v_{jkt}.

Using Proposition 1, we can remove any error component, even with unbalanced data. Here we have to demean over the unbalanced t dimension first to remove ω_{ijk} so that it cannot cause any bias after subsequent demeanings:

$$
\begin{aligned}
& s_t\omega_{ijk} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t\omega_{ijk} - \tfrac{1}{n_t}\sum_{j=1}^{n} s_t\omega_{ijk} + \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t\omega_{ijk} \\
&= s_t\omega_{ijk} - \omega_{ijk} - \tfrac{1}{n_t}\textstyle\sum_{j=1}^{n} s_t\omega_{ijk} + \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t\omega_{ijk} \\
&= (s_t\omega_{ijk} - \omega_{ijk}) - (\tfrac{1}{n_t}\textstyle\sum_{j=1}^{n} s_t\omega_{ijk} - \tfrac{1}{n_t}\sum_{j=1}^{n}\omega_{ijk}) \\
&= (\omega_{ijk} - \omega_{ijk}) - (\tfrac{1}{n_t}\textstyle\sum_{j=1}^{n}\omega_{ijk} - \tfrac{1}{n_t}\sum_{j=1}^{n}\omega_{ijk}) \qquad \text{for } s_t = 1 \\
&= 0
\end{aligned}
$$

Here we make use of the fact that $\frac{1}{T}\sum_{t=1}^{T} s_t = 1$ and that, for observed ijk, $s_t = 1$.

For the remaining demeanings, we make use the balancedness of the i, j and k dimensions to avoid inconsistency. For instance, we can use the t-j demeanings to remove u_{ikt} (after removing ω_{ikt}):

$$
\begin{aligned}
& s_t u_{ikt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t u_{ikt} - \tfrac{1}{n_t}\sum_{j=1}^{n} s_t u_{ikt} + \tfrac{1}{n_t}\sum_{j=1}^{n} \tfrac{1}{T}\sum_{t=1}^{T} s_t u_{ikt} \\
&= \big(s_t u_{ikt} - u_{ikt}\tfrac{1}{n_t}\textstyle\sum_{j=1}^{n} s_t\big) - \big(\tfrac{1}{T}\sum_{t=1}^{T} s_t u_{ikt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t u_{ikt}\tfrac{1}{n_t}\sum_{j=1}^{n} 1\big) \\
&= \big(s_t u_{ikt} - u_{ikt} s_t\big) - \big(\tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t u_{ikt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t u_{ikt}\big) \\
&= 0
\end{aligned}
$$

Building on this, we can use the t-j-i demeanings to further remove v_{jkt}, as follows:

$$
\begin{aligned}
& s_t v_{jkt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t v_{jkt} - \tfrac{1}{n_t}\sum_{j=1}^{n}\big(s_t u_{ikt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt}\big) \\
& -\tfrac{1}{m_t}\textstyle\sum_{i=1}^{m}\big[s_t v_{jkt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt} - \tfrac{1}{n_t}\sum_{j=1}^{n}\big(s_t u_{ikt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt}\big)\big] \\
&= s_t v_{jkt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t v_{jkt} - \tfrac{1}{n_t}\sum_{j=1}^{n} s_t v_{jkt} + \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt} \\
& -\tfrac{1}{m_t}\textstyle\sum_{i=1}^{m} s_t v_{jkt} + \tfrac{1}{m_t}\sum_{i=1}^{m}\tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt} + \tfrac{1}{m_t}\sum_{i=1}^{m}\tfrac{1}{n_t}\sum_{j=1}^{n} s_t v_{jkt} - \tfrac{1}{m_t}\sum_{i=1}^{m}\tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt} \\
&= s_t v_{jkt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t v_{jkt} - \tfrac{1}{n_t}\sum_{j=1}^{n} s_t v_{jkt} + \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt} \\
& -v_{jkt}\tfrac{1}{m_t}\textstyle\sum_{i=1}^{m} s_t + \tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt}\tfrac{1}{m_t}\sum_{i=1}^{m} 1 + \tfrac{1}{n_t}\sum_{j=1}^{n} s_t v_{jkt}\tfrac{1}{m_t}\sum_{i=1}^{m} 1 - \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt}\tfrac{1}{m_t}\sum_{i=1}^{m} 1 \\
&= s_t v_{jkt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t v_{jkt} - \tfrac{1}{n_t}\sum_{j=1}^{n} s_t v_{jkt} + \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt} \\
& -v_{jkt} + \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t v_{jkt} + \tfrac{1}{n_t}\sum_{j=1}^{n} s_t v_{jkt} - \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t v_{jkt} \\
&= 0
\end{aligned}
$$

where we use $\frac{1}{m_t}\sum_{i=1}^{m} 1 = 1$.

Likewise, we can further remove η_{ijt} using t-j-i-k demeanings:

$$
\begin{aligned}
& s_t\eta_{ijt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n}(s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt}) \\
& -\tfrac{1}{m_t}\textstyle\sum_{i=1}^{m}[s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n}(s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt})] \\
& -\tfrac{1}{K_t}\textstyle\sum_{k=1}^{K}\{s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n}(s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt}) \\
& -\tfrac{1}{m_t}\textstyle\sum_{i=1}^{m}[s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n}(s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt})]\} \\
& = s_t\eta_{ijt} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n} s_t\eta_{ijt} + \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt} \\
& -\tfrac{1}{m_t}\textstyle\sum_{i=1}^{m}\{s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n}(s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt}) \\
& -\eta_{ijt} + \tfrac{1}{T}\textstyle\sum_{t=1}^{T} s_t\eta_{ijt} + \tfrac{1}{n_t}\sum_{j=1}^{n} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n}\tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt} \\
& +\tfrac{1}{m_t}\textstyle\sum_{i=1}^{m}[s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt} - \tfrac{1}{n_t}\sum_{j=1}^{n}(s_t\eta_{ijt} - \tfrac{1}{T}\sum_{t=1}^{T} s_t\eta_{ijt})]\} \\
& = 0
\end{aligned}
$$

where for observed ijt, $s_t\eta_{ijt} = \eta_{ijt}$.

It should be pointed out that changing the demeaning order of the balanced dimensions will not affect the results. For instance, the t-k-i-j demeaning can also remove all four error components. □

Remark 10. The proof shows that we can change the order of demeanings among balanced dimensions, but the demeaning over the unbalanced dimension must come first in order to avoid inconsistency.

C Exponential Conditional Mean Function and Unbalanced Data

The CRE terms for 4WEC are quite large. Consider the t-k-j-i demeanings. For each of 15 distinct mean terms, the control function consists of the following dummy indicators and their interactions with mean terms:

(i) For the t-dimensional demeaning term, $\sum_{r_1=2}^{T} 1[T_{ijk} = r_1]\psi_1 + \sum_{r_1=2}^{T} 1[T_{ijk} = r_1]\bar{\mathbf{X}}_{ijk\cdot}\xi_1$;

(ii) For the k-dimensional demeaning term, $\sum_{r_2=2}^{K} 1[K_{ijt} = r_2]\psi_2 + \sum_{r_2=2}^{K} 1[K_{ijt} = r_2]\bar{\mathbf{X}}_{ij\cdot t}\xi_2$;

(iii) For the j-dimensional demeaning term, $\sum_{r_3=2}^{N} 1[N_{ikt} = r_3]\psi_3 + \sum_{r_3=2}^{N} 1[N_{ikt} = r_2]\bar{\mathbf{X}}_{i\cdot kt}\xi_3$;

(iv) For the i-dimensional demeaning term, $\sum_{r_4=2}^{M} 1[M_{jkt} = r_4]\psi_4 + \sum_{r_4=2}^{M} 1[M_{jkt} = r_2]\bar{\mathbf{X}}_{ij\cdot t}\xi_4$;

(v) Fo the t-k-dimensional demeaning term, $\sum_{r_2=2}^{K}\sum_{r_1=2}^{T} 1[T_{ijk} = r_1]1[K_{ijt} = r_2]\psi_5 + \sum_{r_2=2}^{K}\sum_{r_1=2}^{T} 1[T_{ijk} = r_1]1[K_{ijt} = r_2]\bar{\mathbf{X}}_{ij\cdot\cdot}\xi_5$.

Similarly, we can define the CRE terms for the remaining ten mean terms.

D Product-level Gravity Model

We extend the Helpman et al. (2008) (HMR) model designed for aggregate data to one that can be used for product-level data. The HMR model explicitly deals with zero (or missing) trade flows, which are very common in product-level data.

Products are indexed by k ($k = 1, 2, ...K$). For aggregate data, $K = 1$; for HS2 data, $K = 98$; and for HS6 data, $K \approx 5,000$. In the following product-level HMR model, we assume each sector produces one product (but with multiple varieties) so that the terms "sector" and "product" are interchangeable. As in Chaney (2008) and Arkolakis et al. (2012), we further assume that the consumer's utility is a Cobb-Douglas function of products and a Dixit-Stiglitz function of varieties within a product.[23]

A representative consumer in country j ($j = 1, 2, ...n$) has the following utility function at time t:

$$U_{jt} = \prod_{k=1}^{K} \left(\int_{h \in H_{jkt}} x_{jkt}(h)^{(\sigma_k - 1)/\sigma_k} dh \right)^{\theta_k \sigma_k / (\sigma_k - 1)} \tag{29}$$

where $x_{jkt}(h)$ is a differentiated variety, h, of product k available in country j (at time t, which is the same for the other variables), H_{jkt} is the set of product k's varieties available in j, σ_k is the elasticity of substitution between varieties of product k, with $\sigma_k > 1$, and is assumed to vary across products, and θ_k is the share of income, Y_{jt}, spent on product k, with $\sum_k^K \theta_k = 1$.

Although consumption can change over time as Y_{jt} changes, for simplicity, we assume no borrowing or lending, so that the consumer is maximizing consumption period by period. There are also no dynamics on the production side, unlike in Alessandria et al. (2014). With j's income Y_{jt} equal to expenditure in every period, we obtain j's demand for variety h of product k from eq.(29):

$$x_{jkt}(h) = \frac{p^*_{jkt}(h)^{-\sigma_k} E_{jkt}}{P_{jkt}^{1-\sigma_k}} \tag{30}$$

where $p^*_{jkt}(h)$ is the price of variety h of product k in j, $E_{jkt} = \theta_k Y_{jt}$ is the expenditure by j on goods from sector k from all origins, and P_{jkt} is the price index associated with varieties from product k in j. P_{jkt} is given by:

$$P_{jkt} = \left[\int_{h \in H_{jkt}} p^*_{jkt}(h)^{1-\sigma_k} dh \right]^{1/(1-\sigma_k)} \tag{31}$$

Using firms' standard mark-up pricing equation in the monopolistic competition model, we have the same

[23] In the empirical application, we treat HS2-digit items as products, and for each fixed HS2-digit, treat HS6 items as varieties. Following the theoretical model, we assume independence across products (at the HS2-digit level) and substitution across varieties (at the HS6 level). Substitution across varieties for a fixed product is governed by σ_k.

outcome for i's imports of product k from j as in HMR for each k. Due to the assumption of independence of consumption across products k, we obtain a trade flows equation that is the same as in HMR, except that now trade flows are product-specific:

$$M_{ijkt} = \left(\frac{c_{jkt}\tau_{ijkt}}{\alpha_k P_{it}}\right)^{1-\sigma_k} E_{ikt} N_{jkt} V_{ijkt} \tag{32}$$

where M_{ijkt} is the import of product k by i from j, N_{jkt} is the number of firms in sector k in j and V_{ijkt} is a monotonic function of the fraction of j's firms in sector k that export to i (which depends on firms' relative productivity and fixed exporting costs to i). τ_{ijkt} is the variable trade cost of product k exported from j to i, and $(1-\sigma_k)$ is the trade elasticity with respect to the variable trade cost for product k. c_{jkt} is a measure of average product-specific productivity of firms in j, and different sectors have different productivity cut-offs for exporting based on a_{kt}, where the inverse of a_{kt} (i.e. $1/a_{kt}$) represents firms' productivity in sector k. In the baseline specification of HMR, $1/a_t$ determines firm-productivity cut-offs for exporting. Here, we extend firm-productivity cut-offs to be sector-specific, $1/a_{kt}$, using the independence assumption. Following HMR, we further assume that productivity is heterogeneous across firms within a sector and that $G(a_{kt})$ has a truncated Pareto distribution, with the support $[a_{Lkt,}a_{Hkt}]$, where a_{Hk} (a_{Lk}) implies the lowest (highest) productivity in sector k, so that $G(a_{kt}) = (a_{kt}^{\gamma} - a_{Lkt}^{\gamma})/(a_{Hkt}^{\gamma} - a_{Lkt}^{\gamma}), \gamma > (\sigma_k - 1)$.

Then, we have:

$$(P_{ikt})^{1-\sigma_k} = \sum_{j=1}^{n}\left(\frac{c_{jkt}\tau_{ijkt}}{\alpha_k}\right)^{1-\sigma_k} N_{jkt} V_{ijkt} \tag{33}$$

$$\begin{aligned} V_{ijkt} &= \int_{a_{Lkt}}^{a_{ijkt}} (a_{kt})^{1-\sigma_k}\, dG(a_{kt}) \\ &= \frac{\gamma a_{Lt}^{\gamma-\sigma_k+1}}{(\gamma-\sigma_k+1)(a_{H_{kt}}^{\gamma} - a_{L_{kt}}^{\gamma})} W_{ijkt} \end{aligned} \tag{34}$$

$$W_{ijkt} = max\left\{\left(\frac{a_{ijkt}}{a_{L_{kt}}}\right)^{\gamma-\sigma_k+1} - 1, 0\right\} \tag{35}$$

P_{ikt} corresponds to i's MRTs for product k, as introduced by Anderson and van Wincoop (2003), but it additionally takes into account the impact of firm heterogeneity on the price of product k.

E Data Description

Our analysis requires bilateral tariff data at the product level. This type of tariff data, however, is not readily available, and needs to be constructed. We construct a dataset for time-variant bilateral tariffs at the HS6 level for every country-pair in our sample.[24]

First, we download each importing country's HS6 simple average applied Most Favoured Nation (MFN) tariffs from the World Integrated Trade Solution (WITS) database. For countries engaged in a customs union (CU), tariff schedules are available at the union level only. We use these CU tariff data to generate country-level MFN tariff data for the member countries of a CU. Also, for some countries, MFN tariffs are not available for certain years. In those cases, we use the data from the most recent previous available year as a substitute. The MFN tariffs are unilateral instead of bilateral. To obtain bilateral tariff data, for each year, we apply an importing country's MFN tariffs to all of its trading partner countries. This procedure is justifiable, as our sample includes WTO members only.

Preferential tariffs are collected separately for applicable pairs for each year at the HS6 level. If preferential tariffs on certain products imposed by an importing country are applicable to multiple exporting countries simultaneously, data are observed at the group level only. To identify which countries belong to a specific group, we use the Preference Beneficiaries data from the Trade Analysis and Information System (TRAINS), and then generate preferential tariff data for each pair. Like MFN tariffs, for the countries engaged in a CU, we only observe the preferential external tariff schedules at the union level. Therefore, we generate an individual member country's bilateral tariff data using the CU tariff data. Preferential tariffs are also not always reported for all years for many countries, so we use the data from the most recent previous available year as a substitute for an unavailable year.

Trade and tariff data are from 2004 to 2008, so all data over this time period are converted into HS6 2007 product codes using UN correspondence tables.

[24] All countries use the same product codes for tariff data up to the HS6 level. For more disaggregated levels, countries do not always use the same product codes.

References

Abrevaya, J., 2013. The projection approach for unbalanced panel data. The Econometrics Journal 16, 161–178.

Alessandria, G., Choi, H., Ruhl, K., 2014. Trade adjustment dynamics and the welfare gains from trade. NBER Working Paper Series 20663.

Anderson, J. E., van Wincoop, E., 2003. Gravity with gravitas: a solution to the border puzzle. American Economic Review 93 (1), 170–192.

Anderson, J. E., van Wincoop, E., 2004. Trade costs. Journal of Economic Literature 42 (3), 691–751.

Anderson, J. E., Yotov, Y. V., Apr. 2011. Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002. NBER Working Papers 17003, National Bureau of Economic Research.

Anderson, J. E., Yotov, Y. V., Apr. 2012. Gold standard gravity. NBER Working Papers 17835, National Bureau of Economic Research.

Antweiler, W., 2001. Nested random efects estimation in unbalanced panel data. Journal of Econometrics 101, 295–313.

Arkolakis, C., Costinot, A., Rodriguez-Clare, A., February 2012. New trade models, same old gains? American Economic Review 102 (1), 94–130.

Balazsi, L., Matyas, L., Wansbeek, T., 2015. The estimation of multidimensional fixed effects panel data models. Econometric Reviews forthcoming.

Baltagi, B., Song, S., October 2006. Unbalanced panel data: A survey. Statistical Papers 47 (4), 493–523.

Baltagi, B. H., 2005. Econometric Analysis of Panel Data. John WIley and Sons, Chichester; Hoboken, NJ.

Baltagi, B. H., Song, S., Jung, B. C., 2001. The unbalanced nested error component regression model. Journal of Econometrics 101, 357–381.

Bergstrand, J. H., Egger, P., Larch, M., 2013. Gravity redux: Estimation of gravity-equation coefficients, elasticities of substitution, and general equilibrium comparative statics under asymmetric bilateral trade costs. Journal of International Economics 89 (1), 110–121.

Broda, C., Weinstein, D. E., May 2006. Globalization and the gains from variety. The Quarterly Journal of Economics 121 (2), 541–585.

Chaney, T., 2008. Distorted gravity: The intensive and extensive margins of international trade. American Economic Review 98 (4), 1707–21.

Davis, P., 2002. Estimating multi-way error components models with unbalanced data structures. Journal of Econometrics 106, 67–95.

Frisch, R., Waugh, F., 1933. Partial time regressions as compared with individual trends. Econometrica 1, 387–401.

Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: Trading partners and trading volumes. The Quarterly Journal of Economics 123 (2), 441–487.

Krautheim, S., 2012. Heterogeneous firms, exporter networks and the effect of distance on international trade. Journal of International Economics 87 (1), 27–35.

Kwak, D. W., Cheong, J., Tang, K. K., 2012. A within estimator for three-level data: An application to the wto effect on trade flows. mimeo, School of Economics Discussion Paper No. 501, 2012.

Lancaster, T., 2000. The incidental parameter problem since 1948. Journal of Econometrics 95, 391–413.

Lovell, M., 1963. Seasonal adjustment of economic time series. Journal of the American Statistical Association 58, 993–1010.

Matyas, L., Balazsi, L., 2012. The estimation of multi-dimensional fixed effects panel data models. mimeo.

Mundlak, Y., 1978. On the pooling of time series and corss sectional data. Econometrica 56, 69–86.

Neyman, J., Scott, E. L., 1948. Consistent estimation from partially consistent observations. Econometrica 16, 1–32.

Silva, J. S., Tenreyro, S., 2006. The log of gravity. The Review of Economics and Statistics 88 (4), 641–658.

Simonovska, I., Waugh, M. E., 2014. The elasticity of trade: Estimates and evidence. Journal of international Economics 92 (1), 34–50.

Wansbeek, T., Kapteyn, A., 1989. Estimation of the error-components model with incomplete panels. Journal of Econometrics 41, 341–361.

Wooldridge, J., 2009. Correlated random effects models with unbalanced panels. mimeo.