A Within Estimator for Three-Level Data: An Application to the WTO Effect on Trade Flows

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

Abstract

This paper proposes a within estimator for three-level data, such as time-variant bilateral trade flows. The estimator helps to address the computational difficulties in estimating, for instance, the gravity model of bilateral trade that needs to control for unobserved country-pair and country-time heterogeneity using fixed effects. Unlike the traditional within transformation that removes cross-sectional heterogeneity, the proposed transformation method removes all three types of heterogeneity, each of which varies in two dimensions of three-level data. We demonstrate the properties of the estimator using empirical examples and Monte Carlo simulations. Simulation results show that the proposed estimators with the adjusted standard errors consistently estimate coefficient parameters and perform correct inference when no data are missing or the missing data are random. When missing data are not random, the proposed within transformation can reduce bias to various degrees, depending on the order of demeaning and sources of bias. As an empirical application, we investigate the WTO effect puzzle by applying the proposed estimator to show that WTO membership has a definite positive effect on bilateral trade flows.

JEL Code: C23
Keywords: Fixed Effects; Within Estimator; Panel Data; World Trade Organization

[*]School of Economics, University of Queensland,QLD 4072, Australia; e-mail: d.kwak@uq.edu.au.

[†]Corresponding author; School of Economics, University of Queensland, QLD 4072, Australia; e-mail: j.cheong@uq.edu.au; tel: +61-7-3346-7048; fax: +61-7-3365-7299.

[‡]School of Economics, University of Queensland,QLD 4072, Australia; e-mail: kk.tang@uq.edu.au.

1 Introduction

The fixed effects (FEs) model is one of the most commonly used models in panel data analysis. While FEs models, in theory, can be estimated using the conventional Least Squared Dummy Variable (LSDV) estimator, in practice they are often estimated using a within estimator that removes cross-sectional heterogeneity through a within transformation. An advantage of the within estimator over the LSDV one is that it can handle large data-sets and thus a large amount of unobserved heterogeneity. The equivalence of the two estimators for two-level panel data (e.g. $i \times t$) can be shown using the theorem by Frisch and Waugh (1933) and Lovell (1963) (the FWL theorem hereafter), and is well documented in standard econometric texts.[1]

To the best of our knowledge, at least two types of within estimator have been introduced to extend the within estimator to three-level panel data (e.g. $i \times j \times t$). Egger and Pfaffermayr (2004), Cheng and Wall (2005), and Serlenga and Shin (2007), in handling bilateral trade data, treat the $i \times j$ dimension as a cross-sectional country-pair dimension $p = i \times j$, effectively transforming three-level data into two-level data ($p \times t$). On the other hand, Matyas (1997) considers unobserved heterogeneity for each of the three dimensions, but assumes heterogeneity across different dimensions to be uncorrelated. Both approaches, therefore, cannot account for unobserved country-time heterogeneity.

To fill this gap, this paper develops a number of within estimators for three-level panel data that give equivalent estimates to the LSDV estimator. It focuses on the most common time-variant bilateral trade data, but the proposed estimator can be easily extended to other types of three-level data including *any* bilateral *flows panel data* such as migration, foreign direct investment, international aid and remittance. Moreover, our proposed within estimator can be applied to higher-level data. One of our estimators is equivalent to the within estimator suggested by Matyas and Balazsi (2012).[2] However, our paper provides the adjustment of standard error under large n and fixed T for both balanced and unbalanced panel data, as well as much more extensive sensitive analyses for the coefficient estimates for unbalanced panel data. For balanced panel data,[3] the proposed estimators remain algebraically equivalent to the LSDV estimator. However, for unbalanced data they are biased and the magnitude of bias depends on the order of demeaning. In this paper we (i) provide the conditions for the within estimator to be unbiased, (ii) quantify the potential bias of the proposed estimators, (iii) compare the magnitude of bias to other alternative estimators for

[1] See, for instance, Wooldridge (2010, Sec 10.5.3), Hsiao (2003, Chapter 3), Cameron and Trivedi (2005, Chapter 21), and Davidson and MacKinnon (1993, Section 1.4).

[2] They focus on the bias of the within estimator's coefficient in unbalanced panel and dynamic autoregressive models.

[3] There should be no missing data for all $i = 1, 2, ..., n$, $j = 1, 2, ..., n$, and $t = 1, 2, ..., T$ for three-level data to be balanced.

unbalanced panel data, (iv) identify the within estimator with the smallest bias, and (v) provide a simple adjustment of the standard error.

An application of our proposed within estimator is the estimation of the gravity model for bilateral trade flows. The recent literature on such a gravity model emphasizes the importance of accounting for both unobserved country-pair heterogeneity (CPH) and unobserved country-time heterogeneity (CTH).[4] However, it is computationally challenging to estimate the gravity model with both CPH and CTH using the LSDV estimator, since the number of dummy variables is typically large. For instance, for the data-set used by Magee (2008) that has 133 countries and 19 periods, the LSDV estimator entails 22,610 dummies in total. In response, previous studies in this literature either limit the period of the data-set (e.g. Magee, 2008), use the LSDV with five-year data (e.g. Baier and Bergstrand, 2007; Subramanian and Wei, 2007b; Eicher and Henn, 2011), do not control for unobserved CPH (e.g. Subramanian and Wei, 2007b; Roy, 2011), or do not control for unobserved CTH (e.g. Serlenga and Shin, 2007; Liu, 2009).[5] Our proposed within estimator can address this problem, allowing researchers to control for both CPH and CTH. As an illustration, we re-examine the World Trade Organization (WTO) membership effect on trade flows. Previous literature tends to conclude the absence of any discernible WTO effect, running against the theoretical prediction. As the proposed within estimator allows us to overcome a computational obstacle in accounting for both CPH and CTH when using large data-sets, we can obtain precise estimates for the WTO membership effect on trade flows. Our results indicate a definite positive effect of WTO membership on trade flows.

The rest of the paper is organized as follows. Section 2 describes the within estimator we propose for balanced panel data. Section 3 examines robustness properties of the within estimator for unbalanced panel data. Section 4 provides an application and the last section concludes.

2 The within estimator for balanced three-level data

2.1 The theory

For time-variant bilateral data, any observation has a unique identifier ijt, where $i = 1, 2, \dots m$ and $j = 1, 2, \dots n$ are import-country and export-country identifiers respectively, and $t = 1, 2, \dots T$ is the time identifier. We also assume data are balanced, i.e. there are no missing observations

[4] Studies that emphasize CPH include Baldwin and Taglioni (2006); Baier and Bergstrand (2007); Magee (2008); Eicher, Henn, and Papageorgiou (2012); Cheong, Kwak, and Tang (2012), and studies that emphasize CTH include Anderson and van Wincoop (2003); Magee (2008); Subramanian and Wei (2007a); Eicher, Henn, and Papageorgiou (2012); Cheong, Kwak, and Tang (2012).

[5] Those studies do not control for unobserved CTH but account for CPH using the method in Egger and Pfaffermayr (2004), Cheng and Wall (2005), and Serlenga and Shin (2007).

for any variables. We use the following notations for projections: For any matrix $\mathbf{A}$, let $\mathbf{P_A}$ be the projection onto the column space of $\mathbf{A}$, thus $\mathbf{P_A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ is of full column rank. Let $\mathbf{Q_A} = \mathbf{I} - \mathbf{P_A}$ be the projection onto the space orthogonal to $\mathbf{A}$.

2.1.1 Models with country-pair heterogeneity

We first consider a model with CPH only:

$$y_{ijt} \quad = \quad \mathbf{X}_{ijt}\alpha + \mathbf{z}_{it}\gamma_1 + \mathbf{z}_{jt}\gamma_2 + \omega_{ij} + \varepsilon_{ijt} \tag{2.1}$$

where y is a bilateral and time variant dependent variable; $\mathbf{X}$ and $\mathbf{z}$'s are covariates; ω is the CPH term; and ε is an idiosyncratic error. Any time invariant variables are absorbed into ω_{ij}.

We rewrite (2.1) by stacking observations over time:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\alpha + \mathbf{z}_i\gamma_1 + \mathbf{z}_j\gamma_2 + \omega_{ij} \cdot \mathbf{i}_T + \varepsilon_{ij} \tag{2.2}$$

where $\mathbf{y}_{ij} = (y_{ij1}, y_{ij2}, ..., y_{ijt}, ..., y_{ijT})'$, $\mathbf{X}_{ij} = (\mathbf{X}'_{ij1}, \mathbf{X}'_{ij2}, .., \mathbf{X}'_{ijt}, .., \mathbf{X}'_{ijT})'$,
$\mathbf{z}_i = (\mathbf{z}'_{i1}, \mathbf{z}'_{i2}, .., \mathbf{z}'_{it}, .., \mathbf{z}'_{iT})'$, $\mathbf{z}_j = (\mathbf{z}'_{j1}, \mathbf{z}'_{j2}, .., \mathbf{z}'_{jt}, .., \mathbf{z}'_{jT})'$,
$\mathbf{i}_T = (1, 1, ..., 1)'$ and $\varepsilon_{ij} = (\varepsilon_{ij1}, \varepsilon_{ij2}, ..., \varepsilon_{ijt}, ..., \varepsilon_{ijT})'$.

The projection onto the space orthogonal to ω_{ij} transforms (2.2) into:

$$\mathbf{Q}_\omega \mathbf{y}_{ij} = \mathbf{Q}_\omega \mathbf{X}_{ij}\alpha + \mathbf{Q}_\omega \mathbf{z}_i\gamma_1 + \mathbf{Q}_\omega \mathbf{z}_j\gamma_2 + \mathbf{Q}_\omega \varepsilon_{ij} \tag{2.3}$$

We start with a review of the equivalence of the LSDV estimator to the within estimator in two-level panel data. The LSDV estimator applied to (2.1) is equivalent to the conventional within estimator where ij is cross-section and t is time as in (2.3). The LSDV estimator for (2.1) can be applied with a two-stage procedure. In the first stage, we run regressions of each of the variables in $(y_{ijt}, \mathbf{X}_{ijt}, \mathbf{z}_{it}, \mathbf{z}_{jt})$ on ω_{ij} and obtain residuals $(y^*_{ijt}, \mathbf{X}^*_{ijt}, \mathbf{z}^*_{it}, \mathbf{z}^*_{jt})$; in the second stage, we run a regression of y^*_{ijt} on $(\mathbf{X}^*_{ijt}, \mathbf{z}^*_{it}, \mathbf{z}^*_{jt})$. Using the FWL theorem, it can be shown that the residuals from the first stage regressions are:

$$y^*_{ijt} = y_{ijt} - \bar{y}_{ij},\ \bar{y}_{ij} = \frac{1}{T}\sum_{t=1}^{T} y_{ijt} \tag{2.4}$$

and, similarly, $\mathbf{X}^*_{ijt} = \mathbf{X}_{ijt} - \bar{\mathbf{X}}_{ij}$, $\mathbf{z}^*_{it} = \mathbf{z}_{it} - \bar{\mathbf{z}}_i$, and $\mathbf{z}^*_{jt} = \mathbf{z}_{jt} - \bar{\mathbf{z}}_j$. Thus, these residuals are algebraically the same as the transformed variables since, for any matrix $\mathbf{A}_t$, $\mathbf{Q}_\omega \mathbf{A}_t = \mathbf{A}_t - \frac{1}{T}\sum_{t=1}^{T} \mathbf{A}_t$. Therefore, we can establish the following Proposition 1.

Proposition 1: *The LS estimator on the transformed model (2.3) is algebraically equivalent to the LSDV estimator on the model with CPH (2.1).*

2.1.2 Models with country-time heterogeneity

We now consider a model with CTH only:

$$y_{ijt} = \mathbf{X}_{ijt}\alpha + \mathbf{z}_{ij}\gamma_3 + u_{it} + v_{jt} + \varepsilon_{ijt} \tag{2.5}$$

Any country-time variant variables are absorbed into the CTH terms, u_{it} and v_{jt}. We obtain the within estimator with CTH by the following sequential projection procedure.

First, we rewrite (2.5) by stacking over the j-dimension as follows:

$$\mathbf{y}_{it} = \mathbf{X}_{it}\alpha + \mathbf{z}_i\gamma_3 + u_{it}\cdot\mathbf{i}_n + \mathbf{v}_t + \varepsilon_{it} \tag{2.6}$$

where $\mathbf{y}_{it} = (y_{i1t}, ..., y_{ijt}, ..., y_{int})'$, $\mathbf{X}_{it} = (\mathbf{X}'_{i1t}, ..., \mathbf{X}'_{ijt}, .., \mathbf{X}'_{int})'$, $\mathbf{z}_i = (\mathbf{z}'_{i1}, ..., \mathbf{z}'_{ij}, .., \mathbf{z}'_{in})'$, $\mathbf{i}_n = (1, 1, ..., 1)'$, $\mathbf{v}_t = (v_{1t}, ..., v_{jt}, ..., v_{nt})'$ and $\varepsilon_{it} = (\varepsilon_{i1t}, ..., \varepsilon_{ijt}, ..., \varepsilon_{int})'$.

We obtain (2.7) after the projection onto the space orthogonal to u_{it}. Let $\mathbf{U}$=$u_{it}\cdot\mathbf{i}_n$.

$$\mathbf{Q}_u\mathbf{y}_{it} = \mathbf{Q}_u\mathbf{X}_{it}\alpha + \mathbf{Q}_u\mathbf{z}_i\gamma_3 + \mathbf{Q}_u\mathbf{v}_t + \mathbf{Q}_u\varepsilon_{it} \tag{2.7}$$

where $\mathbf{Q}_u = \mathbf{I}_n - \mathbf{U}(\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'$.

We further obtain (2.8) by denoting $\mathbf{Q}_u\mathbf{A}$=$\hat{\mathbf{A}}$ for any matrix $\mathbf{A}$.

$$\hat{\mathbf{y}}_{it} \quad = \quad \hat{\mathbf{X}}_{it}\alpha + \hat{\mathbf{z}}_i\gamma_3 + \hat{\mathbf{v}}_t + \hat{\varepsilon}_{it} \tag{2.8}$$

Second, we rewrite the projection onto the space orthogonal to the u_{it}-transformed (2.8) in the non-stacked form of (2.9).

$$\hat{y}_{ijt} \quad = \quad \hat{\mathbf{X}}_{ijt}\alpha + \hat{\mathbf{z}}_{ij}\gamma_3 + \hat{v}_{jt} + \hat{\varepsilon}_{ijt} \tag{2.9}$$

We now consider the projection onto the space orthogonal to $\hat{v}_{jt}$ and obtain (2.10) after the projection for the equation that stacked (2.9) over the i-dimension.

$$\mathbf{Q}_{\hat{u}}\hat{\mathbf{y}}_{jt} = \mathbf{Q}_{\hat{u}}\hat{\mathbf{X}}_{jt}\alpha + \mathbf{Q}_{\hat{v}}\hat{\mathbf{z}}_j\gamma_3 + \mathbf{Q}_{\hat{u}}\hat{\varepsilon}_{jt} \tag{2.10}$$

where $\hat{\mathbf{y}}_{jt} = (\hat{y}_{1jt},...,\hat{y}_{ijt},...,\hat{y}_{mjt})'$, $\hat{\mathbf{X}}_{jt} = (\hat{\mathbf{X}}'_{1jt},....,\hat{\mathbf{X}}'_{ijt},..,\hat{\mathbf{X}}'_{mjt})'$, $\hat{\mathbf{z}}_j = (\hat{\mathbf{z}}'_{1j},...,\hat{\mathbf{z}}'_{ij},..,\hat{\mathbf{z}}'_{mj})'$, $\mathbf{i}_m = (1,1,...,1)'$ and $\hat{\varepsilon}_{jt} = (\hat{\varepsilon}_{1jt},...,\hat{\varepsilon}_{ijt},...,\hat{\varepsilon}_{mjt})'$.

We further obtain (2.11) by denoting $\mathbf{Q}_{\hat{u}}\mathbf{A}=\acute{\mathbf{A}}$ for any matrix $\mathbf{A}$.

$$\acute{\mathbf{y}}_{jt} = \acute{\mathbf{X}}_{jt}\alpha + \acute{\mathbf{z}}_j\gamma_3 + \acute{\varepsilon}_{jt} \tag{2.11}$$

Finally, we can rewrite (2.11) in the non-stacked form of (2.12).

$$\acute{y}_{ijt} = \acute{\mathbf{X}}_{ijt}\alpha + \acute{\mathbf{z}}_{ij}\gamma_3 + \acute{\varepsilon}_{ijt} \tag{2.12}$$

The LSDV estimator applied to (2.5) can be implemented with a two-stage procedure. In the first stage, we run regressions of each of the variables in $(y_{ijt}, \mathbf{X}_{ijt}, \mathbf{z}_{ij})$ on (u_{it}, v_{jt}); in the second stage, we run a regression of $y^{\#}_{ijt}$ on $(\mathbf{X}^{\#}_{ij,t}, \mathbf{z}^{\#}_{ij})$, where $y^{\#}_{ijt}$, $\mathbf{X}^{\#}_{ij,t}$ and $\mathbf{z}^{\#}_{ij}$ are residuals from the first stage regressions. Applying the FWL theorem, it can be shown that $y^{\#}_{ijt}$ and $\mathbf{X}^{\#}_{ijt}$ are equal to:

$$y^{\#}_{ijt} = y_{ijt} - \check{y}_{it} - \frac{1}{n}\sum_{i=1}^{n}(y_{ijt} - \check{y}_{it}) \tag{2.13}$$

$$\mathbf{X}^{\#}_{ijt} = \mathbf{X}_{ijt} - \check{\mathbf{X}}_{it} - \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_{ijt} - \check{\mathbf{X}}_{it}) \tag{2.14}$$

$$\mathbf{z}^{\#}_{ij} = \mathbf{z}_{ij} - \check{\mathbf{z}}_i - \frac{1}{n}\sum_{i=1}^{n}(\mathbf{z}_{ij} - \check{\mathbf{z}}_i) \tag{2.15}$$

where $\check{y}_{it} = \frac{1}{n}\sum_{j=1}^{n} y_{ijt}$, $\check{\mathbf{X}}_{it} = \frac{1}{n}\sum_{j=1}^{n}\check{\mathbf{X}}_{ijt}$ and $\check{\mathbf{z}}_{it} = \frac{1}{n}\sum_{j=1}^{n}\check{\mathbf{z}}_{ijt}$.

It is shown in the appendix that

$$\acute{y}_{ijt} = y_{ijt} - \check{y}_{it} - \frac{1}{n}\sum_{i=1}^{n}(y_{ijt} - \check{y}_{it}) \tag{2.16}$$

$$\acute{\mathbf{X}}_{ijt} = \mathbf{X}_{ijt} - \check{\mathbf{X}}_{it} - \frac{1}{n}\sum_{i=1}^{n}(\mathbf{X}_{ijt} - \check{\mathbf{X}}_{it}) \tag{2.17}$$

$$\acute{\mathbf{z}}_{ij} = \mathbf{z}_{ij} - \check{\mathbf{z}}_i - \frac{1}{n}\sum_{i=1}^{n}(\mathbf{z}_{ij} - \check{\mathbf{z}}_i) \tag{2.18}$$

Once again, we show that the residuals, (2.13), (2.14) and (2.15), are algebraically identical to the transformed variables, (2.16) , (2.17) and (2.18), respectively, and, therefore, can establish the following Proposition 2.

Proposition 2: *The LS estimator on the transformed model (2.12) is algebraically equivalent to the LSDV estimator on the model with CTH (2.5).*

If we switch the order of projections by starting with the projection onto the space orthogonal to u_{it} first and then to v_{jt}, we will obtain the same within transformation for balanced data. However, the order of projections leads to different within transformations for unbalanced data. This issue is discussed in Section 3.

2.1.3 Models with country-pair and country-time heterogeneity

We now consider a model with both CPH and CTH:

$$y_{ijt} \quad = \quad \mathbf{X}_{ijt}\alpha + u_{it} + v_{jt} + \omega_{ij} + \varepsilon_{ijt} \tag{2.19}$$

We obtain the model for the within estimator by the following sequential projection procedure: first onto the space orthogonal to ω_{ij}, then to v_{jt} (more precisely its transformed version), and lastly to u_{it} (also transformed). The order of projections does not matter for balanced panel data.

First, we rewrite (2.19) by stacking over the t-dimension as follows:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\alpha + \mathbf{u}_i + \mathbf{v}_j + \omega_{ij}\mathbf{i}_T + \varepsilon_{ij} \tag{2.20}$$

where $\mathbf{y}_{ij} = (y_{ij1}, ..., y_{ijt}, ..., y_{ijT})'$, $\mathbf{X}_{it} = (\mathbf{X}'_{ij1}, ..., \mathbf{X}'_{ijt}, .., \mathbf{X}'_{ijT})'$, $\mathbf{u}_i = (u_{i1}, ..., u_{it}, .., u_{iT})'$, $\mathbf{v}_j = (v_{j1}, ..., v_{jt}, .., v_{jT})'$, $\mathbf{i}_T = (1, 1, .., 1)'$ and $\varepsilon_{ij} = (\varepsilon_{ij1}, ..., \varepsilon_{ijt}, ..., \varepsilon_{ijT})'$.

The projection onto the space orthogonal to ω_{ij} transforms (2.20) into:

$$\mathbf{Q}_{\omega}\mathbf{y}_{ij} = \mathbf{Q}_{\omega}\mathbf{X}_{ij}\alpha + \mathbf{Q}_{\omega}\mathbf{u}_i + \mathbf{Q}_{\omega}\mathbf{v}_j + \mathbf{Q}_{\omega}\varepsilon_{ij} \tag{2.21}$$

We can rewrite (2.21) with the notation of $\mathbf{Q}_{\omega}\mathbf{A}=\mathbf{A}^*$ for any matrix $\mathbf{A}$ as (2.22) and as (2.23) by stacking over the i-dimension.

$$y^*_{ijt} \quad = \quad \mathbf{X}^*_{ijt}\alpha + u^*_{it} + v^*_{jt} + \varepsilon^*_{ijt} \tag{2.22}$$

$$\mathbf{y}^*_{jt} \quad = \quad \mathbf{X}^*_{jt}\alpha + \mathbf{u}^*_t + v^*_{jt}\mathbf{i}_m + \varepsilon^*_{jt} \tag{2.23}$$

Next the projection onto the space orthogonal to v^*_{jt} transforms (2.23) into:

$$\mathbf{Q}_{v*}\mathbf{y}^*_{jt} = \mathbf{Q}_{v*}\mathbf{X}^*_{jt}\alpha + \mathbf{Q}_{v*}\mathbf{u}^*_t + \mathbf{Q}_{v*}\varepsilon^*_{jt} \tag{2.24}$$

where $\mathbf{Q}_{v*} = \mathbf{I}_m - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'$ and $\mathbf{V} = v^*_{jt} \cdot \mathbf{i}_m$.

We can rewrite (2.24) with the notation of $\mathbf{Q}_{v*}\mathbf{A}$=$\mathbf{A}^{\#\#}$ for any matrix $\mathbf{A}$ and the unstacked form as (2.25):

$$y^{\#\#}_{ijt} = \mathbf{X}^{\#\#}_{ijt}\alpha + u^{\#\#}_{it} + \varepsilon^{\#\#}_{ijt} \tag{2.25}$$

Finally, the projection onto the space orthogonal to $u^{\#\#}_{it}$ transforms (2.25) into:

$$\mathbf{Q}_{u\#\#}\mathbf{y}^{\#\#}_{it} = \mathbf{Q}_{u\#\#}\mathbf{X}^{\#\#}_{it}\alpha + \mathbf{Q}_{u\#\#}\varepsilon^{\#\#}_{it} \tag{2.26}$$

where $\mathbf{Q}_{u\#\#} = \mathbf{I}_n - \mathbf{U}(\mathbf{U}'\mathbf{U})\mathbf{U}'$ and $\mathbf{U} = u^{\#\#}_{it} \cdot \mathbf{i}_n$.

We can rewrite (2.26) with the notation of $\mathbf{Q}_{u\#\#}\mathbf{A}$=$\widetilde{\mathbf{A}}$ for any matrix $\mathbf{A}$ and the unstacked form as (2.27).

$$\widetilde{y}_{ijt} = \widetilde{\mathbf{X}}_{ijt}\alpha + \widetilde{\varepsilon}_{ijt} \tag{2.27}$$

where $\widetilde{y}_{ijt}$ can be elaborated as:

$$\widetilde{y}_{ijt} = y^{\#\#}_{ijt} - \frac{1}{n}\sum_{j=1}^{n} y^{\#\#}_{ijt};\ y^{\#\#}_{ijt} = y^*_{ijt} - \frac{1}{m}\sum_{i=1}^{m} y^*_{ijt}$$

or

$$\widetilde{y}_{ijt} = y^*_{ijt} - \frac{1}{m}\sum_{i=1}^{m} y^*_{ijt} - \frac{1}{n}\sum_{j=1}^{n} y^*_{ijt} + \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m}\sum_{i=1}^{m} y^*_{ijt} \tag{2.28}$$

In (2.28), unobserved heterogeneity (u_{it}, v_{jt}, ω_{ij}) is eliminated through a covariance transformation to time demeaned variables (Hsiao, 2003, Chapter 3). Demeaning over time eliminates ω_{ij} and a covariance transformation of (2.28) eliminates u_{it} and v_{jt}.

In (2.19), we can apply the LSDV estimator in three steps. First, it can be observed that (2.19) is a special case of (2.1) in that $\mathbf{z}_{it} = u_{it}$ and $\mathbf{z}_{jt} = v_{jt}$. Therefore, by Proposition 1, we can establish that applying the LSDV estimator to (2.19) is equivalent to applying the LS estimator to the following transformed model:

$$y^*_{ijt} = \mathbf{X}^*_{ijt}\alpha + u^*_{it} + v^*_{jt} + \varepsilon^*_{ijt} \tag{2.29}$$

where y^*_{ijt}, $\mathbf{X}^*_{ijt}$, u^*_{it} and v^*_{jt} are residuals from the first stage regressions in which each of the variables are regressed against ω_{ij}.

Second, it can be observed that (2.29) is also a special case of (2.5) because u^*_{it} and v^*_{jt} are of the forms of country-time dummies. Therefore, by Proposition 2, we can establish that applying the LSDV estimator to (2.29) is equivalent to applying the LS estimator to the transformed model (2.27). This is because, by applying the FWL theorem, it can be shown that residuals of $y^{\$}_{ijt}$ and $\mathbf{X}^{\$}_{ij,t}$ from the regressions on dummy variables for u_{it}, v_{jt} and ω_{ij} are

$$y^{\$}_{ijt} = y^*_{ijt} - \frac{1}{m}\sum_{i=1}^{m} y^*_{ijt} - \frac{1}{n}\sum_{j=1}^{n} y^*_{ijt} + \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m}\sum_{i=1}^{m} y^*_{ijt} \tag{2.30}$$

$$\mathbf{X}^{\$}_{ijt} = \mathbf{X}^*_{ijt} - \frac{1}{m}\sum_{i=1}^{m} \mathbf{X}^*_{ijt} - \frac{1}{n}\sum_{j=1}^{n} \mathbf{X}^*_{ijt} + \frac{1}{n}\sum_{j=1}^{n}\frac{1}{m}\sum_{i=1}^{m} \mathbf{X}^*_{ijt} \tag{2.31}$$

From (2.28) and (2.30), we obtain that the residuals of $y^{\$}_{ijt}$ and $\mathbf{X}^{\$}_{ij,t}$ are equivalent to the transformed variables, $\widetilde{y}_{ijt}$ and $\widetilde{\mathbf{X}}_{ijt}$, respectively, in (2.27). Here we can establish the following Proposition 3.

Proposition 3: *The LS estimator on the transformed model (2.27) is algebraically equivalent to the LSDV estimator on the model with CPH and CTH (2.19).*

A more elaborated proof of Proposition 3 is given in the appendix. In conclusion, our proposed within estimator can control for endogeneity problems due to unobserved CPH and CTH.

2.2 Inference and degrees of freedom adjustment

Since the incidential parameters problem causes the standard erorr of the within estimator to be biased (Neyman and Scott, 1948; Lancaster, 2000, among others), the degrees of freedom have to be adjusted. We suggest a simple adjustment that extends the standard error adjustment for two-level panel data and is easily implementable using the regression result output from standard statistical packages, such as Stata and SAS. We verify the validity of our proposed adjustment for the standard error estimate by Monte Carlo analysis.

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^2_ε as:

$$\hat{\sigma}^2_\varepsilon = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{nmT-k} \tag{2.32}$$

instead of the correct one:

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{(n-1)(m-1)(T-1)-k} \tag{2.33}$$

where we lose one degree of freedom for each dimension since we use demeaned variables and $\widehat{\widetilde{\varepsilon}}_{ijt}$ are the residuals from the OLS regression.

Without degrees of freedom adjustment the standard error estimates will be underestimated so that the inference will tend to over-reject the null hypothesis.

The degrees of freedom adjustment for a simple t-test of a univariate case where the null is $H_0 : \alpha_1 = 0$ can be easily done. The statistic that is used for inference after OLS regression of $\widetilde{y}_{ijt}$ on $\widetilde{\mathbf{X}}_{ijt}$ is (2.34), but we should instead use (2.35). We can obtain t^{**} simply by multiplying $\sqrt{\frac{(n-1)(m-1)(T-1)-k}{nmT-k}}$ to t^{*}.

$$t^{*} = \frac{\hat{\alpha}_{1,ols}-0}{\sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{nmT-k}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})_{11}}} \tag{2.34}$$

$$t^{**} = \frac{\hat{\alpha}_{1,ols}-0}{\sqrt{\frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{(n-1)(m-1)(T-1)-k}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})_{11}}} \tag{2.35}$$

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

For the unbalanced data case 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:

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{(nmT-a_1-k)} \tag{2.36}$$

instead of the correct one:

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{t=1}^{T}\widehat{\widetilde{\varepsilon}}_{ijt}}{a_2\cdot[(n-1)(m-1)(T-1)-k]} \tag{2.37}$$

Thus, the degrees of freedom adjustment for a simple t-test of a univariate case requires to multiply $\sqrt{\frac{a_2[(n-1)(m-1)(T-1)-k]}{nmT-a_1-k}}$ to obtain valid test statistics. Similarly, the degrees of freedom adjustment for a standard error requires to multiply the OLS t statistics with $\sqrt{\frac{nmT-a_1-k}{a_2[(n-1)(m-1)(T-1)-k]}}$.

3 Fixed effects transformation for unbalanced three-level data

3.1 Intrinsic and extrinsic missing data

For most bilateral data, a country does not interact with itself by nature, e.g. a country does not give foreign aid to itself, or cannot be a colony of itself. In such cases, y_{iit} and X_{iit} are undefined or 'missing'. We label them as *intrinsic* missing data. In addition, there can be genuine missing data for any country pair ij ($i \neq j$) for any time t due to data reporting errors or collection problems. We label them as *extrinsic* missing data.

For extrinsic missing data, if data is missing completely at random (MCAR), the LSDV estimator is consistent while the within estimator is inconsistent. For intrinsic missing data, although it is not random, we can still obtain a consistent estimate from the within estimator[6] because the 'missing' observations are indeed true observations, thus no imputation is plausible. Therefore, if data is missing completely at random, intrinsic and extrinsic missing data raise the same issue as far as the within transformations are concerned.

With missing data, we can rewrite our benchmark model as follows:

$$s_{ijt}y_{ijt} = s_{ijt}\mathbf{X}_{ijt}\alpha + s_{ijt}u_{it} + s_{ijt}v_{jt} + s_{ijt}\omega_{ij} + s_{ijt}\varepsilon_{ijt} \tag{3.1}$$

where s_{ijt} is a sample indicator which has value 1 if we observe both y_{ijt} and $\mathbf{X}_{ijt}$ and 0 otherwise.

For each of the response (y_{ijt}) and explanatory ($\mathbf{X}_{ijt}$) variables, we calculate the mean only using the observations with balanced data. For instance, in the time demeaning process $\bar{y}_{ij}$ is redefined as $\bar{y}_{ij} = \frac{1}{T_{ij}}\sum_{t=1}^{T} y_{ijt}$, where $T_{ij} = \sum_{t=1}^{T} s_{ijt}$; in the export-country demeaning process, $\bar{y}_{it}$ is defined as $\bar{y}_{it} = \frac{1}{N_{it}}\sum_{j=1}^{N} y_{ijt}$, where $N_{it} = \sum_{j=1}^{N} s_{ijt}$; and in the import-country demeaning process, $\bar{y}_{jt}$ is defined as $\bar{y}_{jt} = \frac{1}{N_{jt}}\sum_{i=1}^{N} y_{ijt}$, where $N_{jt} = \sum_{i=1}^{N} s_{ijt}$.

For balanced panel data, the demeaning orders over dimensions i, j, t have no effect on the finite sample estimates. However, for unbalanced panel data, depending upon the order of demeaning, each within estimator of a different demeaning order provides distinct estimates in finite samples. There are potentially six distinct within estimators depending upon the order of demeaning. The six distinct orders over three dimensions i, j, t are $i-j-t$, $i-t-j$, $j-i-t$, $j-t-i$, $t-i-j$, and $t-j-i$.

3.2 The effect of missing data to the within estimator

When can missing data be ignored? In this section, we provide conditions for the within estimator under which we can ignore the fact that certain variables have missing observations. As an exam-

[6] As the number of countries increases, the effects from intrinsic missing data go to zero.

ple, we focus on the transformations with the demeaning sequences of $t-i-j$. The conditions for the other five within estimators can be derived in similar ways.

3.2.1 Unobserved heterogeneity after the within transformation

For unbalanced data due to missing observations, $t-i-j$ sequential demeaning cannot eliminate u_{it} and v_{jt} completely, but can eliminate ω_{ij} completely. The first order demeaning can eliminate unobserved heterogeneity completely. More specifically, the i-dimensional demeaning eliminates jt-variant unobserved heterogeneity, the j-dimensional demeaning eliminates it-variant unobserved heterogeneity, and the t-dimensional demeaning eliminates ij-variant unobserved heterogeneity, as follows:

$$s_{ijt}\omega_{ij} - \frac{1}{T_{ij}}\Sigma_{t=1}^{T} s_{ijt}\omega_{ij} = s_{ijt}\omega_{ij} - \frac{1}{T_{ij}}T_{ij}s_{ij}\omega_{ij} = 0$$

where the last inequality holds since $s_{ijt} = s_{ij}$, as the sample indicator for ω_{ij} does not depend on t.

$$s_{ijt}u_{it} - \frac{1}{N_{it}}\Sigma_{j=1}^{N} s_{ijt}u_{it} = s_{ijt}u_{it} - \frac{1}{N_{it}}N_{it}s_{it}u_{it} = 0$$

where, similarly, the last inequality holds since $s_{ijt} = s_{it}$, as the sample indicator for u_{it} does not depend on j. And finally, we have

$$s_{ijt}v_{jt} - \frac{1}{N_{jt}}\Sigma_{i=1}^{N} s_{ijt}v_{it} = s_{ijt}v_{jt} - \frac{1}{N_{jt}}N_{jt}s_{jt}v_{jt} = 0$$

where, similarly, the last inequality holds since $s_{ijt} = s_{jt}$, as the sample indicator for v_{jt} does not depend on i.

On the other hand, the second and third order demeaning can only partially eliminate unobserved heterogeneity. Incomplete elimination of u_{it} and v_{jt} occurs because non-first order summations over i, j, t provide distinct values according to the order of summation when there are missing observations in the i and j dimensions. Equations (3.2), (3.3), and (3.4) provide demeaned values for unobserved heterogeneity for ω_{ij}, u_{it} and v_{jt}, respectively, after the $t-i-j$ within transformation.

$$
\begin{aligned}
\widetilde{\omega}_{ij} &= s_{ijt}\omega_{ij} - \frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}\omega_{ij} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}s_{ijt}\omega_{ij} - \frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}\omega_{ij} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}\omega_{ij} \\
&\quad + \frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}\omega_{ij} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}\omega_{ij} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}\omega_{ij} \qquad (3.2) \\
&= s_{ij}\omega_{ij} - s_{ij}\omega_{ij} - \frac{1}{N_i}\Sigma_{j=1}^{N}s_{ij}\omega_{ij} - \frac{1}{N_j}\Sigma_{i=1}^{N}s_{ij}\omega_{ij} + \frac{1}{N_i}\Sigma_{j=1}^{N}s_{ij}\omega_{ij} \\
&\quad + \frac{1}{N_j}\Sigma_{i=1}^{N}s_{ij}\omega_{ij} + \frac{1}{N_i}\Sigma_{j=1}^{N}\frac{1}{N_j}\Sigma_{i=1}^{N}s_{ij}\omega_{ij} - \frac{1}{N_i}\Sigma_{j=1}^{N}\frac{1}{N_j}\Sigma_{i=1}^{N}s_{ij}\omega_{ij} \\
&= 0
\end{aligned}
$$

where, for the second equality in (3.2), we use $\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}\omega_{ij} = \frac{1}{T_{ij}}T_{ij}\cdot s_{ijt}\omega_{ij}$, and for ij-variant variables s_{ijt}, N and T do not depend on t (i.e. $s_{ijt} = s_{ij}$,$N_{it} = N_i$, $N_{jt} = N_j$).

$$
\begin{aligned}
\widetilde{v}_{jt} &= s_{ijt}v_{jt} - \frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}s_{ijt}v_{jt} - \frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}v_{jt} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} \\
&\quad + \frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}v_{jt} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} \\
&= s_{ijt}v_{jt} - \frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}s_{ijt}v_{jt} - s_{ijt}v_{jt} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} \\
&\quad + \frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}s_{ijt}v_{jt} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}v_{jt} \qquad (3.3) \\
&= \left(-\frac{1}{T_j}\Sigma_{t=1}^{T}s_{jt}v_{jt} + \frac{1}{N_t}\Sigma_{j=1}^{N}\frac{1}{T_j}\Sigma_{t=1}^{T}s_{jt}v_{jt}\right) - \left(-\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_j}\Sigma_{t=1}^{T}s_{jt}v_{jt} + \frac{1}{N_t}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_j}\Sigma_{t=1}^{T}s_{jt}v_{jt}\right)
\end{aligned}
$$

where, for the second equality in (3.3), we make use of the facts that $\frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}v_{jt} = \frac{1}{N_{jt}}N_{jt}\cdot s_{ijt}v_{jt} = s_{ijt}v_{jt}$, and that for jt-variant variables s_{ijt}, N and T do not depend on i (i.e. $s_{ijt} = s_{jt}$, $N_{it} = N_t$, $T_{ij} = T_j$). In the last line in (3.3), the first term is not equal to zero if there are missing observations in the j dimension and the second term is not equal to zero if there are missing observations in the i and j dimensions.

$$
\begin{aligned}
\widetilde{u}_{it} &= s_{ijt}u_{it} - \frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}s_{ijt}u_{it} - \frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}u_{it} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} \\
&\quad + \frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}u_{it} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} \\
&= s_{ijt}u_{it} - \frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} - s_{ijt}u_{it} - \frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}u_{it} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} \\
&\quad + \frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}s_{ijt}u_{it} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_{ij}}\Sigma_{t=1}^{T}s_{ijt}u_{it} \qquad (3.4) \\
&= (-\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it} + \frac{1}{N_t}\Sigma_{i=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it}) + (-\frac{1}{N_t}\Sigma_{i=1}^{N}s_{it}u_{it} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_t}\Sigma_{i=1}^{N}s_{it}u_{it}) \\
&\quad + (\frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it} - \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_t}\Sigma_{i=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it}) \\
&= (-\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it} + \frac{1}{N_t}\Sigma_{i=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it}) - (-\frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it} + \frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_t}\Sigma_{i=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it})
\end{aligned}
$$

where, for the second equality in (3.4), we make use of the facts that $\frac{1}{N_{it}}\Sigma_{j=1}^{N}s_{ijt}u_{it} = \frac{1}{N_{it}}N_{it}\cdot s_{ijt}u_{it} = s_{ijt}u_{it}$, and that for it-variant variables s_{ijt}, N and T do not depend on j (i.e. $s_{ijt} = s_{it}$, $N_{jt} = N_t$, $T_{ij} = T_i$). In the last line in (3.4), the first term is not equal to zero if there are missing observations in the i dimension and the second term is not equal to zero if there are missing observations in the i and j dimensions.

Equations (3.2), (3.4), and (3.3) show that the $t-i-j$ demeaning eliminates ω_{ij} completely but $\widetilde{v}_{jt}$ and $\widetilde{u}_{it}$ remain sources of biases. For the one parameter model case, the bias of the within estimator is $\frac{cov(\widetilde{x}_k,\widetilde{u})}{var(\widetilde{x}_k)} + \frac{cov(\widetilde{x}_k,\widetilde{v})}{var(\widetilde{x}_k)}$. It can be also shown that $i-j-t$ and $i-t-j$ transformations can eliminate only v_{jt} completely and transformed ω_{ij} and u_{it} remain sources of biases, while $j-i-t$ and $j-t-i$ transformations can eliminate u_{it} completely but transformed ω_{ij} and v_{jt} remain sources of biases. The biases for the other five within estimators can be derived in similar ways.

Our primary concern is how much bias is left after transformations. Therefore, next we quantify the bias of our within estimators and compare its magnitude to those of alternative estimators.

3.2.2 Bias due to incomplete elimination after the within transformation

For the $t-i-j$ transformation, the bias for the coefficient on x_j due to unobserved heterogeneity can be expressed as:

$$plim\hat{\alpha}_{1k} = \alpha_k + \underbrace{\frac{cov(x_k^@, \widetilde{u}+\widetilde{v})}{var(x_k^@)}}_{bias\ due\ to\ incomplete\ elimination} = \frac{cov(x_k^@, \widetilde{u})}{var(x_k^@)} + \frac{cov(x_k^@, \widetilde{v})}{var(x_k^@)} \tag{3.5}$$

where $x_k^@ = \widetilde{Mx_k}$, $M = I - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$, $\mathbf{W} = \widetilde{\mathbf{X}}_{-k}$, and $\widetilde{\mathbf{X}}_{-k}$ include all $\widetilde{x}_1$, $\widetilde{x}_2, ...,$ $\widetilde{x}_{dim(\mathbf{X})}$ except $\widetilde{x}_k$. We denote $\frac{cov(x_k^@,\widetilde{v})}{var(x_k^@)}$ as v_2 since it is bias due to v_{jt} that is not eliminated after the second order i-dimensional demeaning, and $\frac{cov(x_k^@,\widetilde{u})}{var(x_k^@)}$ as u_3 since it is bias due to u_{it} that is not eliminated after the third order j-dimensional demeaning and v_2.

Similarly, for $t-j-i$ within transformation, the bias for the coefficient on x_k due to unobserved heterogeneity ω_{ij}, u_{it} and v_{jt} can be expressed by:

$$plim\hat{\alpha}_{2k} = \alpha_k + \underbrace{\frac{cov(x_k^{@_2}, \breve{u}+\breve{v})}{var(x_k^{@_2})}}_{bias\ due\ to\ incomplete\ elimination} = \frac{cov(x_k^{@_2}, \breve{u})}{var(x_k^{@_2})} + \frac{cov(x_k^{@_2}, \breve{v})}{var(x_k^{@_2})} \tag{3.6}$$

where $x_k^{@_2} = \breve{M}x_k$, and we denote $\frac{cov(x_k^{@_2},\breve{u})}{var(x_k^{@_2})}$ as u_2 and $\frac{cov(x_k^{@_2},\breve{v})}{var(x_k^{@_2})}$ as v_3.

The bias arises because the transformed variables have distinct values according to the order of summations. This is shown in (3.7) and (3.8), where the only difference between $\widetilde{u}$ and $\breve{u}$ is due to

$$\frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{N_t}\Sigma_{i=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it} \neq \frac{1}{N_t}\Sigma_{i=1}^{N}\frac{1}{N_{it}}\Sigma_{j=1}^{N}\frac{1}{T_i}\Sigma_{t=1}^{T}s_{it}u_{it} \tag{3.7}$$

and the difference between $\widetilde{v}$ and $\breve{v}$ is due to

$$\frac{1}{N_t}\Sigma_{j=1}^{N}\frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{T_j}\Sigma_{t=1}^{T}s_{jt}v_{jt} \neq \frac{1}{N_{jt}}\Sigma_{i=1}^{N}\frac{1}{N_t}\Sigma_{j=1}^{N}\frac{1}{T_j}\Sigma_{t=1}^{T}s_{jt}v_{jt} \tag{3.8}$$

3.3 Monte Carlo simulation

The objective of this section is to quantify how much of the initial bias $\frac{cov(x,\omega+u+v)}{var(x)}$ of the pooled OLS estimator is reduced by each of the six within transformations using Monte Carlo (MC) simulation. As our focus is unbalanced panel data, our data generating process (DGP) also includes a missing observations generating process.

3.3.1 Monte Carlo design

The model we use for all simulations in this section is:

$$y_{ijt} = \alpha x_{1ijt} + \beta z_{1it} + \gamma z_{2jt} + u_{it} + v_{jt} + \omega_{ij} + \varepsilon_{ijt} \tag{3.9}$$

where x_1 is an endogenous variable as it is highly correlated with u_{it}, v_{jt} and ω_{ij}; $u_{it} \sim iidN(0,1)$; $v_{jt} \sim iidN(0,1)$; $\omega_{ij} \sim iidN(0,1)$; idiosyncratic error $\varepsilon_{ijt} \sim iidN(0,1)$; and the number of observations is 2,000 with i =1,...,20 , j =1,...,20 and t =1,...,5 .

Our DGP is:

$$x_{1ijt} = \frac{1}{4}[a \cdot u_{it} + b \cdot v_{jt} + c \cdot \omega_{ij}] + \frac{N(0,1)}{4};\ z_{1it} = d \cdot u_{it} + \frac{N(0,1)}{2};\ z_{2jt} = e \cdot v_{jt} + \frac{N(0,1)}{2} \tag{3.10}$$

We consider the following nine sets of values for (a, b, c, d, e): (i) (0,0,1,0,0), (ii) (1,0,0,0,0), (iii) (0,1,0,0,0), (iv) (1,0,1,0,0), (v) (1,1,0,0,0), (vi) (0,1,1,0,0), (vii) ($\frac{1}{2}$,$\frac{1}{4}$,1,0,0), (viii) (1,$\frac{1}{2}$,$\frac{1}{4}$,0,0), and (ix) ($\frac{1}{2}$,1,$\frac{1}{4}$,0,0). For (i), (ii) and (iii), the bias of within estimators can only be caused by one type of unobserved heterogeneity. For (iv), (v) and (vi), the bias is caused by two types of unobserved heterogeneity with equal magnitude. Finally, for (vii), (viii), (ix), the bias is caused by all three types of unobserved heterogeneity with different magnitudes. The last three processes are to examine how the bias of different within estimators changes with the relative importance of different types of unobserved heterogeneity. [7]

Thus, given x_{1ijt} , z_{1it} , z_{2jt} , u_{it} , v_{jt} , ω_{ij} and ε_{ijt}, and with true parameter values for $\alpha = 1$, $\beta = -.5$, and $\gamma = .2$, we obtain y_{ijt} in (3.9). We are interested in consistently estimating α with the proposed within estimators.

Pooled OLS estimator for α is given by:

$$plim\hat{\alpha} = \alpha + \underbrace{\frac{cov(x_1^*, u_{it}^*)}{var(x_1^*)}}_{\text{bias due to } u_{it}} + \underbrace{\frac{cov(x_1^*, v_{jt}^*)}{var(x_1^*)}}_{\text{bias due to } v_{jt}} + \underbrace{\frac{cov(x_1^*, \omega_{ij}^*)}{var(x_1^*)}}_{\text{bias due to } \omega_{ij}} \tag{3.11}$$

where x_1^* is a residual from the regression of x_1 on z_{1it} and z_{2jt}, and we can similarly define u_{it}^*, v_{jt}^*, and ω_{ij}^* as residuals from the generated data using the regressions of these variables on z_{1it} and z_{2jt}. Our primary focus is to quantify how the within transformation increases/reduces the bias in (3.11) and to identify which within estimator causes the smallest bias.

For the $t-i-j$ within transformation, the within estimator for α is given by:

[7] We also consider DGPs which allow z_{1it} and z_{2jt} to be correlated with unobserved heterogeneity and the x_{1ijt}, but the qualitative implications are the same as the results in Table 1 and 2.

$$plim\hat{\alpha}_1 = \alpha + \underbrace{\frac{cov(\widetilde{x}_1^*,\ \widetilde{u}_{it}^* + \widetilde{v}_{jt}^*)}{var(\widetilde{x}_1^*)}}_{\text{bias after transformation}} = \alpha + v_2 + u_3 \tag{3.12}$$

where v_2 is bias due to v_{jt} which is not eliminated completely after the second order i-dimensional demeaning and u_3 is bias due to u_{it} which is not eliminated completely after the third order j-dimensional demeaning. We can obtain $\hat{\alpha}$ for the other five within estimators in similar ways.

In the first MC simulation, missing observations are assigned randomly to y, x, z_1 and z_2 using the following process. We randomly assign a number q from 1 to 2000 for each observation and then randomly draw 500 missing observations (i.e. 25% missing) from a uniform distribution:

$$d_k* = uniform(0,1)*2000+1 \tag{3.13}$$

$$d_k = [d_k^*];\ s_q = 0 \text{ if } d_k = q$$

where $k = 1,2,...,500$ and $[x]$ is the largest integer not greater than x.

In the second MC simulation, we increase the proportion of missing data to 40%.

3.3.2 Results

Table 1 reports the finite sample behavior of the within estimators for α. On the left hand side of the table are sources of biases, on the right hand side are the remaining biases after each of the transformations. For instance, in the first row where the 198% bias is entirely due to u_{it}, the bias is reduced to 30.5% after the $t-i-j$ transformation, and 25.8% after the $t-j-i$ transformation, and so forth. The following implications can be drawn from Table 1:

(i) As long as the direction of bias due to u_{it}, v_{jt} and ω_{ij} is the same, all six within estimators reduce the bias significantly. Even in the worst case (the fifth row), a within estimator reduces the bias by four-fifths, from 264% to 53.2%.

(ii) The extent of bias reduction varies amongst the six estimators. For instance, in the first row, the bias is reduced to as low as 0% or as high as 30.5%.

(iii) It is possible to completely eliminate a single-sourced bias if an appropriate estimator is used. For instance, in the first row where the bias is entirely due to u_{it}, doing the j-dimensional demeaning first can reduce the bias to 0%, while doing it last will reduce the bias least (with 30.5% bias left).

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

(v) Whenever the bias due to ω_{ij} does not dominate those due to either u_{it} or v_{jt}, doing the t-dimensional demeaning later will lead to 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 $i-j-t$ demeaning sequence tends to be larger than $i-t-j$, which in turn tends to be larger than $t-i-j$ etc.

(vi) A corollary of (v) is that, if biases caused by u_{it}, v_{jt} and ω_{ij} are of comparable magnitudes, as at the seventh row in Table 1, doing the t-dimensional demeaning last is more effective in reducing bias.

Table 2 shows the same results for the data with 40% missing proportion. As missing proportion increases, the remaining bias after within transformations also increases regardless of the order of demeaning dimension. However, the relationship between relative importance of biases due to u_{it}, v_{jt} and ω_{ij} and the order of transformation remains the same as the data with 25% missing proportion in Table 1.

In Tables 1 and 2, biases due to u_{it}, v_{jt} and ω_{ij} are in the same direction and thus will not cancel each other. In this scenario, the within estimators estimate $\hat{\alpha}$ with substantially reduced bias compared to the pooled OLS estimator. However, this is not the case when biases caused by different source of unobserved heterogeneity cancel each other as shown in Table 3. In the table, the sum of absolute biases is 90%, but the total is either approximately 0% or $\pm$30%. If biases cancel each other out, we should observe $\hat{\alpha}_{within,p} > \hat{\alpha}_{pooledOLS}$ for some transformation p and $\hat{\alpha}_{within,q} < \hat{\alpha}_{pooledOLS}$ for some other transformation q with $p \neq q$. For instance, in all nine rows of the table, some within estimators have upward biases while other within estimators have downward biases. In some DGPs and some within estimators, the bias of within estimators can be greater than the bias of pooled OLS estimator since biases cancel out each other for OLS estimator.

Table 4 provides R^2 statistics for six within estimators with 25% missing proportion. R^2 statistics could be useful for choosing among the estimators. Specifically, against the normal rule of thumb that a model with a larger R^2 value is preferable, here it is the opposite. It is obvious from Table 4 that, with one exception, the within estimator with smallest R^2 is the one with smallest bias. The intuition behind this result is that when a within transformation eliminates most 'bad variation' of x that is correlated with unobserved heterogeneity, it is also highly likely to eliminate most 'good variation' of x that is uncorrelated with unobserved heterogeneity. However, this is also a rule of thumb based on simulation analysis and therefore does not necessarily hold true in every single case. For instance, in the third row in which ω_{ij} is the sole source of bias, the $j-t-i$ transformation gives the smallest R^2 but the second smallest remaining bias. However, even in this case, the reduction of bias is still very substantial. Furthermore, choosing the within estimator that gives the smallest R^2 value should not cause concern because, with a large sample, R^2 statistics matter less as the within estimator is precise enough.

<Insert Table 1,2,3,4 here>

3.3.3 Inference accuracy

Table 5 provides inference consistency of the t-test under the null hypothesis that the within estimator is consistent. For the pooled OLS estimator, the bias due to unobserved heterogeneity is provided in the first column of Table 5 and is about .80, as the mean of estimate for α is 1.80 while its true value is one. We examine a consistent within estimator for balanced panel data in the second column and two consistent within estimators for unbalanced panel data with 25% of missing data in third and fourth columns. The mean value of the rejection probability for t-test of $H_0 : \alpha = 1$ is reported in the fourth row and its coverage values are reported in the fifth row. The results show that if a within estimator completely eliminates unobserved heterogeneity, the within estimator is consistent and a usual t-test with appropriate degrees of freedom adjustment in section 2.2 provides a valid inference. The estimated mean value of the rejection probability for a t-test of $H_0 : \alpha = 1$ is very close to .05 and its coverage always contains true type-I error .05.These simulation results confirm that our proposed adjustment properly corrects the bias of standard error estimates.

<Insert Table 5 here>

4 An application: The WTO effect on bilateral trade flows

4.1 The modeling framework

In this section, we apply the proposed within estimator to investigate the effects of GATT/WTO membership (denoted as WTO membership hereafter) on bilateral trade flows.

The members of the WTO and its predecessor the General Agreements on Tariffs and Trade (GATT) are required to liberalize their trade regimes toward other members, although exception clauses are allowed in some cases, especially in relation to developing countries. Accordingly, it is anticipated WTO membership would increase trade flows. Surprisingly, Rose (2004), the first empirical study investigating the WTO effects, finds little evidence to support such a proposition. However, several papers point out that the results in Rose (2004) could be tainted by omission of important variables (Subramanian and Wei, 2007b; Liu, 2009; Eicher and Henn, 2011; Roy, 2011).[8] Firstly, Rose (2004) did not account for the multilateral resistance terms introduced in Anderson and van Wincoop (2003), which capture the effects of trade barriers having a larger impact on relatively small countries. Secondly, both trade and WTO membership may be dependent upon country-time specific conditions for which reliable measures are hard to obtain. For example, China was an initial member of GATT in 1948, but then resigned in 1950 in the aftermath of the civil war, and joined the WTO in 2001. The membership status of China seems to be influenced by political factors as well as economic factors over the years, and so were its bilateral trade flows with other countries. Ignoring unobserved bilateral heterogeneity such as cultural and political ties of pair countries can be another source of potential omitted variable bias (Baldwin and Taglioni, 2006). To account for these omitted but unobserved variables, we incorporate both CPH (Magee, 2003; Baldwin and Taglioni, 2006; Baier and Bergstrand, 2007) and CTH terms (Anderson and van Wincoop, 2003; Subramanian and Wei, 2007a; Magee, 2008; Eicher, Henn, and Papageorgiou, 2012) in the gravity model as in (4.1).

The WTO membership effect on bilateral trade flows is estimated as follows:

$$\begin{aligned} ln(T_{ijt}) &= \alpha_0 + \alpha_1 lnY_{it} + \alpha_2 lnY_{jt} + \alpha_3 lny_{it} + \alpha_4 lny_{jt} \\ &\quad + \mathbf{X}_{ijt}\beta + \gamma WTO_{ijt} + \mu_t + \omega_{ij} + u_{it} + v_{jt} + \varepsilon_{ijt} \end{aligned} \tag{4.1}$$

where the WTO_{ijt} dummy is equal to one if both importer i and exporter j were WTO members in year t; Y is real GDP measured in purchasing power parity (PPP) terms; y is real per capita GDP; $\mathbf{X}$ is a vector of other determinants, such as regional trade agreement (RTA_{ijt}), currency unions (CU_{ijt}), general system of preferences (GSP_{ijt} and GSP_{jit}), hostility, alliance, distance, area, border, landlock, island, common language, common religion, colony, colonizer, current colony, current colonizer, common colony, and remoteness;[9] μ_t represents any unobserved global trend in trade and aggregate shocks in each year; ω_{ij} is country-pair heterogeneity; u_{it} and v_{jt} are importer-time and exporter-time heterogeneity, respectively; and ε_{ijt} is an idiosyncratic error term. Some of the terms, such as μ_t, u_{it} and v_{it} will not be included in the same estimation due to multicollinearity.

Our data-set is obtained from Liu (2009). The data-set covers 216 countries from 1948 to 2003;

[8]Data and model specification may also contribute to Rose (2004)'s unexpected results. Subramanian and Wei (2007b) point out the existence of heterogeneity in the GATT/WTO system, e.g. developing countries vs. developed countries; less liberalized sectors vs. more liberalized sectors; old GATT members vs. new members, and find uneven impact of the WTO memberships on trade flows. Furthermore, Tomz, Goldstein, and Rivers (2007) re-define the WTO membership using actual participation rather than formal membership and find the evidence of the positive WTO effect. More recently, literature also argues that omission of zero observations causes sample selection bias and shows that inclusion of zero bilateral trade flows results in the postive WTO effect (Liu, 2009; Konya, Matyas, and Harris, 2011).

[9]The detailed definitions of the explanatory variables are provided in the appendix.

deployment of the LSDV estimator to control for both CPH and CTH will require 46,440 country-pair indicators and 24,192 import-time and export-time indicators as additional regressors.[10] To the best of our knowledge, no previous studies using such large data-sets have controlled for both CPH and CTH. To circumvent the computational challenges in handling such a large number of dummy variables, we use the proposed within estimator.

4.2 The equivalence of the LSDV and within estimators

In this section we demonstrate the validity of our within estimator using the LSDV estimation as the benchmark.

First, we consider a sub-sample with intrinsic missing data only. The sub-sample covers 22 OECD countries from 1951 to 1970.[11] If the data-set were balanced, there would be 9,680 (=22×22×20) observations. In reality, 4.5% of the observations are (intrinsically) missing, reducing the sample size to 9,240. With this reduced data-set, it is computationally feasible to estimate (4.1) using the LSDV estimator as it involves only 650 country-pair and 1,040 country-time dummy variables. This allows us to compare the results from the two estimators.

The results are reported in Table 6. Besides that of WTO, we also report the results for a number of institutional variables for comparison. In columns (1) to (3), we only control for, respectively, time, country-pair, and country-time FEs, while in column (4) we control for all three types of FEs. The purpose of including different FEs is to demonstrate the importance of controlling for both CPH and CTH. As we control for more unobserved heterogeneity, the effect of WTO membership decreases from 0.67 with time FEs only in column (1) to -.78 with both country-time and country pair FEs in column (4). Overall, the estimates for the effect of WTO membership and other institutional variables are very imprecise due to large standard errors.

Columns (5) to (12) report the results from the within estimator. In columns (5) and (6), CPH is not controlled for, and they are different in terms of the demeaning order. From columns (7) to (12), both CPH and CTH are controlled for, and again each of them is different in terms of the demeaning order. The estimates in columns (3), (5) and (6) are comparable to one another as they all have controlled for CTH, and likewise for the estimates in columns (4), (7) to (12), which all have controlled for both CPH and CTH. Our results show that the estimates of both LSDV and the two within estimators are almost identical, and the order of demeaning makes little difference. These results are not surprising given the small proportion of intrinsic missing data.

[10]The maximum number of variables that can be defined in Stata 12.0 is 32,767 and the maximum size of a matrix is 11,000. Containing 70,569 dummy variables in the model inevitably leads to computational problems.

[11]Out of 24 OECD countries in 1975, our sample includes only 22 OECD countries. Belgium and Luxembourg are dropped because they have extrinsic missing observations.

Next we consider another sub-sample with both intrinsic and extrinsic missing observations. The second sub-sample consists of 24 non-OECD countries from 1961 to 1980.[12] We randomly select sample countries from non-OECD countries that have about 30% to 40% missing data. A completely balanced panel would have 11,520 (=24×24×20) observations, but the actual sample has 36.5% of the data missing, leaving only 7,308 observations.

The results are reported in Table 7. The *GSP* variables drop out as no such arrangements exist amongst the sample countries. Regardless of how we control for unobserved heterogeneity, we do not find any significant effect of WTO membership on bilateral trade flows. Overall, the estimates for the effect of WTO membership on bilateral trade flows are very imprecise due to large standard errors. Though the differences between the LSDV and the two within estimators are small in general, the discrepancies amongst them are more noticeable than in Table 6. This is probably because the proportion of missing observations is much more substantial in this second sub-sample.

For the coefficient on WTO, the comparison between the estimates in (1) and (2) implies that $\frac{Cov(x^*,\omega^*)}{Var(x^*)}$ has a negative sign so that, from the comparison of estimates between (2) and (7)-(12), we can infer that $\frac{Cov(x^*,u^*+v^*)}{Var(x^*)}$ is positive. This implies that biases due to CPH and CTH cancel out each other. Using the "smallest R^2 means smallest bias" rule of thumb, the within estimators in (7) and (8) are likely to eliminate most biases in absolute value. However, we cannot determine which of the two estimates is closer to the true parameter value.

<Insert Table 6,7 here>

4.3 Full sample estimation

The estimation results for the within estimator to the full sample of 216 countries from 1948 to 2003 are reported in Table 8. The LSDV estimator is used to produce columns (1) and (2) and the within estimator is used for columns (3) to (8), as the LSDV is not implementable here due to the large number of indicators. Fifty-four per cent of the data are missing, leaving us 1,184,525 observations. To keep the discussion focused, for the within estimator, only the estimates that account for both CPH and CTH are reported.

Unlike in Table 6 and Table 7, the coefficient estimates of WTO in this full sample are statistically significant at the 1% level and the standard errors are much smaller than those in the previous reduced samples. This is because with a much larger sample there is sufficient variation

[12]Non-OECD sample countries include Nicaragua, Bahamas, Bermuda, Belize, Surinam, Qatar, Myanmar, Laos, Vietnam, Angola, Central African Rep., Guinea, Malawi, Mauritania, Rwanda, Seychelles, Sierra Leone, Burkina Faso, Zambia, Fiji, Papua New Guinea, Albania, Bulgaria, and Cuba.

across countries and/or over time to identify the effect of WTO membership. Controlling for CPH reduces bias substantially, but the demeaning order has little influence on the estimation results. Our results clearly show that WTO membership leads to a statistically significant increase in trade flows between two countries even after accounting for both CPH and CTH.

Based on the "smallest R^2" rule of thumb, the results suggest that column (6) gives the least biased estimates. Despite the fact that the within transformation has removed a substantial portion of variation in the variables, due to the large number of observations, the estimation is still precise enough to have small standard errors. According to column (6), if two trading countries have WTO membership, it will increase their bilateral trade flows by as much as 74%, compared to the baseline that none or one of them are WTO members. In summary, using the proposed within estimators, we show that the WTO effect on bilateral trade is not only statistically significant, but also economically large.

<Insert Table 8 here>

5 Conclusion

This paper proposes a within estimator for three-level data. We show that the proposed within estimator is valid with balanced data. We prove the algebraic equivalence between the proposed within estimator and the widely used LSDV estimator, and also examine whether the former can replace the latter specifically for analysis with large data-sets where the latter is not implementable. For unbalanced panel data, we quantify the potential bias of the proposed six distinct within estimators and show that the bias is much smaller than those of other popular alternative estimators. In simulation, the proposed within estimator reduces the bias of pooled OLS by more than four-fifths, even in the worst case. We also provide a method to select a within estimator with the smallest bias using R^2. As an illustration, we apply the proposed within estimator to estimate the WTO membership effect on bilateral flows. Our results indicate that WTO membership has a statistically significant and economically large positive effect on bilateral trade flows, which is consistent with the theory but in contrast to many previous empirical findings that produce little evidence of the WTO effect. We show that the imprecise estimation in previous studies are probably due to small samples and improper control of unobserved heterogeneity. The WTO membership effect application indicates that the within estimator is especially useful when proper account of unobserved heterogeneity is important while the data-set is too large to apply the LSDV estimator.

Table 1: 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 element 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).

Table 2: 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 element 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).

Table 3: 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 out each other. The number of replications is 2,000. $(\cdot,\cdot,\cdot)$ represent common element 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).

Table 4: R^2 and source of 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%
					(.112)	(.103)	(.059)	(.114)	(.068)	(.059)
(0,v_{jt},0)	197%	0%	197%	0%	25.7%	30.4%	30.6%	0%	0%	5.6%
					(.105)	(.110)	(.101)	(.072)	(.061)	(.066)
(0,0,ω_{ij})	200%	0%	0%	200%	0%	0%	5.9%	5.9%	11.6%	11.6%
					(.073)	(.070)	(.065)	(.080)	(.076)	(.073)
(u_{it},0,ω_{ij})	265%	131%	0%	134%	30.9%	26.2%	6.0%	36.0%	17.1%	11.5%
					(.112)	(.103)	(.065)	(.122)	(.084)	(.073)
(u_{it},v_{jt},0)	264%	132%	132%	0%	53.2%	53.2%	31.2%	31.2%	6.1%	6.1%
					(.149)	(.147)	(.101)	(.114)	(.068)	(.066)
(0,v_{jt},ω_{ij})	265%	0%	131%	134%	26.2%	30.9%	36.0%	6.0%	11.6%	17.0%
					(.105)	(.110)	(.109)	(.079)	(.076)	(.081)
(u_{it},v_{jt},ω_{ij})	299%	99%	99%	101%	52.9%	52.9%	35.9%	35.9%	17.0%	17.0%
					(.149)	(.147)	(.101)	(.114)	(.068)	(.066)

Note: The number of replications is 2,000. R^2 is provided in parentheses. $(\cdot,\cdot,\cdot)$ represent common element 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}. u_{it}, v_{jt}, and ω_{ij} are sources of bias as in equation (3.11).

Table 5: Inference accuracy

	bias	balanced	unbalanced-jit	unbalanced-jti
(u_{it},0,0), mean of bias	.79	.00	.00	.00
(u_{it},0,0), SD	(.043)	(.053)	(.062)	(.063)
(u_{it},0,0), coverage of coefficient		(.997,1.001)	(.997,1.002)	(.997,1.002)
(u_{it},0,0), rejection% mean		.051	.056	.050
(u_{it},0,0), rejection% coverage		(.042,.061)	(.046,.066)	(.040,.059)
	bias	balanced	unbalanced-ijt	unbalanced-itj
(0,v_{jt},0), mean of bias	.79	.00	.00	.00
(0,v_{jt},0), SD	(.044)	(.054)	(.065)	(.065)
(0,v_{jt},0), coverage of coefficient		(.998,1.003)	(.998,1.003)	(.998,1.003)
(0,v_{jt},0), rejection% mean		.049	.052	.057
(0,v_{jt},0), rejection% coverage		(.039,.059)	(.041,.062)	(.046,.067)
	bias	balanced	unbalanced-tij	unbalanced-tji
(0,0,ω_{ij}), mean of bias	.80	.00	.00	.00
(0,0,ω_{ij}), SD	(.040)	(.053)	(.063)	(.063)
(0,0,ω_{ij}), coverage of coefficient		(.998,1.002)	(.997,1.003)	(.997,1.003)
(0,0,ω_{ij}), rejection% mean		.056	.046	.045
(0,0,ω_{ij}), rejection% coverage		(.046,.066)	(.037,.055)	(.036,.054)

Note: The number of replications is 2,000. $(\cdot,\cdot,\cdot)$ represent common element 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}. Balanced column represents the estimates of the within estimator for balanced panel data.

Table 6: The WTO effect on trade flows - OECD countries

	LSDV estimator				Within estimator							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
WTO	.67	.37***	.26	-.78	.27	.26	-.78	-.78	-.78	-.78	-.78	-.78
	(.22)	(.17)	(.80)	(.75)	(.79)	(.79)	(.69)	(.71)	(.71)	(.69)	(.69)	(.71)
RTA	.31**	.08	-.23	.18	-.23	-.23	.18	.18	.18	.18	.18	.18
	(.15)	(.10)	(.20)	(.18)	(.18)	(.20)	(.17)	(.17)	(.17)	(.17)	(.17)	(.17)
CU	1.46	-.68	1.50	-.59	1.50	1.50	-.59	-.59	-.59	-.59	-.59	-.59
	(1.63)	(1.06)	(1.36)	(.62)	(1.36)	(1.36)	(.60)	(.59)	(.59)	(.60)	(.60)	(.59)
GSP_i	-.88	.10	-.16	-.04	-.17	-.17	-.04	-.04	-.04	-.04	-.04	-.04
	(.46)	(.26)	(.97)	(.51)	(.97)	(.98)	(.49)	(.49)	(.49)	(.49)	(.49)	(.49)
GSP_j	-.03	-.10	2.05	-.90	-2.04	2.04	-.90	-.90	-.90	-.90	-.90	-.90
	(.61)	(.25)	(.97)	(.82)	(.97)	(.96)	(.79)	(.79)	(.79)	(.79)	(.79)	(.79)
R^2	.560	-	.647	-	.106	.106	.003	.003	.003	.003	.003	.003
Time FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Country-pair FEs	no	yes	no	yes	no	no	yes	yes	yes	yes	yes	yes
Country-time FEs	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Demean order	no	no	no	no	ji	ij	tij	tji	jti	itj	ijt	jit
Num of Obs	9,240											

Note: Unbalanced panel data only occurs due to intrinsic missing data. Cluster (country-pairs) robust standard errors are reported in parentheses. Other control variables include $\ln(distance_{ijt})$, $\ln(Y_{it})$, $\ln(Y_{jt})$, commom language, common colony, colonizer, colony, shared borders, shared religion, common legal origin, alliance and hostility. These control variables are identical to the ones in Liu(2009). The details for the description of variables are provided in Liu(2009). ***, ** and * mean that the estimated coefficient is statistically significant at the 1%, 5%, and 10% level, respectively. For the within estimates, cluster-robust standard errors are reported with the degrees of freedom adjustment in section 2.2. The column (2) is implemented with the within estimator in two-level data index and the column (4) is implemented with the within estimator in two-level data index and country-time dummy variables.

Table 7: The WTO effect on trade flows - Non OECD countries

	LSDV estimator				Within estimator							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
WTO	-.69	.60	-.54	-.32	-.52	-.54	-.30	-.26	-.23	-.29	-.29	-.24
	(.44)	(.51)	(.55)	(.67)	(.55)	(.55)	(.67)	(.66)	(.67)	(.67)	(.67)	(.67)
RTA	.80	-1.96	.97	-1.52	1.08	.99	-1.58	-1.50	-1.50	-1.55	-1.53	-1.48
	(1.46)	(1.03)	(1.09)	(1.03)	(1.11)	(1.11)	(1.10)	(1.07)	(1.07)	(1.09)	(1.09)	(1.08)
CU	2.23	1.76	2.33**	2.19	2.43**	2.31**	2.15	2.16	2.12	2.19	2.19	2.15
	(1.36)	(1.57)	(1.02)	(1.42)	(1.04)	(1.04)	(1.35)	(1.46)	(1.44)	(1.35)	(1.34)	(1.44)
R^2	.140	-	.440	-	.166	.168	.006	.006	.007	.007	.007	.007
Time FEs	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Country-pair FEs	no	yes	no	yes	no	no	yes	yes	yes	yes	yes	yes
Country-time FEs	no	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Demean order	no	no	no	no	ji	ij	tij	tji	jti	itj	ijt	jit
Num of Obs	7,308											

Note: Unbalanced panel data occurs due to both intrinsic and extrinsic missing data. The rest is the same as Table 5. Cluster (country-pairs) robust standard errors are reported in parentheses. ***, ** and * mean that the estimated coefficient is statistically significant at the 1%, 5%, and 10% level, respectively. For the within estimates, cluster-robust standard errors are reported with the degrees of freedom adjustment in section 2.2. The column (2) is implemented with the within estimator in two-level data index and the column (4) is implemented with the within estimator in two-level data index and country-time dummy variables.

Table 8: The WTO effect on trade flows

	LSDV estimator		Within estimator					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WTO	1.30***	1.07***	.79***	.82***	.76***	.74***	.75***	.75***
	(.04)	(.04)	(.06)	(.06)	(.06)	(.06)	(.06)	(.06)
RTA	1.27***	.67***	.61***	.62***	.64***	.61***	.64***	.64***
	(.07)	(.07)	(.07)	(.07)	(.07)	(.07)	(.07)	(.07)
CU	1.58***	3.53***	2.80***	2.77***	2.67***	2.72***	2.62***	2.62***
	(.15)	(.28)	(.25)	(.25)	(.24)	(.24)	(.24)	(.24)
GSP_i	2.88***	1.25***	1.68***	1.71***	1.83***	1.60***	1.80***	1.78***
	(.07)	(.08)	(.10)	(.10)	(.11)	(.10)	(.11)	(.11)
GSP_j	3.13***	1.69***	1.73***	1.74***	1.56***	1.76***	1.64***	1.68***
	(.06)	(.08)	(.10)	(.10)	(.10)	(.10)	(.10)	(.10)
R^2	.507	-	.034	.035	.020	.017	.018	.020
Time FEs	yes	yes	yes	yes	yes	yes	yes	yes
Country-pair FEs	no	yes	yes	yes	yes	yes	yes	yes
Country-time FEs	no	no	yes	yes	yes	yes	yes	yes
Demean order	no	no	tij	tji	jti	itj	ijt	jit
Num of Obs	1,184,525							

Note: Cluster (country-pairs) robust standard errors are reported in parentheses. Please refer to the footnote in Table 5 for the details of controls. ***, ** and * mean that the estimated coefficient is statistically significant at the 1%, 5%, and 10% level, respectively. For the within estimates, cluster-robust standard errors are reported with the degrees of freedom adjustment in section 2.2. The column (2) is implemented with the within estimator in two-level data index.

A Data descriptions

Data on nominal bilateral trade flows are drawn from the IMF's Direction of Trade Statistics, data on nominal GDP and GDP per capita are drawn from the Penn World Table (PWT) 7.0, and data on GDP deflator are drawn from the U.S. Department of Commerce's Bureau of Economic Analysis. PTA data are constructed from the Regional Trade Agreements Information System (RTA-IS) of the WTO and data on the GATT/WTO membership are also drawn from the WTO website. Our measure of PTAs include customs unions and free trade agreements. Data on distance, shared border, common language, common colony, and common legal origin are from CEPII.

Table 9: Description of variables used in gravity regressions

variables	description
T_{ijt}	is the c.i.f. import of i from j in year t
lnY_{it}	log real GDP for import-country i
$Distance_{ij}$	is the great-circle distance between i and j
$Area$	is the geographic area of a country
$Border_{ij}$	dummy equals one if i and j share a land border
$Landlock_{ij}$	is the number of landlocked countries for a pair (0, 1, or 2)
$Island_{ij}$	is the number of island countries for a pair (0, 1, or 2)
$ComLang_{ij}$	dummy equals one if i and j share a common language
$ComRelig_{ij}$	dummy equals one if i and j share a common religion
$Colony_{ij}$	dummy equals one if i has ever been a colony of j
$Colonizer_{ij}$	dummy equals one if i has ever been a colonizer of j
$CurColony_{ijt}$	dummy equals one if i was currently a colony of j in year t
$CurColonizer_{ijt}$	dummy equals one if i was currently a colonizer of j in year t
$ComColony_{ij}$	dummy equals one if i and j have ever been colonized by the same colonizer
$Remote_{ijt}$	is the distance of a country pair to the rest of the world weighted by all the other countries GDPs in year t
GSP_{ijt} (GSP_{jit})*	dummy equals one if i (j) offered GSP to j (i) in year t
$Hostility_{ij}$**	is the military conflict intensity between i and j
$Alliance_{ijt}$***	dummy equals one if i and j were in a formal alliance in year t
WTO_{ijt}	dummy equals one if both i and j were GATT/WTO members in year t
$OneWTO_{ijt}$	dummy equals one if either i or j was a GATT/WTO member in year t
CU_{ijt}	dummy equals one if i and j used the same currency in year t
RTA_{ijt}	dummy equals one if i and j belonged to the same regional trade agreement in year t

(Note) *: The Generalized System of Preferences (GSP) was adopted at UNCTAD II in New Delhi, India, in 1968. Its objective is to allow developing and least-developed countries to increase their export earnings, promote their industrialization, and accelerate their rates of economic growth through reduced or zero tariff rates over MFN rates in more developed countries. **: The Militarized Interstate Dispute Dataset (Ghosn and Palmer, 2003) provides information on conflicts in which one or more states threatened, displayed, or used force against one or more other states. They assign an ordinal number to each conflict, with higher numbers representing higher hostility. ***: The Formal Alliance data-set (Gibler and Sarkees, 2004) seeks to identify each formal alliance between at least two states that fall into the classes of defense pact, neutrality or non-aggression treaty, or entente agreement.

B A proof of Proposition 3

We consider the following equation for transformation:

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

where $i = 1, 2, ...n$ and $j = 1, 2, ...m$ are country identifiers and $t = 1, 2, ...T$ the time identifier.

In matrix format, we can rewrite:

$$\mathbf{y} = \mathbf{X}\alpha + \mathbf{u} + \mathbf{v} + \mathbf{D_w}\omega + \varepsilon \tag{B.1}$$

where $\mathbf{y} = (y_{111}, y_{112}, ..., y_{11T}, y_{121}, ..., y_{12T}, y_{131} ..., y_{1m1}, ..., y_{1mT}, y_{211}, ..., y_{21T}, y_{221}, .., y_{nm1}, .., y_{nmT})$ is a $m \times n \times T$ by 1 vector. In matrix format, we stack over t first, then over j and, finally, over i.

$$\mathbf{Q_w} = \mathbf{I_T} - \mathbf{D_w}(\mathbf{D'_w D_w})^{-1}\mathbf{D'_w}$$

$$\mathbf{Q_w} = \begin{pmatrix} \mathbf{I}_T & & & \\ & \mathbf{I}_T & & \\ & & \ddots & \\ & & & \mathbf{I}_T \end{pmatrix} - \begin{pmatrix} \mathbf{i}_T & & & \\ & \mathbf{i}_T & & \\ & & \ddots & \\ & & & \mathbf{i}_T \end{pmatrix} \cdot$$

$$\left[\begin{pmatrix} \mathbf{i}'_T & & & \\ & \mathbf{i}'_T & & \\ & & \ddots & \\ & & & \mathbf{i}'_T \end{pmatrix} \begin{pmatrix} \mathbf{i}_T & & & \\ & \mathbf{i}_T & & \\ & & \ddots & \\ & & & \mathbf{i}_T \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{i}'_T & & & \\ & \mathbf{i}'_T & & \\ & & \ddots & \\ & & & \mathbf{i}'_T \end{pmatrix}$$

$$\mathbf{Q_w} = \begin{pmatrix} \mathbf{I}_T & & & & \\ & \mathbf{I}_T & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \mathbf{I}_T \end{pmatrix} - \begin{pmatrix} \frac{1}{T}\mathbf{i}_T\mathbf{i}'_T & & & & \\ & \frac{1}{T}\mathbf{i}_T\mathbf{i}'_T & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \frac{1}{T}\mathbf{i}_T\mathbf{i}'_T \end{pmatrix}$$

$$\mathbf{Q_w} = \begin{pmatrix} \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T' & & & & \\ & \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T' & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T' \end{pmatrix}$$

where $\mathbf{i}_T$ is a $T \times 1$ column vector of 1, $\mathbf{Q_w}$ is a $m \times n \times T$ by $m \times n \times T$ matrix and $\mathbf{D_w}$ is a $m \times n \times T$ by $m \times n$ matrix.

Therefore, we obtain:

$$\mathbf{Q_w y} = \begin{pmatrix} \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T' & & & & \\ & \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T' & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \mathbf{I}_T - \frac{1}{T}\mathbf{i}_T\mathbf{i}_T' \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1m} \\ y_{21} \\ \vdots \\ y_{nm} \end{pmatrix}$$

where $\mathbf{y_{11}} = (y_{111}, y_{112}, ..., y_{11T})'$ and, similarly, $\mathbf{y_{nm}} = (y_{nm1}, y_{nm2}, ..., y_{mnT})'$.

$$\mathbf{Q_w y} = \begin{pmatrix} y_{111} - \bar{y}_{11} \equiv y^*_{111} \\ y_{112} - \bar{y}_{11} \equiv y^*_{112} \\ \vdots \\ y_{1mT} - \bar{y}_{1m} \equiv y^*_{1mT} \\ y_{211} - \bar{y}_{21} \equiv y^*_{211} \\ \vdots \\ y_{nmT} - \bar{y}_{mn} \equiv y^*_{nmT} \end{pmatrix}$$

where $\bar{y}_{11} = \frac{1}{T}\sum_{t=1}^{T} y_{11t}$, $\bar{y}_{1m} = \frac{1}{T}\sum_{t=1}^{T} y_{1mt}$, and $\bar{y}_{nm} = \frac{1}{T}\sum_{t=1}^{T} y_{nmt}$. Therefore, pre-multiplying by $\mathbf{Q_w}$ demeans variables using the mean over t.

We can rewrite (B.1), after pre-multiplying it by $\mathbf{Q_w}$, as follows:

$$y^*_{ijt} = \mathbf{X}^*_{ijt}\alpha + u^*_{it} + v^*_{jt} + \varepsilon^*_{ijt}$$

In matrix format, we can rewrite:

$$\mathbf{y}^* = \mathbf{X}^*\alpha + \mathbf{D_u u}^* + \mathbf{v}^* + \varepsilon^* \tag{B.2}$$

where $\mathbf{y}^* = (y^*_{111}, y^*_{121}, ..., y^*_{1m1}, y^*_{112}, ..., y^*_{1m2}, y^*_{113} ..., y^*_{1m3}, ..., y^*_{1mT}, y^*_{211}, ..., y^*_{2m1}, y^*_{212}, .., y^*_{n1T}, .., y^*_{nmT})$ is a $m \times n \times T$ by 1 vector. In matrix format, we stack over j first, then over t and, finally, over i.

$$\mathbf{Q_u} = \mathbf{I_m} - \mathbf{D_u}(\mathbf{D_u'}\mathbf{D_u})^{-1}\mathbf{D_u'}$$

$$\mathbf{Q_u} = \begin{pmatrix} \mathbf{I}_m & & & \\ & \mathbf{I}_m & & \\ & & \ddots & \\ & & & \mathbf{I}_m \end{pmatrix} - \begin{pmatrix} \mathbf{i}_m & & & \\ & \mathbf{i}_m & & \\ & & \ddots & \\ & & & \mathbf{i}_m \end{pmatrix} \cdot$$

$$\left[\begin{pmatrix} \mathbf{i}_m' & & & \\ & \mathbf{i}_m' & & \\ & & \ddots & \\ & & & \mathbf{i}_m' \end{pmatrix} \begin{pmatrix} \mathbf{i}_m & & & \\ & \mathbf{i}_m & & \\ & & \ddots & \\ & & & \mathbf{i}_m \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{i}_m' & & & \\ & \mathbf{i}_m' & & \\ & & \ddots & \\ & & & \mathbf{i}_m' \end{pmatrix}$$

$$\mathbf{Q_u} = \begin{pmatrix} \mathbf{I}_m & & & & \\ & \mathbf{I}_m & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \mathbf{I}_m \end{pmatrix} - \begin{pmatrix} \frac{1}{m}\mathbf{i}_m\mathbf{i}_m' & & & & \\ & \frac{1}{m}\mathbf{i}_m\mathbf{i}_m' & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \frac{1}{m}\mathbf{i}_m\mathbf{i}_m' \end{pmatrix}$$

where $\mathbf{i}_m$ is a $m \times 1$ column vector of 1, $\mathbf{Q_u}$ is a $m \times n \times T$ by $m \times n \times T$ matrix and $\mathbf{D_u}$ is a $m \times n \times T$ by $n \times T$ matrix.

Therefore, we obtain:

$$\mathbf{Q_u}\mathbf{y}^* = \begin{pmatrix} \mathbf{I}_m - \frac{1}{m}\mathbf{i}_m\mathbf{i}_m' & & & & \\ & \mathbf{I}_m - \frac{1}{m}\mathbf{i}_m\mathbf{i}_m' & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \mathbf{I}_m - \frac{1}{m}\mathbf{i}_m\mathbf{i}_m' \end{pmatrix} \begin{pmatrix} y^*_{11} \\ y^*_{12} \\ \vdots \\ y^*_{1T} \\ y^*_{21} \\ \vdots \\ y^*_{nT} \end{pmatrix}$$

where $\mathbf{y^*_{11}} = (y^*_{111}, y^*_{121}, ..., y^*_{1m1})'$ and, similarly, $\mathbf{y^*_{nT}} = (y^*_{n1T}, y^*_{n2T}, ..., y^*_{nmT})'$.

$$
\mathbf{Q_u}\mathbf{y}^* = \begin{pmatrix} y^*_{111} - \overline{y}^*_{11} \equiv y^{\#}_{111} \\ y^*_{121} - \overline{y}^*_{11} \equiv y^{\#}_{121} \\ \vdots \\ y^*_{1mT} - \overline{y}^*_{1T} \equiv y^{\#}_{1mT} \\ y^*_{211} - \overline{y}^*_{21} \equiv y^{\#}_{211} \\ \vdots \\ y^*_{nmT} - \overline{y}^*_{nT} \equiv y^{\#}_{nmT} \end{pmatrix}
$$

where $\overline{y}^*_{11} = \frac{1}{m}\sum_{j=1}^{m} y^*_{1j1}$, $\overline{y}^*_{1T} = \frac{1}{m}\sum_{j=1}^{m} y^*_{1jT}$, and $\overline{y}^*_{nT} = \frac{1}{m}\sum_{j=1}^{m} y_{njT}$. Therefore, pre-multiplying by $\mathbf{Q_u}$ demeans variables using the mean over j. We can rewrite (B.2), after pre-multiplying it by $\mathbf{Q_u}$, as follows:

$$
y^{\#}_{ijt} = \mathbf{X}^{\#}_{ijt}\alpha + v^{\#}_{jt} + \varepsilon^{\#}_{ijt}
$$

In matrix format, we can rewrite:

$$
\mathbf{y}^{\#} = \mathbf{X}^{\#}\alpha + \mathbf{D_v}\mathbf{v}^{\#} + \varepsilon^{\#} \tag{B.3}
$$

We stack over i first, then over t and, finally, over j.

$$
\mathbf{Q_v} = \mathbf{I_n} - \mathbf{D_v}(\mathbf{D_v'}\mathbf{D_v})^{-\mathbf{1}}\mathbf{D_v'}
$$

$$
\mathbf{Q_v} = \begin{pmatrix} \mathbf{I}_n & & & \\ & \mathbf{I}_n & & \\ & & \ddots & \\ & & & \mathbf{I}_n \end{pmatrix} - \begin{pmatrix} \frac{1}{n}\mathbf{i}_n\mathbf{i}_n' & & & \\ & \frac{1}{n}\mathbf{i}_n\mathbf{i}_n' & & \\ & & \ddots & \\ & & & \frac{1}{n}\mathbf{i}_n\mathbf{i}_n' \end{pmatrix}
$$

where $\mathbf{i}_n$ is a $n \times 1$ column vector of 1, $\mathbf{Q_v}$ is a $m \times n \times T$ by $m \times n \times T$ matrix and $\mathbf{D_v}$ is a $m \times n \times T$ by $m \times T$ matrix.

Therefore, we obtain:

$$\mathbf{Q_v y^{\#}} = \begin{pmatrix} \mathbf{I}_n - \frac{1}{n}\mathbf{i}_n\mathbf{i}_n' & & & & \\ & \mathbf{I}_n - \frac{1}{n}\mathbf{i}_n\mathbf{i}_n' & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \mathbf{I}_n - \frac{1}{n}\mathbf{i}_n\mathbf{i}_n' \end{pmatrix} \begin{pmatrix} y_{11}^{\#} \\ y_{12}^{\#} \\ \vdots \\ y_{1T}^{\#} \\ y_{21}^{\#} \\ \vdots \\ y_{mT}^{\#} \end{pmatrix}$$

where $\mathbf{y_{11}^{\#}} = (y_{111}^{\#}, y_{211}^{\#}, ..., y_{n11}^{\#})'$ and, similarly, $\mathbf{y_{mT}^{\#}} = (y_{1mT}^{\#}, y_{2mT}^{\#}, ..., y_{nmT}^{\#})'$.

$$\mathbf{Q_v y^{\#}} = \begin{pmatrix} y_{111}^{\#} - \overline{y}_{11}^{\#} \equiv \widetilde{y}_{111} \\ y_{211}^{\#} - \overline{y}_{11}^{\#} \equiv \widetilde{y}_{211} \\ \vdots \\ y_{n1T}^{\#} - \overline{y}_{1T}^{\#} \equiv \widetilde{y}_{n1T} \\ y_{121}^{\#} - \overline{y}_{21}^{\#} \equiv \widetilde{y}_{121} \\ \vdots \\ y_{nmT}^{\#} - \overline{y}_{mT}^{\#} \equiv \widetilde{y}_{nmT} \end{pmatrix}$$

Therefore, pre-multiplying by $\mathbf{Q_v}$ demeans variables using the mean over i. We can rewrite (B.3), after pre-multiplying it by $\mathbf{Q_v}$ as follows:

$$\widetilde{y}_{ijt} = \widetilde{\mathbf{X}}_{ijt}\alpha + \widetilde{\varepsilon}_{ijt} \tag{B.4}$$

In conclusion, we establish that sequential least square dummy variables approach on (B.1), (B.2) and (B.3) is equivalent to the within estimator with our proposed transformations of variables as in (B.4).

Acknowledgements

We would like to thank Xuepeng Liu for providing us his dataset. We are thankful to Pravin Trivedi, Prasada Rao and the seminar participants at the University of Queensland for their constructive comments.

References

ANDERSON, J. E., AND E. VAN WINCOOP (2003): "Gravity with gravitas: a solution to the border puzzle," *American Economic Review*, 83(1), 170–192.

BAIER, S. L., AND J. H. BERGSTRAND (2007): "Do free trade agreements actually increase members' international trade?," *Journal of International Economics*, 71(1), 72–95.

BALDWIN, R., AND D. TAGLIONI (2006): "Gravity for Dummies and Dummies for Gravity Equations," *NBER Working Paper Series*, (12516).

CAMERON, A. C., AND P. K. TRIVEDI (2005): *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

CHENG, I.-H., AND H. WALL (2005): "Controlling for Heterogeneity in Gravity Models of Trade and Integration," *Federal Reserve Bank of St Louis Review*, 87, 49–63.

CHEONG, J., D. W. KWAK, AND K. K. TANG (2012): "Can trade agreement curtail trade creation and prevent trade diversion? Evidence on cross trade agreement effects," *mimeo, The 6th Empirical Investigations in Trade and Investment (EITI) 2012*.

DAVIDSON, R., AND J. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford.

EGGER, P., AND M. PFAFFERMAYR (2004): "Distance, Trade and FDI: A Hausman-Taylor SUR Approach," *Journal of Applied Econometrics*, 19, 227–246.

EICHER, T., C. HENN, AND C. PAPAGEORGIOU (2012): "Trade Creation and Diversion Revisited: Accounting for Model Uncertainty and Unobserved bilateral heterogeneity," *Journal of Applied Econometrics*, 27(2), 296–321.

EICHER, T. S., AND C. HENN (2011): "In search of WTO trade effects: Preferential trade agreements promote trade strongly, but unevenly," *Journal of International Economics*, 83(2), 137–153.

FRISCH, R., AND F. WAUGH (1933): "Partial time regressions as compared with individual trends," *Econometrica*, 1, 387–401.

HSIAO, C. (2003): *Analysis of Panel Data*. Cambridge University Press, 2nd edn.

KONYA, L., L. MATYAS, AND M. HARRIS (2011): "GATT/WTO membership does promote international trade after all: Some new impirical evidence," *MPRA Working Paper No. 34978*.

LANCASTER, T. (2000): "The incidental parameter problem since 1948," *Journal of Econometrics*, 95, 391–413.

LIU, X. (2009): "GATT/WTO promotes trade strongly: sample selection and model specification," *Review of International Economics*, 17(3), 428–446.

LOVELL, M. (1963): "Seasonal adjustment of economic time series," *Journal of the American Statistical Association*, 58, 993–1010.

MAGEE, C. S. (2003): "Endogenous preferential trade agreements: an empirical analysis," *Contributions to Economic Analysis and Policy*, 2(1), Article 15.

——— (2008): "New measures of trade creation and trade diversion," *Journal of International Economics*, 75(2), 349–362.

MATYAS, L. (1997): "Proper Econometric Specification of the Gravity Model," *The World Economy*, 20(3), 363–368.

MATYAS, L., AND L. BALAZSI (2012): "The Estimation of Multi-dimensional Fixed Effects Panel Data Models," *mimeo*.

NEYMAN, J., AND E. L. SCOTT (1948): "Consistent estimation from partially consistent observations.," *Econometrica*, 16, 1–32.

ROSE, A. K. (2004): "Do We Really Know That the WTO Increases Trade?," *American Economic Review*, 94(1), 98–114.

ROY, J. (2011): "Is the WTO mystery really solved?," *Economics Letters*, 113, 127–130.

SERLENGA, L., AND Y. SHIN (2007): "Gravity Models of Intra-EU Trade: Application of the CCEP-HT Estimation in Heterogeneous Panels with Unobserved Common Time-Specific Factors," *Journal of Applied Econometrics*, 22, 361–381.

SUBRAMANIAN, A., AND S.-J. WEI (2007a): "The WTO Promotes Trade, Strongly But Unevenly," *Journal of International Economics*, 72(1), 151–175.

——— (2007b): "The WTO Promotes Trade, Strongly But Unevenly," *Journal of International Economics*, 72(1), 151–175.

TOMZ, M., J. GOLDSTEIN, AND D. RIVERS (2007): "Do we really know that the WTO increases Trade? Comment," *American Economics Review*, 97 (5), 2005–2018.

WOOLDRIDGE, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, vol. 1 of *MIT Press Books*. The MIT Press, 2nd edn.