The Role of STC in International Trade Patterns: A Dynamic Panel Data Analysis with Attrition*

Singa Wang† University of Auckland

Xuewan Xu‡ University of Auckland

Ping Yu§ University of Hong Kong

Started: November 2013

First Draft: December 2014

This Version: June 2015

Abstract

This paper studies the effects of special trade concerns (STC) on trade occurrence and trade volume. We emphasize the endogeneity of STC and dynamics of the trading process, which were much ignored in the literature, by applying the correlated random effects methods to a dynamic panel with attrition. Main results are as follows. First, STC has a negative effect (-5.73%) on the probability of trade occurrence. Second, STC has a positive effect (1.2%) on the trade volume conditioning on trade happening even if a STC is proposed. Third, the effects of STC are time-specific - there are three regimes in its effect process: the premature period (1997-2001), the mature period (2002-2008) and the post-crisis period (2009-2012). We attribute the first result to the cost effect of STC and the second result to its signaling effect.

KEYWORDS: special trade concern, dynamic binary choice panel, dynamic panel with sample selection, correlated random effects, generalized estimating equations, unit root, signaling

JEL-CLASSIFICATION: C23, C24, C25

*We thank the participants at SETA2015 for helpful comments.

†Department of Economics, 12 Grafton Road, University of Auckland, New Zealand; Email: xingang.wang@auckland.ac.nz.

‡Ministry of Agriculture of the People's Republic of China; Email: xuewan.xu@auckland.ac.nz.

§School of Economics and Finance, The University of Hong Kong, Pokfulam Road, Hong Kong; Corresponding Author Email: pingyu@hku.hk.

1 Introduction

Since the establishment of the WTO agreement on the application of sanitary and phytosanitary measures (SPS Agreement hereinafter) in 1995, SPS measures have been attracting increasing attention from global general public, especially from policy makers of national governments. SPS measures have also been broadly believed to be one of the most effective instruments to promote plant, animal and human health and fair trade under the WTO framework; see Appendix A for what SPS agreement tries to do and why it is necessary. However, policy making regarding SPS measures typically involves two conflicting objectives - the introduction of SPS measures or related food safety regulations may protect plant, animal health and human safety but meanwhile restricts the trade of agricultural and food products. This throws the policy makers to a dilemma where they have to balance between high level of trade liberalization and high level of health protection. This paper does not intend to propose any policy suggestion such as whether any SPS measure is suitable, rather, it tries to provide a detailed empirical analysis on the impacts of SPS measures on the pattern of international trade.

Intuitively, SPS measures have two effects on the trade pattern. On one hand, they impose higher costs on the exporters, so reduce the possibility of trade occurrence. On the other hand, if the exporters pass the SPS measures, domestic consumers will be more confident about the safety and quality of the imported products, which may boost the demand for international trade. We examine these two effects in details below.

It is widely acknowledged that implementation of quite a lot of SPS measures or related regulations has strong negative influences on the trade of agricultural and food products; see, inter alia, Petrey and Johnson (1993), Ndayisenga and Kinsey (1994), National Research Council (1995), Sykes (1995), Digges et al. (1997), Hillman (1997), Thilmany and Barrett (1997), Jaffee (1999) and Unnevehr (1999). The negative impacts take place in two ways. First, the importers can completely prohibit trade by imposing critical bans in the name of environment or health protection; this mainly involves prohibiting the use of special production processes, such as beef products treated by hormone, food from genetically modified organism and agricultural products from non-free-disease zone.[1] Second, some restrictive and prohibitive SPS measures or regulations can also impede or ban trade in agricultural and food products by raising production or marketing costs of exporting producers. Such measures include a series of conformity assessment requirements (certification, sampling, testing, inspection, verification, registration, accreditation and approval procedures) and requirements for the characteristics of the product (plant and animal cold or quarantine requirements, non-zero-tolerance limits for residues of chemicals, contaminants and microbes, labeling and packaging, and geographical application of measures). From the perspective of production costs, some restrictive SPS measures like increased test and certification requirements entail internal self-testing before exporting, which increases both the fixed cost (e.g., purchase of new testing equipments) and the variable cost (e.g., more testing staff and time). This large cost increase has a growing tendency to distort the original appropriate allocation of resources, reduce production efficiency, undermine the international competitiveness of export products and eventually cause the tariff effect, which substantially and effectively restricts imports of foreign products. We label these negative effects of SPS measures as the cost effect.

On the other hand, proper and reasonable SPS measures are necessary to ensure food safety. With rapid growth of global economics and increasing consuming capability over the past decades, international trade in the food and agricultural sector has expanded significantly, which demands for higher level of food safety particularly in developed countries. Generally speaking, the more stringent the SPS measures, the higher the

[1] In fact, all beef products treated by hormone have been prohibited to export to European markets since the beef hormone problem arose, which brought substantial losses to the main beef product exporters such as the US and Canada.

safety level of the imported products. Consequently, some restrictive or prohibitive SPS requirements may give consumers a signal that the commodities complying with the designated SPS measures are safer, and thereby enhance their confidence in imported food and boost the trade. In addition, in order to meet strict SPS standards, the exporting producers have to improve their production efficiency and effectiveness, and thus raise the competitiveness of exporting commodities. In this case, stringent SPS measures necessarily pose a positive effect on trade, due to lower relative production costs of importing products and the consumers' high willingness to pay for safer and quality products (see WTO (2012)). We label this positive effect of SPS measures as the signaling effect.

Given these two opposite-directioned effects of SPS measures on trade, an interesting empirical question is what is the net effect of SPS agreement. Of course, as argued above, suitable SPS measures are necessary for food safety, so it is better to concentrate on the effect of "excessive" SPS requirements on trade. However, excessiveness is quite subjective. Fortunately, there is an objective measure of "excessiveness", so-called special trade concerns (STC hereinafter). Examples of specific trade concerns include a range of topics related to BSE, better known as "mad cow disease". Many countries imposed BSE-related trade barriers, and some exporters feel they are unfairly affected. Another example is Mexican restrictions on Thai milled rice, allegedly intended to avoid the introduction of pests.[2] The goal of this paper is to study the effects of STC on the international trade patterns. We use a panel data set with attrition to tackle this problem. Attrition is obvious since there may not be trade in some trade routes during some years in the period of interest. Most existing literature on this problem uses the Poisson pseudo maximum likelihood (PPML) gravity model (see, e.g., Schlueter et al. (2009)) or the static sample selection model (see, e.g., Crivelli and Gröschl (2012)) to estimate the effects of STC; see Crivelli and Gröschl (2012) for a summary of relevant literature. However, two key features of the data are missing in such models: the dynamics of the trading process and the endogeneity of STC. The former means that this year's trade is related to last year's trade in both occurrence and volume, and the latter means that STC is related to some historical or practical factors which are left in the error term. We analyze the dynamic participation process (how trade occurrence is determined) and the dynamic trading process (how trade volume is determined) using correlated random effects (CRE) techniques which explicitly consider observed omitted factors in the error term. In addition, we apply the generalized estimating equations (GEE) technique of Liang and Zeger (1986) to the dynamic participation process to make the estimation robust to serial correlation among the error terms; we also estimate the dynamic trading process by applying Heckman's correction on the level rather than the difference of the trade volume as in the first-difference instrumental variables (FD-IV) estimator of Anderson and Hsiao (1981, 1982) to avoid the weak IV problem. Our main results are (i) STC negatively affects the frequency of trade occurrence; (ii) STC positively affects the trade volume conditional on that trade happens even if a STC is proposed; (iii) the effects of STC are time-specific - there are three regimes in the effect process from 1997 to 2012. Although result (i) and/or (ii) may appear in some previous papers, the drawbacks in their econometric methods make their conclusions unreliable.

The contribution of this paper is better understood in the following way. We do not intend to propose any new econometric technique in this paper; rather, we show how to modify and adapt the existing methods to a new environment. Since many panel data sets with attrition are expected to have similar features as ours, we hope our solution is generic and can be applied to other similar problems. Also, our empirical results seem interesting.

The rest of this paper is organized as follows. Section 2 summarizes the main features of our data set, reviews the existing dynamic panel methods for binary choice and sample selection, and argues why our methods are best suitable to the current problem. Section 3 analyzes the dynamic participation process:

[2] In this case, continued Thai pressure at SPS Committee meetings led to a change in the Mexican measure.

we first provide the specification and estimation technique of our model and then report the empirical results, especially the average partial effects (APEs). Section 4 analyzes the dynamic trading process by first describing our specification and identification scheme and then reporting the empirical results. Finally, Section 5 concludes and discusses directions for future research. A word on notation: different from most econometric literature, vectors in this paper are defined as row vectors to simplify notations; for two vectors x and y, xy means their inner product; $\phi(\cdot)$ and $\Phi(\cdot)$ are standard normal probability density function (pdf) and cumulative distribution function (cdf), respectively.

	Full Sample		$STC = 1$		$STC = 0$	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
STC	.00804	.0893	1	0	0	0
FTA	.229	.420	.331	.471	.228	.419
POP	32.405	2.705	34.095	1.961	32.392	2.706
GDP	17.254	2.189	18.937	1.472	17.240	2.188
$\overline{\text{STC}}$	.00804	.0633	.506	.159	.004	.0425
$\overline{\text{FTA}}$	.229	.375	.270	.290	.228	.376
$\overline{\text{POP}}$	32.405	2.700	34.044	1.961	32.392	2.701
$\overline{\text{GDP}}$	17.254	2.117	18.729	1.466	17.242	2.118
DIST	8.501	.921	8.473	.934	8.501	.921
LANG	.593	.491	.706	.456	.592	.492
REL	.179	.383	.0793	.270	.179	.384
COL	.0317	.175	.0503	.219	.0316	.175
CONT	.0376	.190	.0276	.164	.0377	.190
d_{i0}	.296	.457	.680	.467	.293	.455
No. of Obs	180512		1451		179061	

Table 1: Descriptive Statistics, 1997-2012

Note: $\bar{\cdot}$ is the sample mean of $\cdot$ for a trade route during 1997-2012;

d_{i0} is the indicator for the initial trade occurrence in 1996

2 Summary of the Data and Econometric Methods

The SPS agreement was formally established after the Uruguay Round in 1995; see Appendix A for a historical description of SPS agreement. So theoretically, STC can be raised by exporters from 1996 and implemented from 1997. We will first report descriptive statistics of control variables in Table 1 to illustrate the endogeneity of STC if time-invariant variables are omitted. Given that the main goal of this paper is to analyze the effect of STC on the international trade patterns, we will then summarize the basic information on STC and trade during 1997-2012 in Table 2 to show why modeling dynamics of the trading process is necessary and what should be paid attention to in such modeling. Appendix B describes the construction method of our data set and provides definitions of variables used in the rest of the paper.

From Table 1, we can draw a few interesting conclusions. First, STC is a rare event with the occurring probability 0.8% (although the absolute number of its occurrence is not small - 1451). This implies that correctly and robustly modeling its relationship with other variables (either dependent or independent variables) is critical to measure its effects. Second, STC, Free Trade Agreement (FTA), population (POP) and GDP are the only four time-varying covariates, and all others are time invariant. Obviously, there are important correlations between STC and these time-invariant variables. For example, STC is more likely to be proposed between FTA members and between partners that are more wealthy, with common language, and have colonial heritage, and less likely to be proposed between partners that have common religion and are contiguous. Population and distance between the partners seem not to correlate with STC much. These results are intuitively interpretable. As a result, neglecting these time-invariant factors will seriously bias the effects of STC. Finally and most importantly, STC seems inherent and historically dependent rather than randomly proposed. For example, among the trade routes and time periods with STC proposed, the average STC numbers during 1997-2012 is about 0.5, which is, however, roughly null for the trade routes and time periods without STC proposed. Also, STC is more likely to be proposed when there was trade in 1996 - the initial period, which indicates that historical information, as embodied in the initial trade, is very informative about the subsequent STC behaviors. It is thus very important to incorporate all these factors in our model; this is why we use the CRE methods to study the effects of STC, where the effects of time-invariant factors are explicitly controlled.

Table 2 summarizes the temporal information on STC and trade. First, the European Union (EU) is the biggest complainer and complainee; other popular complainers include Egypt, Côte d'Ivoire, Brail and USA and complainees include Australia, USA, Slovakia, Israel, Czech Rep., Brazil, Indonesia and Japan. Second, the number of STCs increases during 1997-2008 but becomes unstable after the financial crisis; a similar pattern appears in the total trade occurrence and volume. Third, it seems that the STC system becomes mature around 2001, e.g., some popular complainers and complainees started to operate. Overall the data manifest a three-regime pattern: the premature period (1997-2001), the mature period (2002-2008) and the post-crisis period (2009-2012).

For the effects of STC on the trade pattern, we study two aspects: the effect on trade occurrence d_{it} and the effect on trade volume y_{it}, where $d_{it} \in \{0, 1\}$ with 0 indicating no trade happened and 1 indicating the converse, y_{it} is the logarithm of trade volume as $d_{it} = 1$, i indexes the trade route and t indexes the time period. In the following, we first state some basic facts about d_{it} and y_{it} and then review related methods of modeling them in the literature.

First, we must model d_{it} in a dynamic way. To appreciate why, check the Markov transition matrix of d_{it} in Table 3. This transition matrix indicates that there is strong persistence in d_{it}, i.e., if there is trade in the last period, it tends to have trade in the current period, and vice versa. This is understandable since it is much harder to start or terminate a trade relationship than maintain the present state due to political or historical reasons. Note that in Table 3, we did not normalize the sum of the numbers in each row as 1; rather, we normalize the sum of all transition probabilities as 1 to provide more information about the relative proportion of each transition. Second, if we regress y_{it} on $y_{i,t-1}$ when both of them are available, we have

$$\underset{}{\widehat{y}_{it} = \underset{(0.0004)}{0.9985} \cdot y_{i,t-1},}$$

which obviously indicates that the model has a unit root. As noted by Blundell and Bond (1998), the FD-IV estimator of Anderson and Hsiao will suffer from the weak IV problem in this case;[3] see also Binder et al. (2005) and Hsiao et al. (2002) for likelihood methods (which do not suffer from the weak IV problem) in the case without sample selection. Another problem associated with the FD-IV estimator is that it requires observability of at least three consecutive periods. The following Table 4 summarizes the observability of y_{it}

[3] Intuitively, if $y_{it} = y_{i,t-1} + \varepsilon_{it}$ with $Cov(\varepsilon_{it}, \varepsilon_{is}) = 0$ for $s \neq t$, then $y_{it} - y_{i,t-1} = \varepsilon_{it}$ is uncorrelated with the usual IVs - past values $(y_{i,t-2}, \cdots, y_{i0})$.

Exporter (179)	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	Total
EU	0	14	28	28	28	14	27	42	60	43	61	61	61	61	61	22	611
Egypt	0	0	0	0	0	15	15	15	24	24	24	26	26	26	26	26	247
Côte d'Ivoire	0	0	0	0	0	15	15	15	21	21	21	22	22	22	22	22	218
Brazil	0	0	0	0	0	0	0	15	23	23	23	25	25	26	26	26	212
USA	0	1	1	1	1	1	2	3	4	5	5	4	4	4	4	4	44
Importer (189)																	
EU	0	0	0	0	0	30	30	60	91	68	68	68	68	68	68	68	687
Australia	0	0	0	0	0	0	0	14	21	21	21	23	23	23	23	0	169
USA	0	0	0	0	0	0	0	0	0	22	22	24	24	24	24	24	164
Slovakia*	0	0	15	15	15	15	15	15	26	3	3	3	3	3	3	3	137
Israel	0	0	0	0	0	0	0	0	0	0	18	20	20	20	20	0	98
Czech Rep.	0	14	14	14	14	0	0	0	3	3	3	3	3	3	3	3	80
Brazil	0	0	0	0	0	0	14	14	17	0	0	0	0	0	0	0	45
Indonesia	0	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	41
Japan	0	0	0	0	0	0	2	2	2	3	3	3	3	5	5	5	33
Total-STC	0	16	31	31	31	49	67	97	139	124	143	152	152	154	154	111	1451
Total-Trade Occurence ($\times 10^3$)	3.51	3.58	4.03	4.67	4.78	4.94	5.08	5.26	5.43	5.56	5.68	5.74	5.64	5.74	5.75	5.38	80.75
Total-Trade Volume ($\times 10^{10}$)	2.31	2.41	2.57	2.40	2.50	2.63	2.96	3.21	3.38	3.67	4.21	4.36	4.20	4.77	5.05	4.57	55.22

Table 2: The Pattern of STC and Trade During 1997-2012

Note*: Slovakia is in EU since 2004

in our data set, where n is the total number of trade routes and $T = 16$ is the total number of periods.[4] Obviously, from 2 consecutive periods (which is required for dynamic modeling of y_{it}) to 3 consecutive periods, we lose about 16% ($(0.299 - 0.355)/0.355$) data points. In summary, it is better to model y_{it} in a dynamic way and estimate the model based on levels rather than differences.

$d_{i,t-1} \backslash d_{it}$	0	1
0	0.472	0.092
1	0.081	0.355

Table 3: Transition Matrix of d_{it}

proportion$\backslash s$	1	2	3	y_{i0}	$1 + y_{i0}$	$2 + y_{i0}$
$\cdot/nT$	0.447	0.355	0.299	0.297	0.237	0.222
$\cdot/n$	0.977	0.672	0.553	0.297	0.288	0.280

Table 4: Observability for s Consecutive Periods and/or y_{i0}: $nT = 180512, n = 11282$

There is tremendous literature on the dynamic participation model originated by Heckman (1981a,b,c). Hyslop (1999) uses the simulation method to estimate a CRE model where the idiosyncratic error follows an $AR(1)$ process. Since $T = 16$ in our setup, this method is too time-consuming. Honoré and Kyriazidou (2000) identify a fixed effects (FE) model by conditioning on a set of covariate values with positive probability, which implicitly excludes time dummies.[5] However, from Table 2, time dummies are necessary to incorporate the trend in STC numbers. Actually, Honoré and Tamer (2006) show that the model is generally only partially identified. Honoré and Lewbel (2002) identify the FE model using a special covariate which is not available in our data. Arellano and Carrasco (2003) propose a class of semiparametric random effects estimators when the covariates are only predetermined, but their method suffers from the curse of dimensionality when T or the dimension of covariates is large. Woutersen (2002) and Carro (2007) propose bias correction methods in the FE model when T is large by adjusting the first order conditions (FOCs) of the likelihood; Fernández-Val (2009) and Hahn and Kuersteiner (2011) make adjustments in the estimator and Arellano and Hahn (2006) in the likelihood; Dhaene et al. (2006) extend the Jackknife method used by Hahn and Newey (2004) in the static model to the dynamic model. We do not use the large-T framework since in our data $T/n \approx 0.0014$ which is quite small and also the large-T correction excludes time dummies. Bartolucci and Nigro (2010) consider the quadratic exponential model which allows to use the conditional likelihood to identify the model; although time dummies are allowed, the effect of time-invariant covariates cannot be identified. Our estimation method is based on Wooldridge (2005a); see also Wooldridge (2000, 2005b). We modify Wooldridge's method to be robust to the autocorrelation among idiosyncratic errors; details are given in Section 3.

There is also some literature on the dynamic panel data with sample selection.[6] Bover and Arellano (1997) propose a two-step within-group method based on estimated reduced form predictions of the latent endogenous variables, but the restrictions on the reduced form seem too strong. Arellano et al. (1999) estimate a model without other covariates except lagged dependent variables. Their method is based on differencing and the minimum distance estimation of Chamberlain (1982). Also, their method relies on the existence of a set of contiguous data for which the latent variables are observed. Both Bover and Arellano (1997) and Arellano et al. (1999) take a parametric "random effects approach". Kyriazidou (2001)

[4] T is the total number of periods. Since we use $AR(1)$ model for d_{it} and y_{it}, there are only 15 equations for each i.

[5] Also, the performance of their estimator deteriorates with the number of exogenous variables.

[6] Most review papers on sample selection such as Verbeek and Nijman (1996), Vella (1998) and Honoré et al. (2008) concentrate on cross-sectional data or static panel data.

develops semiparametric FE estimators by extending the identification idea of the static model in Kyriazidou (1997). Kyriazidou (2001) does not allow for predetermined regressors except $y_{i,t-1}$; Gayle and Viauroux (2007) allow for this extension but impose restrictions on the individual effects of the participation equation. Both methods are based on differencing and involve some nonparametric techniques, which makes them unattractive to our data. Fernández-Val and Vella (2011) introduce large-T bias corrected estimators when endogeneity results from both time invariant and time varying heterogeneity, but they do not allow for time dummies in both the selection and primary equations. In our estimation, we extend Semykina and Wooldridge (2013)'s method to adapt to our specific context; details are provided in Section 4.

3 Dynamic Participation Process

In this section, we first specify the empirical model we will use, and then report the empirical results and APEs.

3.1 Empirical Specification and Identification

We specify the dynamic participation process as

$$d_{it} = 1(d_{it}^* \geq 0),$$

where

$$d_{it}^* = \phi d_{i,t-1} + z_{it}\gamma + \delta_t D_t + \eta_i + u_{it} \tag{1}$$

is the latent dependent variable, z_{it} is a vector of time-varying explanatory variables including STC, indicator for Free Trade Agreement (FTA) between the two trade partners, GDP and population, η_i is the time-invariant characteristic of trade route i, D_t is the time dummy for period t, and $u_{it} \sim N(0, \sigma_u^2)$.[7] In other words, we assume that whether there is trade during this period depends on whether there was trade during the last period. This is usually due to inertia in trade bargaining. We are especially interested in ϕ and the coefficient of STC. In standard (uncorrelated) random effect (RE) models, η_i is assumed to be uncorrelated with z_{it}. Instead, following Mundlak (1978) and Chamberlain (1980, 1982, 1984), correlation between η_i and the observed characteristics in the model can be allowed by assuming a relationship between η and either the time means of the z variables or a combination of their lags and leads. For example, $\eta_i = \overline{z}_i\psi + \zeta_i$, where $\overline{z}_i$ is the sample average of z_{it} for trade route i, and ζ_i is i.i.d. and independent of z_{it} and u_{it} for all i, t.[8] In this paper, we assume that η_i is related to not only z_{it}, $t = 1, \cdots, T$, but also other time-invariant variables which are often called gravity controls. These time-invariant variables include historical factors (e.g., common language, common religion, previous colony) and factors related to the fixed costs of trade (e.g., adjacency, geographical distance between capitals). We collect all these variables in z_i and assume

$$\eta_i = z_i\psi + \zeta_i,$$

where z_i includes a constant, $\overline{z}_i$ and other time-invariant variables. We denote $\mathbf{z}_{it} = (z_{it}, z_i)$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}, \cdots, \mathbf{z}_{iT})$.

For a dynamic binary choice CRE model, we need to consider two further problems: the time dependence

[7] Usually, Logit or Probit is only of personal preference. Nevertheless, Probit has some computational and theoretical advantages over Logit; see, e.g., Section 2 of Papke and Wooldridge (2008) for some related discussions.

[8] We assume η_i is only related to $\overline{z}_i$ rather than $(z_{i1}, \cdots, z_{iT})$ because $T = 16$ in our case such that the dimension of $(z_{i1}, \cdots, z_{iT})$ is quite large.

among u_{it} and the initial conditions (IC) problem. First, assume u_{it} is i.i.d. and consider how to deal with the IC problem. For identification, normalize $u_{it} \sim N(0,1)$. Then the likelihood of $(d_{i1}, \cdots, d_{iT})$ given d_{i0} and ζ_i is

$$f(d_{i1}, \cdots, d_{iT} | d_{i0}, \zeta_i, \mathbf{Z}_i; \phi, \gamma, \delta) = \prod_{t=1}^{T} \Phi\left(\left(\phi d_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_t D_t + \zeta_i\right)(2d_{it} - 1)\right),$$

where $\delta = (\delta_1, \cdots, \delta_T)$ and $\boldsymbol{\gamma} = (\gamma, \psi)$. Since ζ_i is unobservable, we must integrate it out to get a feasible likelihood function. In a dynamic model, d_{i0} and ζ_i are usually not independent (intuitively, η_i, and hence ζ_i, affects d_{it} for any t from (1)), so we must model the joint distribution $f(d_{i0}, \zeta_i)$. There are two methods in the literature. The first method is proposed by Heckman (1981b) who models $f(d_{i0}|\zeta_i)$. Specifically, assume

$$d_{i0}^* = \mathsf{z}_{i0}\pi + \tau\zeta_i + u_{i0},$$

where z_{i0} includes the presample information (and at least includes $\mathbf{z}_{i0}$), $u_{i0} \sim N(0,1)$ is independent of ζ_i and all other u_{it}'s, and τ captures any change in error variance. Now, the likelihood of $(d_{i0}, d_{i1}, \cdots, d_{iT})$ given ζ_i is

$$f(d_{i0}, d_{i1}, \cdots, d_{iT} | \zeta_i, \mathsf{Z}_i; \phi, \gamma, \delta, \pi, \tau) = \Phi\left(\left(\mathsf{z}_{i0}\pi + \tau\zeta_i\right)(2d_{i0} - 1)\right) \prod_{t=1}^{T} \Phi\left(\left(\phi d_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_t D_t + \zeta_i\right)(2d_{it} - 1)\right),$$

where $\mathsf{Z}_i = (\mathbf{Z}_i, \mathsf{z}_{i0})$ and the marginal likelihood of $(d_{i0}, d_{i1}, \cdots, d_{iT})$ is

$$\int \Phi\left(\left(\mathsf{z}_{i0}\pi + \tau\sigma_\zeta\zeta^*\right)(2d_{i0} - 1)\right) \prod_{t=1}^{T} \Phi\left(\left(\phi d_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_t D_t + \sigma_\zeta\zeta^*\right)(2d_{it} - 1)\right) dF_{\zeta^*}(\zeta^*),$$

where F_{ζ^*} is the cdf of $\zeta^* = \zeta/\sigma_\zeta$ with σ_ζ being the standard deviation of ζ. When $\zeta_i \sim N(0, \sigma_\zeta^2)$, the integration can be evaluated using Gaussian-Hermite quadraure (Butler and Moffitt (1982)). The second method is proposed by Wooldridge (2005a) who models $f(\zeta_i|d_{i0})$. Specifically, he assumes

$$\zeta_i = \pi d_{i0} + a_i,$$

where a_i is independent of d_{i0} and $\mathbf{Z}_i$ but may be correlated with $d_{i,t-1}$. As a result,

$$d_{it}^* = \phi d_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_t D_t + \pi d_{i0} + a_i + u_{it}. \tag{2}$$

Now, the likelihood of $(d_{i1}, \cdots, d_{iT})$ given d_{i0} and a_i is

$$f(d_{i1}, \cdots, d_{iT} | d_{i0}, a_i, \mathbf{Z}_i; \phi, \gamma, \delta, \pi) = \prod_{t=1}^{T} \Phi\left(\left(\phi d_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_t D_t + \pi d_{i0} + a_i\right)(2d_{it} - 1)\right)$$

and the likelihood of $(d_{i1}, \cdots, d_{iT})$ given d_{i0} is

$$\int \prod_{t=1}^{T} \Phi\left(\left(\phi d_{i,t-1} + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_t D_t + \pi d_{i0} + \sigma_a a^*\right)(2d_{it} - 1)\right) dF_{a^*}(a^*)$$

where F_{a^*} and a^* are similarly defined as F_{ζ^*} and ζ^*. When $a_i \sim N(0, \sigma_a^2)$, the quadrature method of Butler and Moffitt (1982) can be used to calculate the integration. In both methods, the resulting likelihood has exactly the same structure as in the standard RE Probit model, except that the explanatory vairables at time period t are adjusted suitably.

We next consider the serial correlation among u_{it}. For each trade route, the idiosyncratic error u_{it} may be serially correlated due to factors not covered in the model. When u_{it} is serially correlated, the estimators based on the above two methods are not consistent. A popular way to model the serial correlation is to assume u_{it} follows an $AR(1)$ process, i.e., $u_{it} = \varrho u_{i,t-1} + \epsilon_{it}$, where $\varrho \in (0,1)$ and $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$; see, e.g., Hyslop (1999). However, construction of the likelihood involves high-dimensional integral and is numerically difficult. Maximum simulated likelihood (MSL) is a natural alternative estimator to use. A popular simulator is the so-called GHK algorithm of Geweke, Hajivassiliou and Keane. The MSL estimation routine provides a consistent estimator of the vector of parameters as the number of simulation draws tends to infinity (and is asymptotically equivalent to the ML estimator). Strictly speaking, for the simulation error to disappear asymptotically, the number of simulation draws needs to increase at a rate greater than the square root of the sample size. Another restriction associated with the $AR(1)$ modeling of u_{it} is that d_{it}^* will satisfy the "common factor" restricted form. For example, suppose d_{it}^* follows (1); then

$$d_{it}^* = \varrho d_{i,t-1}^* + \phi d_{i,t-1} + \phi_{-1} d_{i,t-2} + z_{it}\gamma + z_{i,t-1}\gamma_{-1} + \delta_t D_t + \delta_{-1} D_{t-1} + f_i + \epsilon_{it},$$

where $\phi_{-1} = -\varrho\phi, \gamma_{-1} = -\varrho\gamma$, $\delta_{-1} = -\varrho\delta_{t-1}$ and $f_i = (1-\alpha)\eta_i$. In contrast, we do not impose the implied common factor restrictions. As an alternative, we treat the dynamics as an empirical approximation to some more general adjustment process and use the GEE technique to get robust to the unknown correlation structure of the error terms. In Appendix C, we adapt the GEE method to the current context of this paper.

In summary, we apply the GEE mechanism to (2) and the empirical results are reported in the next subsection.

3.2 Empirical Results and Average Partial Effects

	STC Only	STC with D_t	Standard	Correlated	With d_0 and z_i	GEE	$\cdot \big/ \sqrt{1+\widehat{\sigma}_a^2}$
d_{t-1}	1.169***	1.048***	1.043***	1.038***	.991***	.935***	.862
STC	.261***	.0415	−.0867	−.276***	−.289***	−.225***	−.251
FTA			.277***	.0421*	.0434*	.0362*	.0378
POP			.166***	−.0249	−.0221	.00839	−.0192
GDP			.209***	.178***	.185***	.151***	.161
d_0					.908***	.733***	.790
$\frac{\sigma_a^2}{\sigma_a^2+1}$	.419***	.471***	.365***	.363***	.243***		

Table 5: Parameter Estimates Under Different Specifications and Methods

Note: *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

We first summarize the GEE estimates in the following Table 5 and then calculate a few APEs to illustrate the effects of STC on trade occurrence. Since we compare many other methods and specifications with ours, we do not report standard errors and all coefficients in Table 5 for brevity, but use $*$ to indicate significance of selected coefficients (mainly the coefficients of z_{it}) of different estimates at different significance levels. Nevertheless, all implementation details are available upon request. In Table 5, "STC only", "STC with D_t" and "Standard" are all based on (1) with η_i and u_{it} independent of z_{it}, but the specifications of z_{it} are different. "Correlated" is the same as "Standard", but assumes $\eta_i = \overline{z}_i\psi + \zeta_i$ with ζ_i independent of z_{it}. "With d_0 and z_i" is the same as "Correlated" but adds in d_{i0} and z_i as additional regressors, i.e., it is based on (2). Finally, the setup of "GEE" is the same as "With d_0 and z_i", but a different estimation method is used.

From Table 5, we can get a few interesting results. First, when more time-invariant regressors are present, $\frac{\sigma_a^2}{\sigma_a^2+1}$ gets smaller. This is because more variation of η_i is explained. Second, the coefficients of $d_{i,t-1}$ for the first four methods are greater than 1, which seems unreasonable. This coefficient is less than 1 only for the estimates based on (2), which means that including time-invariant variables is important for our data. Third, the effect of STC critically depends on whether observed time-invariant effects are controlled, which implies that there is important correlation between STC and z_i. When this correlation is controlled, the effect of STC is significantly negative as expected, due to the cost effect as mentioned in the Introduction. Fourth, the FTA has only weak effect on trade occurrence, population does not have significant effects, while GDP per capita has very significant effect.[9] It seems that the wealth of trade partners rather than population or FTA determines whether a trade will happen. Fifth, d_{i0} has very significant effect on d_{it} as expected, but the effect is less than that of $d_{i,t-1}$. Sixth, note that GEE estimates Wooldridge's coefficients divided by $\sqrt{1+\widehat{\sigma}_a^2}$. For comparison, we list the corresponding (Wooldridge's) estimates in the last column of Table 5. It seems that the GEE estimates are comparable to Wooldridge's estimates, which indicates that autocorrelation among u_{it} does not bias Wooldridge's estimates much in our data.

We next report APEs of STC in different scenarios. For reference, we describe how different APEs are calculated here. Our calculation is based on the GEE estimation. To be specific, we describe how the APE of STC when $d_{i,t-1} = d_{t-1}$, FTA$= f$ and GDP$= gdp$ is calculated; other APEs are similar or simpler. This APE is denoted as $APE(d_{t-1}, f, gdp)$ and defined as

$$\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Phi\left(\widehat{\gamma}_{STC} + \widehat{\phi}d_{t-1} + \widehat{\gamma}_{FTA}f + \widehat{\gamma}_{GDP}gdp + \underline{z}_{it}\underline{\widehat{\gamma}} + z_i\widehat{\psi} + \widehat{\delta}_t D_t + \widehat{\pi}d_{i0}\right)$$
$$- \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\Phi\left(\widehat{\phi}d_{t-1} + \widehat{\gamma}_{FTA}f + \widehat{\gamma}_{GDP}gdp + \underline{z}_{it}\underline{\widehat{\gamma}} + z_i\widehat{\psi} + \widehat{\delta}_t D_t + \widehat{\pi}d_{i0}\right),$$

where $\underline{z}_{it}$ is defined as z_{it} excluding $(STC_{it}, FTA_{it}, GDP_{it})$, and $\underline{\gamma}$ is defined correspondingly.

We conduct the following average partial analyses. First, in Figure 1, we show the APEs of STC for each year in our sample. It is quite surprising to observe that the three-regimes division in Section 2 is roughly correct for the effects of STC on trade occurrence. During the premature period, the effect of STC is quite volatile and relatively small; during the mature period, the effect remains stable at about -5.55%; during the post-crisis period, the effect becomes unstable again and gets smaller (maybe) to stimulate the world economy. Second, in Table 6, we report the APEs of STC for different combinations of FTA and $d_{i,t-1}$ which seem to be the most important covariates except STC. It is interesting to observe that whether there is FTA does not affect the STC effect, while whether there is trade during the last period will affect the STC effect in this period; specifically, STC has larger effect when there was trade during the last period than when there was no trade. This may be because more information is revealed when trade happened during the last period. This result matches the estimation in Table 5 that $\widehat{\gamma}_{FTA}$ is small while $\widehat{\phi}$ is relatively large. The unconditional APE of STC is about -5.73%. Third, we calculate the STC effect at different deciles of GDP and population in Figure 2 and 3 respectively. Matching the estimation in Table 5 that $\widehat{\gamma}_{GDP}$ is much larger than $\widehat{\gamma}_{POP}$, the STC effect depends on GDP more significantly than on population, and this result is

[9] GDP and POP are the log sums of the GDP per capita and population of the two trade partners, respectively. They proxy for the supply capacities and market capacities of the exporting and the importing countries. Actually, we also included importer and exporter's population and GDP as separate regressors, but the results are qualitatively similar.

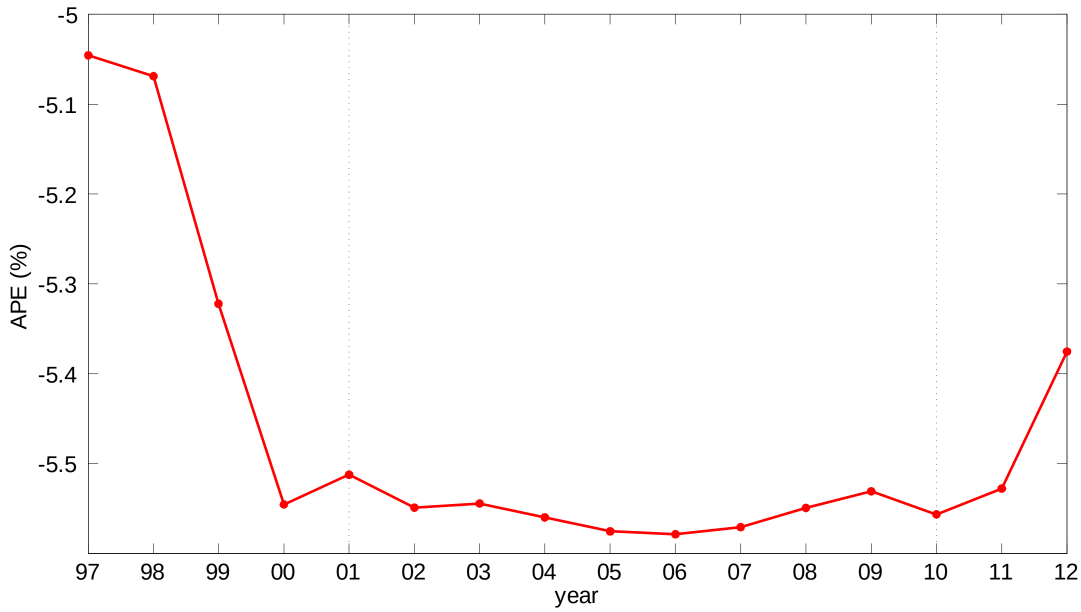

Figure 1: APEs (%) as a Function of Year

invariant to whether conditioning on $d_{i,t-1}$ and/or FTA. To understand this result, note that

$$\frac{dAPE(d_{t-1}, f, gdp)}{dgdp} = \widehat{\gamma}_{GDP}\left[\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\phi\left(\widehat{\gamma}_{STC} + \widehat{\phi}d_{t-1} + \widehat{\gamma}_{FTA}f + \widehat{\gamma}_{GDP}gdp + \underline{z}_{it}\underline{\widehat{\gamma}} + z_i\widehat{\psi} + \widehat{\delta}_t D_t + \widehat{\pi}d_{i0}\right)\right.$$
$$\left. -\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\phi\left(\widehat{\phi}d_{t-1} + \widehat{\phi}d_{t-1} + \widehat{\gamma}_{FTA}f + \widehat{\gamma}_{GDP}gdp + \underline{z}_{it}\underline{\widehat{\gamma}} + z_i\widehat{\psi} + \widehat{\delta}_t D_t + \widehat{\pi}d_{i0}\right)\right],$$

so the slope $APE(d_{t-1}, f, gdp)$ depends on both $\widehat{\gamma}_{GDP}$ and the average density difference in the bracket; results about other APEs can be similarly understood. From Figure 2, it is interesting to observe that when there was trade during the last period, STC has a smaller effect for the richer trade partners, while when there was no trade during the last period, the converse is true.

$APE(d_{t-1}, f)$	$f = 0$	$f = 1$	$\cdot$
$d_{t-1} = 0$	-6.406^{***}	-6.497^{***}	-6.415^{***}
$d_{t-1} = 1$	-7.078^{***}	-7.023^{***}	-7.057^{***}
$\cdot$	-5.731^{***}	-5.764^{***}	-5.727^{***}

Table 6: APEs of STC in Percentage

Note: ***: significant at 1% level

4 Dynamic Trading Process

Based on our data structure, we extend the method proposed in Semykina and Wooldridge (2013) to estimate the effect of STC on the trade volume. Semykina and Wooldridge (2013) apply the ideas in Wooldridge (2005a,b) and Heckman (1976, 1979) to the dynamic panel with sample selection. Their estimation is based

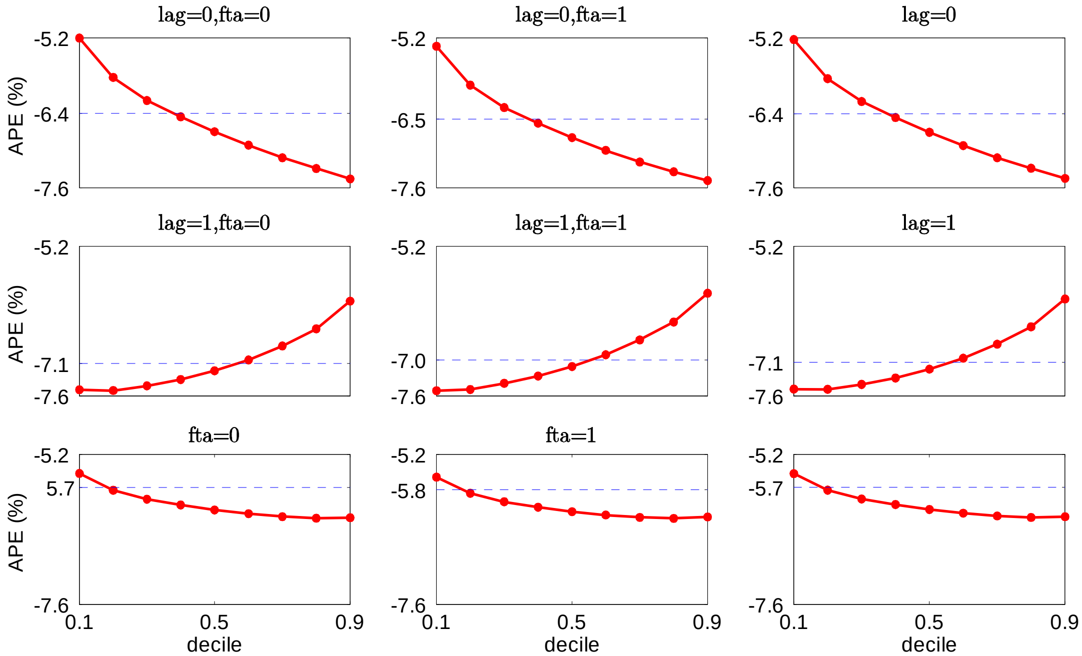

Figure 2: APEs (%) as a Function of the Deciles of GDP Per Capita Conditional on $d_{i,t-1}$ and FTA: Horizontal Dashed Lines for APEs in Table 6

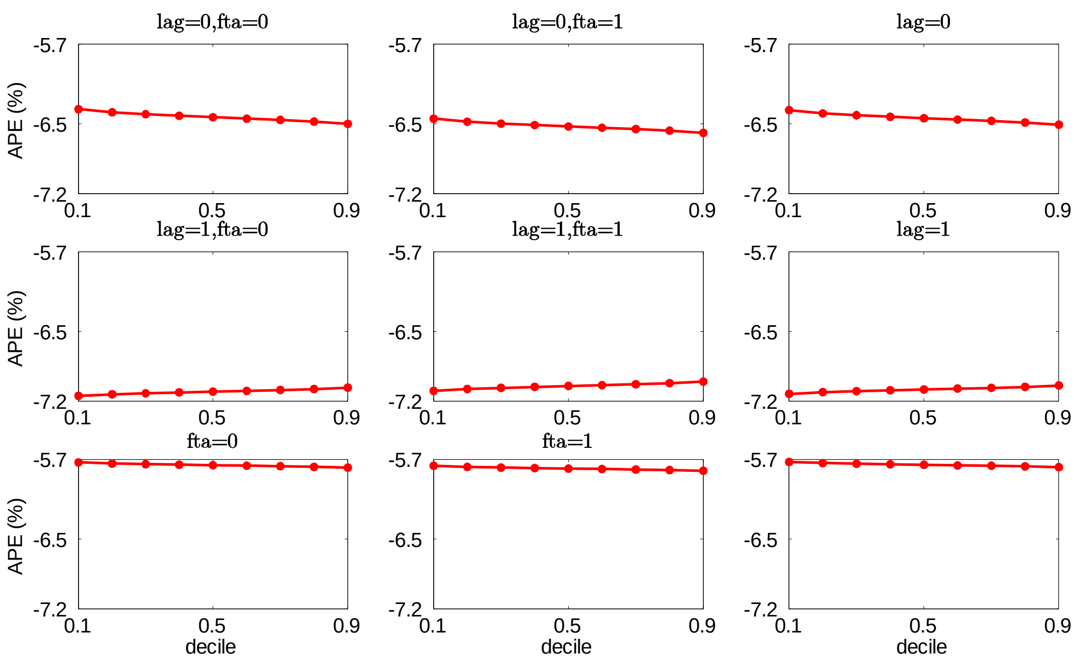

Figure 3: APEs (%) as a Function of the Deciles of Population Conditional on $d_{i,t-1}$ and FTA

on levels rather than first differences as in Kyriazidou (2001) and Gayle and Viauroux (2007). Since there does not exist ready-made computer code for our purpose, we design our own Matlab code which is available upon request. For comparison, we also report the GEE and FD-IV estimates.

4.1 Empirical Specification and Identification

We specify the outcome equation as

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \delta_{1t}D_t + \alpha_i + \varepsilon_{it}, \tag{3}$$

where x_{it} is the same as z_{it} in the participation equation which is specified as

$$d_{it} = 1(\phi d_{i,t-1} + z_{it}\gamma + \delta_{2t}D_t + \eta_i + u_{it} \geq 0). \tag{4}$$

We essentially use the CRE model, that is, we assume the unobserved heterogeneities are controlled by

$$\alpha_i = x_i\varphi + y_{i0}\pi_1 + a_{i1}, \tag{5}$$

and

$$\eta_i = z_i\psi + d_{i0}\pi_2 + a_{i2}, \tag{6}$$

where x_i is a subset of z_i,[10] and a_{i1} and a_{i2} are independent of all control variables. Recursively substituting back for $y_{i,t-1}$, we have

$$y_{it} = \rho^t y_{i0} + \left(\sum_{j=0}^{t-1} \rho^j x_{i,t-j}\right)\beta + \sum_{j=0}^{t-1} \rho^j \delta_{1,t-j} D_{t-j} + \alpha_i \sum_{j=0}^{t-1} \rho^j + \sum_{j=0}^{t-1} \rho^j \varepsilon_{i,t-j}, t = 1, \cdots, T,$$

when y_{i0} is observed. If y_{i0} is not observed, we approximate it by

$$y_{i0} = \sum_{s=0}^{T_0} \underline{x}_{i,-s}\kappa_s + \underline{x}_i\varkappa + b_i, \tag{7}$$

where $\underline{x}_{i,-s}$ is the presample value of $\underline{x}_{it}$ with $\underline{x}_{it}$ being x_{it} excluding the STC variable,[11] $\underline{x}_i$ is x_i excluding $\overline{x}_i$,[12] and b_i is independent of all control variables. To reduce the number of parameters, we let $T_0 = 1$; i.e., we use $\underline{x}_{i,t}$ at 1995 and 1996 to approximate y_{it} at 1996. A similar strategy to approximate y_{i0} is used in Heckman (1981b) to control the endogeneity of d_{i0} in a dynamic binary choice process. The difference of our model and Semykina and Wooldridge (2013) includes three aspects. First, in their model, $\psi = 0$, so there is no observable heterogeneity. Second, they assume α_i, η_i and y_{i0} are linear functions of $z_{i1}, \cdots, z_{iT}$, while we model them in a different way (especially for y_{i0}). Third, they assume the time-varying variable x_{it} is a subset of z_{it}, while in our case, they are the same but x_i and z_i are different. The key assumption of our model is that the endogeneity of the outcome equation is only from the time-invariant characteristics of the two trade partners, so we need only include extra time-invariant instruments in the participation equation for identification.

[10] x_i includes distance but excludes common language, common religion, previous colony and adjacency which seem to affect trade occurrence but not trade amount.

[11] We exclude STC because $STC_{i0} = 0$ for all i. Also, as mentioned at the beginning of Section 2, STC is available only from 1997.

[12] We here implicitly assume y_{i0} is affected by information at $t = 0$ and before.

Under the above assumptions, when y_{i0} is observable, the conditional mean of y_{it} given all control variables (and y_{i0}) and $d_{it}=1$ is

$$\begin{aligned} m_{it}^0(\theta^0;\vartheta) &\equiv \rho^t y_{i0} + \left(\sum_{j=0}^{t-1}\rho^j x_{i,t-j}\right)\beta + \sum_{j=0}^{t-1}\rho^j \delta_{1,t-j}D_{t-j} + \frac{1-\rho^t}{1-\rho}\left(x_i\varphi + y_{i0}\pi_1\right) + \xi_t^0\lambda_{it} \\ &= \left(\sum_{j=0}^{t-1}\rho^j x_{i,t-j}\right)\beta + \sum_{j=0}^{t-1}\rho^j\delta_{1,t-j}D_{t-j} + \left(\rho^t + \frac{1-\rho^t}{1-\rho}\pi_1\right)y_{i0} + \frac{1-\rho^t}{1-\rho}x_i\varphi + \xi_t^0\lambda_{it}, \end{aligned}$$

where $\theta^0 = \left(\rho,\beta,\delta_1,\varphi,\pi_1,\xi^0\right)$ with $\delta_1 = (\delta_{11},\cdots,\delta_{1T})$ and $\xi^0 = (\xi_1^0,\cdots,\xi_T^0)$, $\vartheta = (\phi,\gamma,\delta_2,\psi,\pi_2)$ with $\delta_2 = (\delta_{21},\cdots,\delta_{2T})$ collects the parameters in λ_{it} which is the Heckman correction term, $t=1,\cdots,T$, and

$$\xi_t^0\lambda_{it} = E\left[v_{it1}^0|d_{it}=1\right]$$

with

$$v_{it1}^0 = \frac{1-\rho^t}{1-\rho}a_{1i} + \sum_{j=0}^{t-1}\rho^j\varepsilon_{i,t-j}.$$

When y_{i0} is unobservable, the conditional mean of y_{it} given all control variables and $d_{it}=1$ is

$$\begin{aligned} m_{it}(\theta;\vartheta) &\equiv \rho^t\sum_{s=0}^{T_0}\underline{x}_{i,-s}\kappa_s + \rho^t x_i\varkappa + \left(\sum_{j=0}^{t-1}\rho^j x_{i,t-j}\right)\beta + \sum_{j=0}^{t-1}\rho^j\delta_{1,t-j}D_{t-j} + \frac{1-\rho^t}{1-\rho}\left[x_i\varphi + \pi_1\sum_{s=0}^{T_0}\underline{x}_{i,-s}\kappa_s + \pi_1 x_i\varkappa\right] + \xi_t\lambda_{it} \\ &= \left(\sum_{j=0}^{t-1}\rho^j x_{i,t-j}\right)\beta + \sum_{j=0}^{t-1}\rho^j\delta_{1,t-j}D_{t-j} + \left(\rho^t + \frac{1-\rho^t}{1-\rho}\pi_1\right)\left[\sum_{s=0}^{T_0}\underline{x}_{i,-s}\kappa_s + \underline{x}_i\varkappa\right] + \frac{1-\rho^t}{1-\rho}x_i\varphi + \xi_t\lambda_{it}, \end{aligned}$$

where $\theta = (\rho,\beta,\delta_1,\varphi,\pi_1,\kappa,\varkappa,\xi)$ with $\kappa = (\kappa_0,\cdots,\kappa_{T_0})$ and $\xi = (\xi_1,\cdots,\xi_T)$, and

$$\xi_t\lambda_{it} = E\left[v_{it1}|d_{it}=1\right],$$

with

$$v_{it1} = \rho^t b_i + \frac{1-\rho^t}{1-\rho}\left[a_{1i} + b_i\pi_1\right] + \sum_{j=0}^{t-1}\rho^j\varepsilon_{i,t-j}.$$

In our data, there is no trade in year 1996 (or $t=0$) for 7937 routes among the totally 11282 trade routes. So missing initial observation is a serious problem for our estimation. From Table 4, by approximating y_{i0}, we increase about 90% $(0.447/0.237-1)$ data points.

Note that $d_{it} = 1(\phi d_{i,t-1} + z_{it}\gamma + \delta_{2t}D_t + z_i\psi + d_{i0}\pi_2 + v_{it2} \geq 0)$ with $v_{it2} = a_{i2} + u_{it}$. If the joint distribution of v_{it1}^0 and v_{it2} is normal,[13] then

$$E\left[v_{it1}^0|d_{it}=1\right] = \frac{\sigma_{12t}^0}{\sigma_{2t}}\lambda\left(\frac{\phi d_{i,t-1} + z_{it}\gamma + \delta_{2t}D_t + z_i\psi + d_{i0}\pi_2}{\sigma_{2t}}\right)$$

where $\sigma_{12t}^0 = Cov(v_{it1}^0, v_{it2})$, $\sigma_{2t}^2 = Var(v_{it2})$ and $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio. In this case,

$$\xi_t^0 = \frac{\sigma_{12t}^0}{\sigma_{2t}} \text{ and } \lambda_{it} = \lambda\left(\frac{\phi d_{i,t-1} + z_{it}\gamma + \delta_{2t}D_t + z_i\psi + d_{i0}\pi_2}{\sigma_{2t}}\right).$$

[13] This assumption can be relaxed to $v_{it1}^0 = \xi_t^0 v_{it2} + \epsilon_{it}$ with ϵ_{it} independent of v_{it2} and v_{it2} following a nonnormal distribution.

Here, note that v_{it1}^0 does not take the error component form, so σ_{12t}^0 and ξ_t^0 should be time-varying in general. Similarly, $\xi_t = \frac{\sigma_{12t}}{\sigma_{2t}}$ with $\sigma_{12t} = Cov(v_{it1}, v_{it2})$ should be time-varying, and $\xi_t \neq \xi_t^0$ in general. Note further that since the corrections $E\left[v_{it1}^0|d_{it}=1\right]$ and $E\left[v_{it1}|d_{it}=1\right]$ are calculated for each t, the serial correlation among ε_{it} is irrelevant to our estimation. Also, even $Var(u_{it})$ is allowed to be time-varying because σ_{2t} may depend on t.

In summary, the conditional mean of y_{it} given all control variables, d_{i0} and $d_{it}=1$ is

$$m_{it}^0(\theta^0;\vartheta)d_{i0} + m_{it}(\theta;\vartheta)(1-d_{i0}).$$

Given this form of the conditional mean of y_{it}, we can estimate θ^0 and θ based on a two-step procedure. The first step is to estimate the nuisance parameter ϑ. This can be done in two methods. First, we can estimate ϑ by employing the error component structure of v_{it2} as in the last section; i.e., $\left(\phi,\gamma,\delta_2,\psi,\pi_2,\sigma_{2t}^2\right)$ can be estimated by $\left(\widehat{\phi},\widehat{\gamma},\widehat{\delta}_2,\widehat{\psi},\widehat{\pi}_2,1+\widehat{\sigma}_a^2\right)$. Then λ_{it} can be estimated by $\widehat{\lambda}_{it} \equiv \lambda\left(\frac{\widehat{\phi}d_{i,t-1}+z_{it}\widehat{\gamma}+\widehat{\delta}_{2t}D_t+z_i\widehat{\psi}+d_{i0}\widehat{\pi}_2}{\sqrt{1+\widehat{\sigma}_a^2}}\right)$. Second, we can estimate ϑ separately for each time period; i.e., $\left(\frac{\phi}{\sigma_{2t}},\frac{\gamma}{\sigma_{2t}},\frac{\psi}{\sigma_{2t}},\frac{\pi_2}{\sigma_{2t}}\right)$ can be estimated by $\left(\widehat{\phi}_t,\widehat{\gamma}_t,\widehat{\psi}_t,\widehat{\pi}_{2t}\right)$ which are the coefficients in a probit regression of d_{it} on $(d_{i,t-1},z_{it},z_i,d_{i0})$ for each t. The second method is more robust since it even allows (ϕ,γ,ψ,π_2) and/or $Var(u_{it})$ to depend on t, while the first method is more efficient when the error component structure of v_{it2} is satisfied. To be consistent with the participation equation estimation in Section 3, we use the GEE estimator to calculate the inverse Mills ratio. Here, we must point out a serious problem in the two estimation procedures above. In calculating $E\left[v_{it1}^0|d_{it}=1\right]$ and $E\left[v_{it1}|d_{it}=1\right]$, we are implicitly conditioning also on the endogenous variable $d_{i,t-1}$ which in turn depends on $d_{i,t-2}$ etc. This is why the complicated estimation procedure in Gayle and Viauroux (2007) is required. To avoid this complication, we follow the suggestion of Semykina and Wooldridge (2013) - replace $d_{i,t-1}$ by $z_{i,t-1}$. This replacement is a reasonable approximation because $d_{i,t-1}$ is determined by $d_{i,t-2}, z_{i,t-1}, D_{t-1}, z_i$ and d_{i0} and much variation in $d_{i,t-2}$ can be explained by z_i and d_{i0} which have already been included as regressors.

The second step is to estimate θ^0 and θ by pooled nonlinear least squares (NLS) on the selected sample with λ_{it} substituted by $\widehat{\lambda}_{it}$. Specifically, $\left(\theta^0,\theta\right)$ is estimated by the solution to the minimization problem

$$\min_{\theta^0,\theta}\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}d_{it}\left[y_{it}-m_{it}^0(\theta^0;\widehat{\vartheta})d_{i0}-m_{it}(\theta;\widehat{\vartheta})(1-d_{i0})\right]^2,$$

where $\widehat{\vartheta}$ is the estimated nuisance parameter in the first step. Of course, we can improve the efficiency of our estimation of $\left(\theta^0,\theta\right)$ by using a GMM estimator as in Section 4 of Semykina and Wooldridge (2013), but we will not pursue this in this paper. Our final estimation procedure is a combination of the GEE method in the first step and the pooled NLS in the second step. Our decision is based on two considerations. First, we have already made many assumptions on our model to make the estimation feasible, so it is better to make the estimation procedure as robust as possible in our framework. In general, a more complicated estimation procedure tends to be less robust. Second, the standard errors of a more complicated procedure is harder to compute. From the simulation studies in Semykina and Wooldridge (2013), when n is large (as in our case), the performances of NLS and GMM are quite close.

The objective function is nonlinear in θ and θ^0. Given that the dimension of $\left(\theta,\theta^0\right)$ equals 147, the usual optimization routine is hard to apply to find the minimizer. Carefully checking $m_{it}^0(\theta^0;\widehat{\vartheta})$ and $m_{it}(\theta;\widehat{\vartheta})$, we find that they are nonlinear only in (ρ,π_1) and linear in the remaining parameters, so a concentrated procedure can be used. Figure 4 shows the concentrated objective function as a function of (ρ,π_1). From

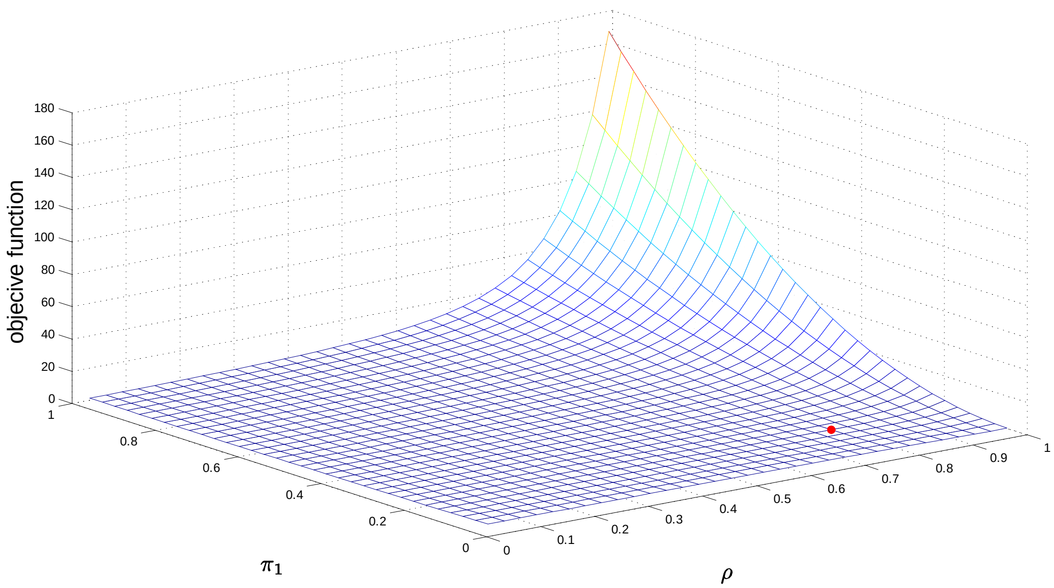

Figure 4: Concentrated Objective Function

the figure, $\widehat{\pi}_1 = .153$ is smaller than $\widehat{\rho} = .769$ as expected - the effect of $y_{i,t-1}$ to y_{it} should be larger than that of y_{i0}.

4.2 GEE and FD-IV Estimators

We consider three GEE estimators. For all the estimators, we assume the data missing is random, so no Heckit correction is required. The first estimator uses the observations with $(y_{it}, y_{i,t-1})$ observed and neglects y_{i0}. This estimator does not consider the impact of initial conditions, so

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \delta_{1t} D_t + x_i \varphi + a_{i1} + \varepsilon_{it}.$$

From Table 4, this estimator uses 35.5% of the total observations. The second estimator uses only the observations with $(y_{it}, y_{i,t-1}, y_{10})$ observed, so from Table 4, only 22.2% of the total observations are used. We assume

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \delta_{1t} D_t + x_i \varphi + y_{i0}\pi_1 + a_{i1} + \varepsilon_{it}. \tag{8}$$

The third estimator uses the observations with $(y_{it}, y_{i,t-1})$ observed and approximates y_{i0} by (7). So when y_{i0} is available, y_{it} follows (8); when y_{i0} is not available,

$$y_{it} = \rho y_{i,t-1} + x_{it}\beta + \delta_{1t} D_t + x_i \varphi + \left(\sum_{s=0}^{T_0} \underline{x}_{i,-s} \kappa_s + \underline{x}_i \varkappa \right) \pi_1 + b_i \pi_1 + a_{i1} + \varepsilon_{it}.$$

This estimator also uses 35.5% of the total observations. Comparing to the three GEE estimators, our estimator uses more observations (44.7% of all observations), and also corrects the sample selection bias.[14]

[14] Note also that our estimator uses more variations in the time-varying variable x_{it} through the term $\sum_{j=0}^{t-1} \rho^j x_{i,t-j}$.

In Section 5 of Semykina and Wooldridge (2013), the authors "nominally" specify the sample selection in our context as

$$1(d_{it} = 1, d_{i,t-1} = 1) = 1(z_{it}\gamma + z_{i,t-1}\gamma_{-1} + \delta_{2t}D_t + z_i\psi + d_{i0}\pi_2 + a_{i2} + u_{it} \geq 0)$$

in the first and third GEE estimators and

$$1(d_{it} = 1, d_{i,t-1} = 1, d_{i0} = 1) = 1(z_{it}\gamma + z_{i,t-1}\gamma_{-1} + z_{i0}\gamma_0 + \delta_{2t}D_t + z_i\psi + a_{i2} + u_{it} \geq 0),^{15}$$

in the second GEE estimator. However, their specification is only for testing purpose and may be seriously misspecified in estimation.

For completeness, we also report the FD-IV estimator of Anderson and Hsiao. Our estimators are the two-step estimators of Arellano and Bond (1991) associated with the Windmeijer (2005) bias-corrected robust standard error. Besides the weak IV problem mentioned in Section 2, this estimator does not correct the selection effect and uses only 29.9% of the total observations.

	FD-IV	GEE, no y_0	GEE, only y_0	GEE, appr. y_0	Heckit	Heckit, with d_{t-1}
y_{t-1}	.212***	.584***	.591***	.609***	.769***	.744
STC	−.00889	.00879	.0172	.00323	.0150***	.0482
FTA	.0858*	.114***	.113***	.139***	.0832	.0967
POP	.370**	.231***	.102	.184***	.177*	.257
GDP	.330***	.290***	.291***	.277***	.182***	.155
y_0	-	-	.278***	.263***	.153***	.168

Table 7: Parameter Estimates Under Different Specifications and Methods

Note: *: significant at 10% level, **: significant at 5% level, ***: significant at 1% level

4.3 Empirical Results

The estimates of all methods in the last two subsections are summarized in Table 7. From Table 7, we can draw a few conclusions. First, interestingly, the most different estimates among all methods occur in the coefficient of STC. In fact, all estimates of the STC effect except our method are insignificant. The FD-IV estimator suffers from both the weak IV and sample selection problem, so its estimates of almost all coefficients are very different from ours. The three GEE estimators suffer only from the sample selection problem, so most estimates are quite close to ours except the coefficient of STC. Second, similar as in the dynamic participation process, y_{t-1}, GDP and the initial condition have significant effects on the trade volume, and the effects of FTA and population are not very significant. Differently, STC has a positive effect on the trade volume, but a negative effect on the trade occurrence. In other words, the signaling effect of STC seems to dominate the cost effect when the trade happens even if a STC is proposed. Also, this positive effect (1.5%) of STC is less than its negative effect to trade occurrence (−5.73%) in absolute value. Third, for comparison, we also report the results when $d_{i,t-1}$ rather than $z_{i,t-1}$ is used in the participation equation; this is reported in the last column of Table 7. Given that the correlation between the two λ_{it}'s is 0.922, most coefficients under these two specifications are similar except the coefficient of STC. From this point and the first point above, the estimation of STC is sensitive to the specification of the model. This is understandable given that STC appears only 1451 times among all 180512 observations. To study the effects of such rare events, correct specification of the model is critical. Finally, based on the testing results unreported in Table

[15] Note that $STC_{i0} = 0$ for all i so should be omitted to avoid multicollinearity.

7, we cannot reject the null that $\xi^0 = \xi$ (with the p-value equal to 0.257) but strongly reject the null that $\left(\xi^0, \xi\right) = 0$ (with the p-value equal to 0.000). These testing results are intuitively understandable: after approximating y_{i0} by presample information, the remaining unexplained variation in y_{i0} is neglectable; on the other hand, the selection effect is indeed statistically significant and nonneglectable.

The effect of STC on trade volume, i.e., 1.5% in Table 7, is for all trade routes. As suggested by Vella (1988) in the cross-sectional environment, it is more interesting to compare the average trade volumes only among trade routes with trade happening. Specifically, our estimation of the STC effect on trade volume is as follows:

$$\begin{aligned}\Delta = & \frac{1}{\sum_{i=1}^{n}\sum_{t=1}^{T} 1(d_{it}=1)} \sum_{i=1}^{n}\sum_{t=1}^{T} d_{it} \exp\left\{\widehat{E}\left[y_{it}|d_{it}=1, STC_{it}=1\right]\right\} \\ & - \frac{1}{\sum_{i=1}^{n}\sum_{t=1}^{T} 1(d_{it}=1)} \sum_{i=1}^{n}\sum_{t=1}^{T} d_{it} \exp\left\{\widehat{E}\left[y_{it}|d_{it}=1, STC_{it}=0\right]\right\},\end{aligned}$$

where

$$\widehat{E}\left[y_{it}|d_{it}=1, STC_{it}\right] = m_{it}^0(\widehat{\theta}^0; \widehat{\vartheta}, STC_{it})d_{i0} + m_{it}(\widehat{\theta}; \widehat{\vartheta}, STC_{it})(1-d_{i0}),$$

$m_{it}^0(\widehat{\theta}^0; \widehat{\vartheta}, STC_{it})$ is $m_{it}^0(\widehat{\theta}^0; \widehat{\vartheta})$ with STC_{it} taking the specified value, and $m_{it}(\widehat{\theta}; \widehat{\vartheta}, STC_{it})$ is similarly defined. To provide an impression on the relative magnitude of Δ, we only report the percentage change in trade volume with the scenario of $STC_{it} = 0$ as the base case, i.e., Δ divided by the second term in its definition.[16]

Vella (1988) evaluates Δ at the average values of covariates and even λ_{it} in each group ($STC = 0$ and $STC = 1$). In our case, the group with $STC = 1$ is very small, so we suggest a measure similar to the APE in Section 3.2. In our measure, STC has two effects on Δ: first, it affects Δ indirectly through λ_{it}, i.e., through the participation probability; second, it affects Δ directly. We also measure the effect of STC on trade volume for each year as

$$\begin{aligned}\Delta_t = & \frac{1}{\sum_{i=1}^{n} 1(d_{it}=1)} \sum_{i=1}^{n} d_{it} \exp\left\{\widehat{E}\left[y_{it}|d_{it}=1, STC_{it}=1\right]\right\} \\ & - \frac{1}{\sum_{i=1}^{n} 1(d_{it}=1)} \sum_{i=1}^{n} d_{it} \exp\left\{\widehat{E}\left[y_{it}|d_{it}=1, STC_{it}=0\right]\right\}.\end{aligned}$$

Note that

$$\Delta = \sum_{t=1}^{T} w_t \Delta_t \text{ with } w_t = \frac{\sum_{i=1}^{n} 1(d_{it}=1)}{\sum_{i=1}^{n}\sum_{t=1}^{T} 1(d_{it}=1)}.$$

Similar to Δ, we only report the percentage change of Δ_t. This is shown in Figure 5. As a benchmark, the percentage change of Δ is shown as a dashed line in the figure.

From Figure 5, the percentage change of Δ is 1.2% which is smaller than 1.5%, that is, the selection effect slightly dampens the trade volume, but not in a significant way. As in the participation process in Section 3, there seems to be three regimes in the effect of STC on trade volume. Before 2001, the selection effect boosts the effect of STC ($\Delta_t > 1.5\%$), while after the financial crisis, the converse claim seems to be more suitable. At the beginning of financial crisis (2008-2010), there seems to be a downward jump in the effect of STC. After 2010, the signaling effect of STC on trade seems to start recovering.

[16] Another advantage of the percentage change is that we can neglect the constant term in inverting the log transformation; see p212-215 of Wooldridge (2012).

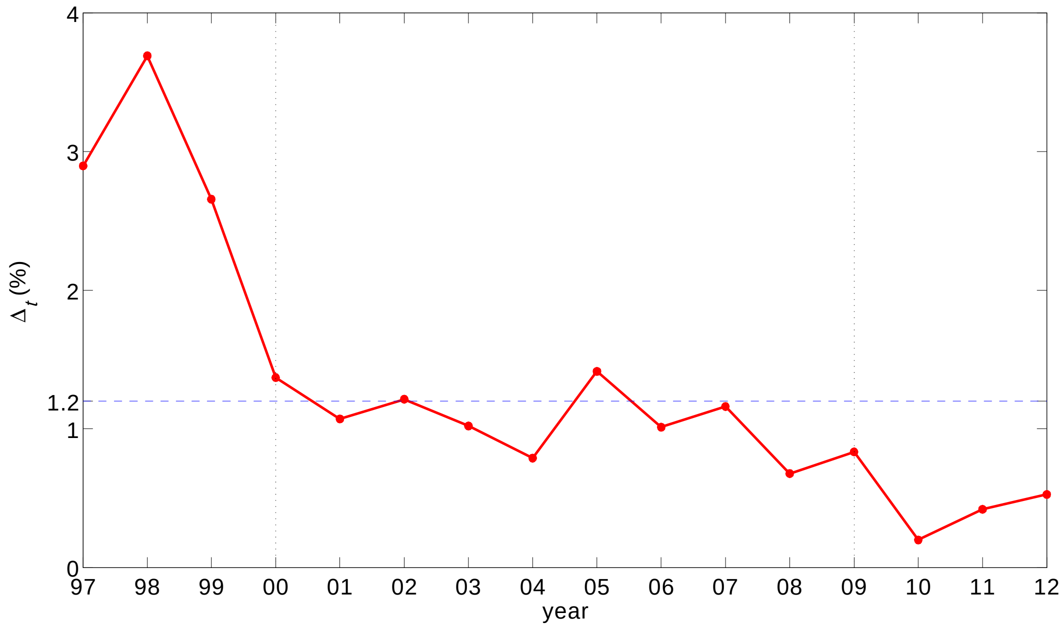

Figure 5: Effect of STC on Trade Volume Among Trade Routes with $d_{it} = 1$: Horizontal Dashed Line for Δ

5 Conclusion

This paper studies the effects of STC on the trade occurrence and trade volume based on a dynamic panel data model with attrition. Different from the existing literature, we provide a rigorous treatment on the econometric aspects of the problem. Our estimation scheme is a combination of the correlated random effects specification and the robustness to serial correlation of idiosyncratic errors. Some interesting conclusions are drawn based on our analysis. For example, STC has a negative effect (-5.73%) on the probability of trade occurrence; STC has a positive effect (1.2%) on the trade volume conditioning on trade happening even if a STC is proposed; the effects of STC are time-specific - there are three regimes in its effect process: the premature period (1997-2001), the mature period (2002-2008) and the post-crisis period (2009-2012).

Our paper serves only as a rigorous starting point of this interesting topic. There are many other problems untouched in this paper due to data constraints and we conclude by outlining some of these problems here. First, we do not distinguish different SPS measures in this paper. This is because STC is a rare event and finer classification of STC makes it even rarer, leading to unreliable inference. Crivelli and Gröschl (2012) study the STC effect for two types of SPS measures - measures related to conformity assessment and measures related to product characteristics. Since they compile all products together, their data size is 5,452,530, which may alleviate the sparsity of STC. Second, one STC may last for a few years and we assume its effect remains the same during its existence in this study. However, intuitively, this effect should decline with time elapsing. How to model such a declining process is intriguing. Third, a STC is sometimes only partially resolved, but it is hard to judge how much is resolved. It is better to describe its state as an interval in $(0, 1)$ based on expert opinion. Such interval data may only generate partially identified estimates (see, e.g., Manski and Tamer (2002)) and the estimation and inference in this new context is a promising area for future research.

References

Anders, S.M. and J.A. Caswell, 2009, Standards as Barriers versus Standards as Catalysts: Assessing the Impact of HACCP Implementation on US Seafood Imports, *American Journal of Agricultural Economics*, 91, 310-321.

Anderson, T.W., and C. Hsiao, 1981, Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association*, 76, 598-606.

Anderson, T.W., and C. Hsiao, 1982, Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18, 47-82.

Arellano, M. and S. Bond, 1991, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.

Arellano, M., O. Bover and J.M. Labeaga, 1999, Autoregressive Models with Sample Selectivity for Panel Data, *Analysis of Panels and Limited Dependent Variable Models*, in honor of G.S. Maddala, C. Hsiao, K. Lahiri, L.-F. Lee and M.H. Pesaran, eds., Cambridge: Cambridge University Press, 23-48.

Arellano, M. and R. Carrasco, 2003, Binary Choice Panel Data Models with Predetermined Variables, *Journal of Econometrics*, 115, 125-157.

Arellano, M. and J. Hahn, 2006, A Likelihood-Based Approximate Solution to the Incidental Parameter Problem in Dynamic Nonlinear Models with Multiple Effects, mimeo, CEMFI.

Bank, A.D., 2008, How to Design, Negotiate, and Implement a Free Trade Agreement in ASIA, Philippines: Asian Development Bank.

Bartolucci, F. and V. Nigro, 2010, A Dynamic Model for Binary Panel Data with Unobserved Heterogeneity Admitting a $\sqrt{n}$-Consistent Conditional Estimator, *Econometirca*, 78, 719-733.

Binder, M., C. Hsiao and M.H. Pesaran, 2005, Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, *Econometric Theory*, 21, 795-837.

Blundell, R.. and S. Bond, 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 87, 115-143.

Bover, O. and M. Arellano, 1997, Estimating Dynamic Limited Dependent Variable Models from Panel Data, *Investigaciones Economicas*, 21, 141-165.

Butler, J.S. and R. Moffitt, 1982, A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model, *Econometrica*, 50, 761-764.

Carro, J.M., 2007, Estimating Dynamic Panel Data Discrete Choice Models with Fixed Effects, *Journal of Econometrics*, 140, 503-528.

Chamberlain, G., 1980, Analysis of Covariance with Qualitative Data, *Review of Economic Studies*, 47, 225-238.

Chamberlain, G., 1982, Multivariate Regression Models for Panel Data, *Journal of Econometrics*, 18, 5-46.

Chamberlain, G., 1984, Panel Data, *Handbook of Econometrics*, Vol. 2, Z. Griliches and M. Intriligator, eds., Elsevier Science B.V., Ch. 22, 1247-1318.

Codex-Secretariat, 2003-2012, The Reports of the 26th to 35th Sessions of Codex Alimentarius Commission.

Crivelli, P. and J. Gröschl, 2012, SPS Measures and Trade: Implementation Matters, Staff Working Paper ERSD-2012-05, WTO.

Dhaene, G., K. Jochmans and B. Thuysbaert, 2006, Jackknife Bias Reduction for Nonlinear Dynamic Panel Data Models with Fixed Effects, mimeo, K.U. Leuven.

Digges, P., A. Gordon, and A. Marter, 1997, *International Markets for African Agricultural Exports: Agricultural Policy Reform and Agricultural Exports*, Greenwich: NRI.

Fernández-Val, I., 2009, Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models, *Journal of Econometrics*, 150, 71-85.

Fernández-Val, I. and F. Vella, 2011, Bias Corrections for Two-Step Fixed Effects Panel Data Estimators, *Journal of Econometrics*, 163, 144-162.

Gayle, G.L. and C. Viauroux, 2007, Root-N Consistent Semiparametric Estimators of a Dynamic Panel-Sample-Selection Model, *Journal of Econometrics*, 141, 179-212.

Hahn, J. and G. Kuersteiner, 2011, Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects, *Econometric Theory*, 27, 1152-1191.

Hahn, J. and W. Newey, 2004, Jackknife and Analytical Bias Reduction for Nonlinear Panel Models, *Econometrica*, 72, 1295-1319.

Heckman, J.J., 1976, The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models, *Annals of Economic and Social Measurement*, 5, 475-492.

Heckman, J.J., 1979, Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-161.

Heckman, J.J., 1981a, Statistical Models for Discrete Panel Data, in *Structural Analysis of Discrete Data with Econometric Applications*, C.F. Manski and D. McFadden, eds., Cambridge, MA: MIT Press, 114-178.

Heckman, J.J., 1981b, The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process, *Structural Analysis of Discrete Data with Econometric Applications*, C.F. Manski and D. McFadden, eds., Cambridge, MA: MIT Press, 179-195.

Heckman, J.J., 1981c, Heterogeneity and State Dependence, *Studies of Labor Markets,* S. Rosen, eds., Chicago: University of Chicago Press, 91-140.

Hillman, J., 1997, Non-Tariff Agricultural Trade Barriers Revisited, *Understanding Technical Barriers to Agricultural Trade*, D. Orden and D. Roberts, eds., St. Paul, MN: University of Minnesota, 1-32.

Honoré, B.E. and A. Lewbel, 2002, Semiparametric Binary Choice Panel Models without Strictly Exogeneous Regressors, *Econometrica*, 70, 2053-2063.

Honoré, B.E. and E. Tamer, 2006, Bounds on Parameters in Panel Dynamic Discrete Choice Models, *Econometrica*, 74, 611-629.

Honoré, B. F. Vella and M. Verbeek, 2008, Attrition, Selection Bias and Censored Regressions, *The Econometrics of Panel Data : Fundamentals and Recent Developments in Theory and Practice*, 3rd ed., Mátyás, L. and P. Sevestre, eds., Boston: Kluwer, Chapter 12, 385-418.

Hsiao, C., M.H. Pesaran and A.K. Tahmiscioglu, 2002, Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, *Journal of Econometrics*, 109, 107-150.

Hyslop, D.R., 1999, State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women, *Econometrica*, 67, 1255-1294.

Jaffee, S., 1999, *Southern African Agribusiness: Gaining Through Regional Collaboration*, Washington DC: World Bank.

Kyriazidou, E., 1997, Estimation of A Panel Data Sample Selection Model, *Econometrica*, 65, 1335-1364.

Kyriazidou, E., 2001, Estimation of Dynamic Panel Data Sample Selection Models, *Review of Economic Studies*, 68, 543-572.

Liang K.-Y. and S.L. Zeger, 1986, Longitudinal Data Analysis Using Generalized Linear Models, *Biometrika*, 73, 13-22.

Manski, C.F. and E. Tamer, 2002, Inference on Regeressions with Interval Data on a Regressor or Outcome, *Econometrica*, 70, 519-546.

Mundlak, Y., 1978, On the Pooling of Time Series and Cross Section Data, *Econometrica*, 46, 69-85.

National Research Council, 1995, *Standards, Conformity Assessment and Trade*, Washington DC: National-Research Council.

Ndayisenga, F. and J. Kinsey, 1994, The Structure of Nontariff Trade Measures on Agricultural Products in High-Income Countries, *Agribusiness*, 10, 275-292.

Papke, L.E. and J.M. Wooldridge, 2008, Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates, *Journal of Econometrics*, 145, 121-133.

Petrey, L.A. and R.W.M. Johnson, 1993, Agriculture in the Uruguay round: sanitary and phytosanitary measures, *Review of Marketing and Agricultural Economics*, 61, 433-442.

Schlueter, S.W., W. Christine, and T. Heckelei, 2009, Regulatory Policies in Meat Trade: Is There Evidence for Least Trade-Distorting Sanitary Regulations?, *American Journal of Agricultural Economics*, 91, 1484-1490.

Semykina, A. and J.M. Wooldridge, 2013, Estimation of Dynamic Panel Data Models with Sample Selection, *Journal of Applied Econometrics*, 28, 47-61.

Sykes, A.O., 1995, *Product Standards for Internationally Integrated Goods Markets*, Washington DC: Brookings Institution.

Stewart, M., 2006, Maximum Simulated Likelihood Estimation of Random-Effects Dynamic Probit Models with Autocorrelated Errors, *The Stata Journal*, 6, 256-272.

Stewart, M., 2007, The Interrelated Dynamics of Unemployment and Low-Wage Employment, *Journal of Applied Econometrics*, 22, 511-531.

Thilmany, D.D., and C.B. Barrett, 1997, Regulatory Barriers in an Integrating World Food Market, *Review of Agricultural Economics*, 19, 91–107.

Unnevehr, L.J., 1999, Food Safety Issues and Fresh Food Product Exports from LDCs, Paper presented at the conference Agro-Industrialization, Globalization and Development, Nashville.

Vella, F., 1988, Generating Conditional Expectations from Models with Selectivity Bias, *Economics Letters*, 28, 97-103.

Vella, F., 1998, Estimating Models with Sample Selection Bias: A Survey, *Journal of Human Resources*, 33, 127-169.

Verbeek, M. and T.E. Nijman, 1996, Incomplete Panels and Selection Bias, *The Econometrics of Panel Data : A Handbook of the Theory with Applications*, 2nd ed., Mátyás, L. and P. Sevestre, eds., Boston: Kluwer, 449-490.

Windmeijer, F., 2005, A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators, *Journal of Econometrics*, 126, 25-51.

Wooldridge, J.M., 2000, A Framework for Estimating Dynamic, Unobserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables, *Economics Letters*, 68, 245-250.

Wooldridge, J.M., 2005a, Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity, *Journal of Applied Econometrics*, 20, 39-54.

Wooldridge, J.M., 2005b, Unobserved Heterogeneity and Estimation of Average Partial Effects, in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D.W.K. Andrews and J.H. Stock, Cambridge, U.K.: Cambridge University Press, Chapter 3, 27-55.

Wooldridge, J.M., 2012, Introductory Econometrics: A Modern Approach, 5th edition, Mason, Ohio: South-Western Cengage Learning.

Woutersen, T.M., 2002, Robustness Against Incidental Parameters, mimeo, University of Western Ontario.

WTO, 2012, World Trade Report 2012, Geneva: World Trade Organization.

Appendix A: The Background of SPS Agreement

Over the past century, international trade has been expanding significantly. Consumers across the world have more choices of products, and the commodities have become cheaper and cheaper. Agriculture, as a very special and sensitive sector, especially in most developing countries, remains a cornerstone of many economies. Agricultural production and processing offer many low-income countries the possibility to trade their way out of poverty.

The increased movements of products, as a result of rapid growth of international trade (particularly in the food and agricultural sector), may pose health risks to human, animal or plant life of the importing region. Examples of such risks include the spread of insect pests hosted by imported fruit or vegetables, the transfer of infectious animal diseases carried by imported animals or animal products, and food safety risks from inadequate hygienic standards in the production or transportation of the exported food. For consumers throughout the world, it is one of the most fundamental requirements that internationally imported agricultural products are safe.

To ensure food safety, and to avoid introduction of diseases and pests through trade, countries have intentions of imposing regulations to protect human and animal health (sanitary measures) and plant health (phytosanitary measures). In fact, GATT (the General Agreement on Tariffs and Trade) 1947 recognized the need to introduce trade restrictions. According to its Article XX (b), GATT members were permitted to impose SPS measures which, however, have to be **necessary** to protect human, animal or plant life or health. This article did not apply if SPS measures were used as a means of arbitrary or unjustifiable discrimination between countries, or a disguised restriction on international trade. However, this provision is too broad to easily and precisely implement by GATT members. Especially, there is not a proper definition and objective measure for the key word "necessary" in the article. In practice, in the circumstance of decreasing tariffs, quotas and prohibitions (due to successive rounds of multilateral and bilateral negotiations within the GATT framework), the provisions of Article XX (b) in GATT 1947 were increasingly used to justify non-tariff barriers (including SPS measures) to protect domestic industries. As a result, an increasing number of countries and regions began to impose more severe technical regulations and standards to avoid the alleged potential health risks, regardless of whether these measures are scientifically justifiable.

To solve this problem, in the Kennedy Round GATT negotiations, all of the non-tariff barriers (including SPS measures) were considered to be developed as separate agreements or "Codex". Regretfully, such provisions were covered only in the so-called "Plurilateral Trade Agreement", which was binding only on the members who accepted them and did not create either obligations or rights otherwise.

After the Kennedy Round, GATT members increasingly concerned about the growing number and variety of forms of domestic standards, which poses a potential threat to international trade. Therefore, more and more countries believed that the best way to minimize the negative impact of SPS measures on international trade is to cooperate among members in developing international or domestic standards.

In the later Tokyo Round, based on the proposals of the United States to develop uniform technical standards and quality certification agreement, 47 members finally reached and signed the Agreement on Technical Barrier to Trade (TBT Agreement, hereinafter), which includes a range of mandatory and voluntary technical standards on industrial and agricultural products. These technical standards specifically prohibit any arbitrary or unjustifiable discrimination between countries where the same conditions prevail or a disguised restriction on international trade, and urge the signatories to base national measures on international standards. However, no consensus on provisions regarding SPS measures was reached. In the decade after the Tokyo Round, although "TBT Agreement" covered conformity assessment procedures in standard settings of animal and plant products, these procedures did not apply to the technical regulations specific to quarantine measures of food safety and animals and plants health. Practically, these measures have their particularities, and are very different from the general technical trade measures, continuing to cause trade restrictions.

After the Tokyo Round, with the rapid development of the world economy, trade protectionism all over the world gradually rose and especially some developed countries made use of non-traditional methods to protect domestic enterprises under the GATT Article XX (b). Due to more complexity and elusiveness compared to other measures, most required standards related to quality and safety became practical and effective instruments of trade protectionism. As a result, mandatory SPS measures and unjustifiable health standards in import and export trade prevailed. This aroused another round of trade debates particularly between developed countries, and these debates could not be resolved under the existing GATT "Standards Code" or through the prevailing GATT dispute settlement procedures. In the face of this serious impediment to international trade, the SPS issue became a key agenda during the Uruguay Round. After successive and repeated negotiations, the SPS Agreement was signed formally by all WTO members in 1995. The agreement sets out the principles that WTO members can use in establishing national technical regulations and standards for food safety and animal and plant health.

The SPS Agreement has a two-fold objective. First of all, the sovereign rights of WTO members to provide the level of health protection they deem appropriate is recognized. The other aim is to ensure that SPS measures do not represent unnecessary, arbitrary, scientifically unjustifiable, or disguised restrictions on international trade. This is to say, the SPS Agreement admits the need for WTO members to protect themselves from the risks posed by the entry of pests and diseases, but also seeks to minimize any unjustifiable effects of SPS measures on trade. The key principles or provisions of the SPS Agreement include risk assessment, harmonization, equivalence, regional conditions and transparency. Among these, risk assessment, sometimes known as the scientific principle, and harmonization are the core principles of the agreement; we provide more details on them below.

Risk assessment states that members have the right to adopt SPS measures that establish the health protection level they deem appropriate, but requires such measures to be based on appropriate assessment of risks and applied *only* to the extent necessary to protect life or health. It requires consistency in the application of appropriate levels of SPS protection and that measures are not more trade-restrictive than necessary. SPS measures cannot be justified in the absence of sufficient scientific evidences (except emergency or provisional measures). A member cannot unjustifiably discriminate between national and foreign, or among foreign sources of supply. To meet the requirement for science-based measures, members may base their measures on either international standards or scientific risk assessment. This clearly embodies the underlying aim of the SPS Agreement, namely, to balance the sovereign right of WTO members of protecting health in their territories against SPS risks and the goal of promoting free trade and preventing protectionism.

Another key provision is harmonization. The Agreement encourages WTO members to use international standards, guidelines or recommendations where they exist and wherever possible, and it recognizes that the measures conforming to these standards are "deemed necessary" and "consistent with the SPS Agreement". The measures recognized in this way consist of "standards", "guidelines" and "recommendations" of three international standard-setting bodies: the Codex Alimentarius Commission (CAC) on food safety regulations, the Office International des Epizooties (OIE) on animal health measures, and the Secretariat of the International Plant Protection Convention (IPPC) on guidelines for plant health measures. Measures based on international standards, guidelines or recommendations developed by these three sister organizations are presumed to be consistent with the SPS Agreement. This principle, however, does not preclude a WTO member from applying stricter measures if there is scientific justification or if a higher level of SPS protection is required. In addition, the SPS Agreement also sets forth other important rules such as equivalence, regionalization, control, inspection and approval procedures and technical assistance.

Appendix B: Data Description

The data used in this paper are obtained from the websites of WTO, United Nations and the world bank. The paper concerns about the trade pattern of agricultural products, especially, editable vegetables. Analysis for other products can be similarly conducted.

The time range is from 1995 to 2012. $t = 1$ corresponds to 1997. Each i indexes a trade route. Due to different tariff policies and trade barriers in different countries, we give different identities for the trade route from country A to country B and the reverse route. We tried our best to fill in the covariates values (i.e., values of z_{it} and x_{it}). If some key covariates values (such as GDP and population) for a trade route cannot be found, we delete it because the pattern for such trade routes would be very different given that even the GDPs of the trade partners are not observable. As a result, all z_{it}'s are observable in our data but the trade volume y_{it} may not. y_{it} is measured in 2005 US dollars if it is observable.

We list below the covariates z_{it} and z_i in the dynamic participation process and x_{it} and x_i in the dynamic trading process:

z_{it}: STC (dummy for stc), FTA (dummy for fta), POP (sum of log populations of importer and exporter), GDP (sum of log GDPs of importer and exporter)

z_i: $1, \overline{z}_i$ (mean of z_{it} from $t = 1$ to T), DIST (log distance between the capitals of the importer and exporter), LANG (dummy for common language), REL (dummy for common religion), COL (dummy for previous colony), CONT (dummy for contiguity), CM (indicator for importer country), CX (indicator for exporter country)

x_{it}: same as z_{it}

x_i: $1, \overline{z}_i$, DIST, CM and CX

Appendix C: GEE Estimation

We review the GEE method of Liang and Zeger (1986) in this appendix, with emphasis on the two cases used in this paper, i.e., balanced binary outcome (Case I) and unbalanced continuous outcome (Case II). We roughly follow the notations of Liang and Zeger (1986) with some adaptions to our context. Let $Y_i = (y_{i1}, \cdots, y_{iT_i})'$ be the $T_i \times 1$ vector of outcome values and let $W_i = (w_{i1}, \cdots, w_{iT_i})'$ be the $T_i \times p$ matrix of convariate values for the ith subject, $i = 1, \cdots, n$. In Case I, $y_{it} = d_{it}$, $T_i = T$. $w_{it} = (d_{i,t-1}, \mathbf{z}_{it}', D_t, d_{i0})'$ in Case I and $w_{it} = (y_{i,t-1}, x_{it}', D_t, x_i', y_{i0})'$ in Case II. Also, p in Case I is greater than that in Case II. We assume that the marginal density for y_{it} may be written in exponential family notation as

$$f(y_{it}) = \exp\left\{[y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})]\,\phi\right\},$$

where $\theta_{it} = h(\eta_{it})$, $\eta_{it} = w_{it}'\beta$. In Case I, $h(\eta) = \ln\frac{\Phi(\eta)}{1-\Phi(\eta)}$, $a(\theta) = -\ln\frac{1}{1+e^{\theta}}$, $b(y) = 1$ and $\phi = 1$; in Case II, $h(\eta) = \eta$, $a(\theta) = \theta^2/2$, $b(y) = -y^2/2 - \sigma^2\left(\log\sigma + \log(2\pi)/2\right)$ and $\phi = 1/\sigma^2$ with $\sigma^2 = Var(\alpha_i + \varepsilon_{it})$. Under this formulation, the first two moments of y_{it} are given by

$$E[y_{it}] = a'(\theta_{it}), Var(y_{it}) = a''(\theta_{it})/\phi.$$

In Case I, $E[y_{it}] = \Phi(w_{it}'\beta)$ and $Var(y_{it}) = \Phi(w_{it}'\beta)(1-\Phi(w_{it}'\beta))$; in Case II, $E[y_{it}] = w_{it}'\beta$ and $Var(y_{it}) = \sigma^2$. Define

$$\begin{aligned}
\Delta_i &= \operatorname{diag}\left(\frac{d\theta_{it}}{d\eta_{it}}\right), \; T_i \times T_i, \\
A_i &= \operatorname{diag}\left(a''(\theta_{it})\right), \; T_i \times T_i, \\
S_i &= Y_i - a'(\theta_i), \; T_i \times 1, \\
D_i &= A_i \Delta_i W_i, \; T_i \times p, \\
V_i &= A_i^{1/2} R_i(\alpha) A_i^{1/2}, \; T_i \times T_i,
\end{aligned}$$

where $\frac{d\theta_{it}}{d\eta_{it}} = \frac{\phi(\eta_{it})}{\Phi(\eta_{it})(1-\Phi(\eta_{it}))}$ in Case I and $\frac{d\theta_{it}}{d\eta_{it}} = 1$ in Case II, and $R_i(\alpha)$ is a $T_i \times T_i$ working correlation matrix parametrizied by α; then the GEE estimator of β is the solution to

$$\sum_{i=1}^{n} D_i' V_i^{-1} S_i = 0.$$

The idea of the GEE estimation is as follows. If we are only interested in the marginal expectation of y_{it} and not interested in the correlation across time, then the GEE method is suitable since it only requires the specification of $E[y_{it}]$ to be correct and treats correlation as nuisance. In other words, the GEE estimator is consistent even if the working correlation matrix is not correct as long as $E[y_{it}]$ is specified correctly.[17]

To compute $\widehat{\beta}_{GEE}$, we use the following iteration,

$$\widehat{\beta}_{j+1} = \widehat{\beta}_j - \left\{\sum_{i=1}^{n} D_i(\widehat{\beta}_j)' \widetilde{V}_i^{-1}(\widehat{\beta}_j) D_i(\widehat{\beta}_j)\right\}^{-1} \sum_{i=1}^{n} D_i(\widehat{\beta}_j)' \widetilde{V}_i^{-1}(\widehat{\beta}_j) S_i(\widehat{\beta}_j).$$

Define $Z = D\beta - S$ with $D = (D_1', \cdots, D_n')'$ and $S = (S_1', \cdots, S_n')'$; then the iteration above is equivalent to performing an iteratively reweighted linear regression of Z on D with weight $\widetilde{V}^{-1}$, where $\widetilde{V} = \operatorname{diag}\left(\widetilde{V}_i\right)$. At a given iteration, the correlation parameter α and scale parameter ϕ can be estimated from the current Pearson residuals (or standardized errors), defined by

$$\widehat{r}_{it} = \left\{y_{it} - a'(\widehat{\theta}_{it})\right\} \Big/ \left\{a''(\widehat{\theta}_{it})\right\}^{1/2},$$

where $\widehat{\theta}_{it}$ depends on the current value for $\widehat{\beta}$. We can then estimate ϕ by

$$\widehat{\phi}^{-1} = \sum_{i=1}^{n} \sum_{t=1}^{T_i} \widehat{r}_{it}^2 \Big/ (N - p),$$

where $N = \sum_{i=1}^{n} T_i$. From Theorem 2 of Liang and Zeger (1986), the asymptotic variance of $\widehat{\beta}_{GEE}$ can be consistently estimated by

$$\left(\frac{1}{n}\sum_{i=1}^{n} \widehat{D}_i' \widehat{V}_i^{-1} \widehat{D}_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{D}_i' \widehat{V}_i^{-1} \widehat{S}_i \widehat{S}_i' \widehat{V}_i^{-1} \widehat{D}_i\right) \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{D}_i' \widehat{V}_i^{-1} \widehat{D}_i\right)^{-1},$$

where $\widehat{D}_i$, $\widehat{V}_i$ and $\widehat{S}_i$ are evaluated at the convergent parameter estimates. This estimator provides valid

[17] Another motivation for GEE is that the joint distribution for the maximum likelihood estimation is not easily available besides the multivariate Gaussian. For example, it is hard to model the joint binary distribution.

standard errors even if the correlations within group are not as hypothesized by the specified correlation structure.

In our estimation, we use the exchangeable $R_i(\alpha)$, which is given by

$$R_i(s,t) = \begin{cases} 1, & s = t, \\ \alpha, & o/w, \end{cases}$$

and α is estimated by

$$\widehat{\alpha} = \frac{\sum_{i=1}^{n}\left(\sum_{t=1}^{T_i}\sum_{s\neq t}\widehat{r}_{it}\widehat{r}_{is}\right)}{\sum_{i=1}^{n}T_i(T_i-1)} \Bigg/ \frac{\sum_{i=1}^{n}\sum_{t=1}^{T_i}\widehat{r}_{it}^2}{N}.$$

This working correlation matrix is correct in the random effect model. Other working correlation matrices can be found in Section 4 of Liang and Zeger (1986).

Appendix D: Asymptotic Variance of the NLS Estimator

Recall from the main text that the minimization problem in the NLS estimator is

$$\min_{\theta^0,\theta}\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}d_{it}\left[y_{it} - m_{it}^0(\theta^0;\widehat{\vartheta})d_{i0} - m_{it}(\theta;\widehat{\vartheta})(1-d_{i0})\right]^2,$$

where

$$\begin{aligned} m_{it}^0(\theta^0;\vartheta) &= \left(\sum_{j=0}^{t-1}\rho^j x_{i,t-j}\right)\beta + \sum_{j=0}^{t-1}\rho^j\delta_{1,t-j}D_{t-j} + \left(\rho^t + \frac{1-\rho^t}{1-\rho}\pi_1\right)y_{i0} + \frac{1-\rho^t}{1-\rho}x_i\varphi + \xi_t^0\lambda_{it}, \\ m_{it}(\theta;\vartheta) &= \left(\sum_{j=0}^{t-1}\rho^j x_{i,t-j}\right)\beta + \sum_{j=0}^{t-1}\rho^j\delta_{1,t-j}D_{t-j} + \left(\rho^t + \frac{1-\rho^t}{1-\rho}\pi_1\right)\left[\sum_{s=0}^{T_0}\underline{x}_{i,-s}\kappa_s + \underline{x}_i\varkappa\right] + \frac{1-\rho^t}{1-\rho}x_i\varphi + \xi_t\lambda_{it}, \\ \theta^0 &= \left(\rho,\beta,\delta_1,\varphi,\pi_1,\xi^0\right) \text{ with } \delta_1 = (\delta_{11},\cdots,\delta_{1T}) \text{ and } \xi^0 = (\xi_1^0,\cdots,\xi_T^0), \\ \theta &= (\rho,\beta,\delta_1,\varphi,\pi_1,\kappa,\varkappa,\xi) \text{ with } \kappa = (\kappa_0,\cdots,\kappa_{T_0}) \text{ and } \xi = (\xi_1,\cdots,\xi_T), \end{aligned}$$

and $\widehat{\vartheta}$ is an estimator of ϑ from the first step. We assume the overlapped parameters $(\rho,\beta,\delta_1,\varphi,\pi_1)$ in θ^0 and θ are the same and denote the "net" parameter as $\boldsymbol{\theta} = \left(\rho,\beta,\delta_1,\varphi,\pi_1,\kappa,\varkappa,\xi^0,\xi\right)$. Define

$$g_{it}\left(\boldsymbol{\theta};\vartheta\right) = d_{it}\left[y_{it} - m_{it}^0(\theta^0;\vartheta)d_{i0} - m_{it}(\theta;\vartheta)(1-d_{i0})\right].$$

Then following Wooldridge (2010, Section 12.3), we can write

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = -D^{-1}n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}g_{it}(\boldsymbol{\theta};\widehat{\vartheta})'g_{it}(\boldsymbol{\theta};\widehat{\vartheta}) + o_p(1),$$

where

$$D = E\left[\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}}g_{it}(\boldsymbol{\theta};\vartheta)'\nabla_{\boldsymbol{\theta}}g_{it}(\boldsymbol{\theta};\vartheta)\right]$$

can be consistently estimated by

$$\widehat{D} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\widehat{\boldsymbol{\theta}};\widehat{\vartheta})'\nabla_{\boldsymbol{\theta}} g_{it}(\widehat{\boldsymbol{\theta}};\widehat{\vartheta}).$$

The mean-value expansion of $n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\boldsymbol{\theta};\widehat{\vartheta})' g_{it}(\boldsymbol{\theta};\widehat{\vartheta})$ around ϑ gives

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\boldsymbol{\theta};\widehat{\vartheta})' g_{it}(\boldsymbol{\theta};\widehat{\vartheta}) \\ &= n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\boldsymbol{\theta};\vartheta)' g_{it}(\boldsymbol{\theta};\vartheta) + Q\sqrt{n}\left(\widehat{\vartheta}-\vartheta\right) + o_p(1), \end{aligned}$$

where

$$Q = E\left[\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\boldsymbol{\theta};\vartheta)'\nabla_{\vartheta} g_{it}(\boldsymbol{\theta};\vartheta)\right]$$

can be consistently estimated by

$$\widehat{Q} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\widehat{\boldsymbol{\theta}};\widehat{\vartheta})'\nabla_{\vartheta} g_{it}(\widehat{\boldsymbol{\theta}};\widehat{\vartheta}).$$

From Appendix C, we can get

$$\sqrt{n}\left(\widehat{\vartheta}-\vartheta\right) = n^{-1/2}\sum_{i=1}^{n} d_i(\vartheta) + o_p(1),$$

where $d_i = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{D}_i'\widehat{V}_i^{-1}\widehat{D}_i\right)^{-1}\widehat{D}_i'\widehat{V}_i^{-1}\widehat{S}_i$. Combining the results above, we have

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) = -D^{-1}n^{-1/2}\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\boldsymbol{\theta};\vartheta)' g_{it}(\boldsymbol{\theta};\vartheta) + Q d_i(\vartheta)\right] + o_p(1),$$

so by the central limit theorem,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \xrightarrow{d} N\left(0, D^{-1}\Omega D^{-1}\right),$$

where

$$\Omega = E\left[r_i r_i'\right] \text{ with } r_i = \sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\boldsymbol{\theta};\vartheta)' g_{it}(\boldsymbol{\theta};\vartheta) + Q d_i(\vartheta)$$

can be consistently estimated by

$$\widehat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\widehat{r}_i\widehat{r}_i' \text{ with } \widehat{r}_i = \sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} g_{it}(\widehat{\boldsymbol{\theta}};\widehat{\vartheta})' g_{it}(\widehat{\boldsymbol{\theta}};\widehat{\vartheta}) + \widehat{Q} d_i(\widehat{\vartheta}).$$

We now derive the explicit formula for $\nabla_{\boldsymbol{\theta}} g_{it}(\boldsymbol{\theta};\vartheta)$ and $\nabla_{\vartheta} g_{it}(\boldsymbol{\theta};\vartheta)$:

$$\begin{aligned}
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\rho} &= -d_{it}\left\{\left(\sum_{j=1}^{t-1} j\rho^{j-1}x_{i,t-j}\right)\beta + \frac{(1-\rho^t)-t\rho^{t-1}(1-\rho)}{(1-\rho)^2}x_i\varphi\right.\\
&\quad\left.+\left(t\rho^{t-1}+\frac{(1-\rho^t)-t\rho^{t-1}(1-\rho)}{(1-\rho)^2}\pi_1\right)\left[y_{i0}d_{i0}+\left(\sum_{s=0}^{T_0}\underline{x}_{i,-s}\kappa_s+\underline{x}_i\varkappa\right)(1-d_{i0})\right]\right\},\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\beta} &= -d_{it}\sum_{j=0}^{t-1}\rho^j x_{i,t-j},\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\varphi} &= -d_{it}\frac{1-\rho^t}{1-\rho}x_i,\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\delta_{1s}} &= -d_{it}\rho^{t-s}D_s 1(1\le s\le t), s=1,\cdots,T,\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\pi_1} &= -d_{it}\frac{1-\rho^t}{1-\rho}\left[y_{i0}d_{i0}+\left(\sum_{s=0}^{T_0}\underline{x}_{i,-s}\kappa_s+\underline{x}_i\varkappa\right)(1-d_{i0})\right],\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\kappa_s} &= -d_{it}(1-d_{i0})\left(\rho^t+\frac{1-\rho^t}{1-\rho}\pi_1\right)\underline{x}_{i,-s},\ s=0,\cdots,T_0,\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\varkappa} &= -d_{it}(1-d_{i0})\left(\rho^t+\frac{1-\rho^t}{1-\rho}\pi_1\right)\underline{x}_i,\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\xi_s^0} &= -d_{it}d_{i0}\lambda_{is}1(s=t), s,t=1,\cdots,T,\\
\frac{\partial g_{it}(\boldsymbol{\theta};\vartheta)}{\partial\xi_s} &= -d_{it}(1-d_{i0})\lambda_{is}1(s=t), s,t=1,\cdots,T,
\end{aligned}$$

and

$$\begin{aligned}
\nabla_{\vartheta} g_{it}(\boldsymbol{\theta};\vartheta) &= -d_{it}\left[d_{i0}\xi_t^0+(1-d_{i0})\,\xi_t\right]\frac{\partial\lambda_{it}}{\partial\vartheta}\\
&= d_{it}\left[d_{i0}\xi_t^0+(1-d_{i0})\,\xi_t\right]\lambda\left(Z_{it}\vartheta\right)\left(\lambda\left(Z_{it}\vartheta\right)+Z_{it}\vartheta\right)Z_{it}.
\end{aligned}$$

where $Z_{it}=(d_{i,t-1},\mathbf{z}_{it},D_t,d_{i0})$.