Gravity models of trade: unobserved heterogeneity and endogeneity*

Felix Chan†, Mark N. Harris†, William Greene#, and László Kónya♮
†School of Economics and Finance, Curtin University, Australia
#Economics Department, Stern School of Business, New York University, USA
♮School of Economics, La Trobe University, Australia

March, 2014

Abstract

We consider the estimation of the usual Gravity model of trade, which involves flows of trade, say exports, from country i to country j in time period t. We suggest an easy-to-impliment *generalised method of moments* estimator that avoids the issues associated with the usual *fixed effects* treatment of the unobserved heterogeneity in this type of models and at the same time provides consistent parameter estimates in the face of potentially endogenous covariates.

JEL Classification: C33, F14, F15

Keywords: Gravity model, unobserved heterogeneity, Generealised Method of Moments.

*Corresponding author: Mark Harris, Mark.Harris@Curtin.edu.au. The usual caveats apply.

1 Introduction

With the increasing integration of national economies, societies, cultures and ideas, the current phase of globalization has seen international trade at unprecedented levels: in spite of the recent global recession and debt crisis, world trade of goods and services amounted to about 27 to 31 percent of world GDP every year between 2007 and 2012. Given the sheer size of international trade and the important role it has in improving productivity and efficiency by providing access to enlarged world markets, stimulating stronger competition, and generating technological spillover, specialization and division of labour, understanding the main empirical drivers of international bilateral trade is clearly a key issue for policymakers. This is true both at the national level, and also at the international level, for groups and trading blocs of countries, such as APEC and the European Union.

There are several approaches formally modelling trade behaviour, like for example partial equilibrium models if the primary interest is in the effects of policy on a specific sector, or computable general equilibrium models if interest is more economy wide and the focus is on the relationship between production, consumption, goods and factors of production. Still, since the pioneering studies of Tinbergen (1962) and Pyhnen (1963) the so-called *gravity model* has proved to be the workhorse of empirical models of bilateral trade flows. Originally developed from Isaac Newton's Law of Universal Gravitation, this model embodies the somewhat vague notion that the strength of interaction between two units is primarily determined by their sizes and the distance or friction (in physical or non-physical sense) between them. In particular, in its simplest form the gravity model of international trade posits that export (import) flows are positively related to the "masses" of the two bodies (i.e. to the "masses" of the trading partners as proxied, for example, by their GDP and population), and inversely related to the distance between them.[1]

The popularity of the basic gravity model of international trade and of its various extensions is mainly due to their empirical success in capturing the impact of such factors as common language, religion, border, trade liberalization etc. on trade, and in assessing the

[1]For a useful summary see Anderson (2011). Apart from bilateral trade flows the gravity model has been successfully adapted to a wide range of research topics ranging from Reilly's law of retail gravitation (Reilly, William J., 1931, The Laws of Retail Gravitation, New York, Knickerbocker Press) through cross-border equity flows (Magee 2008) and productivity flows (Anderson 2009) to the movement of people and ideas between places (see e.g. Karemera et al., 2000, A gravity model analysis of international migration to North America, Applied Economics, 32, 1745-1755).

effect of geographic regions and international agreements on trade. Moreover, by considering many more goods than production factors and allowing for complete specialization in different product varieties across countries, several different theoretical models, like e.g. the monopolistic competition model and the Heckscher-Ohlin-Samuelson model with a continuum of goods, are consistent with the gravity model of international trade. Nevertheless, in spite of its versatility and commendable empirical performance, the gravity model has long been criticised by theorists for the lack of a solid theoretical foundation. In fact, the simple analogy with Newtons law, which makes it so attractive and easily adaptable to many different issues and conditions, also makes it void of serious economic theory, but, at the same time, consistent with several potentially contradicting theories.

Recent advances in this field have focused on the underlying economic theory (e.g. Anderson and Wincoop (2003)), on the empirical specification (e.g. Rose (2004), Liu (2009), Subramanian and Wei (2007)), and on various econometric issues (e.g. Egger (2000), Anderson and Wincoop (2003), Silva and Tenreyro (2006), Baier and Bergstrand (2007)). The current paper belongs to this latter branch of the literature. In particular, it focuses on the treatment of the unobserved heterogeneity that has been the topic of many recent papers (such as Wall and Cheng (1999), Egger (2000), Egger (2002), Egger and Pfaffermayr (2003), Baier and Bergstrand (2007)). We suggest a parsimonious and consistent, in the presence of endogeneity, estimation procedure based on a generalised method of moments approach. We illustrate our technique with an application to a model of export flows within the OECD group of countries.

2 Gravity Models with Multilateral Resistance

Although already the 1960s saw some early attempts to create a theoretical underpinning of the gravity model, the first sound micro-economic foundations are due to Anderson (1979) and Bergstrand (1985). They were subsequently extended by Anderson and van Wincoop (2003), which introduced price indices termed as multilateral resistance terms (MRT) to measure the trade barriers trading partners impose not only on each other but on all other countries in the world. In spite of the theoretical appeal of this extended gravity model, its empirical application was hindered for a while because by definition the MRT of any country is related to the MRTs of all other countries. To solve this problem, Anderson and Wincoop (2003) restricted trade costs to be symmetrical between trading partners and derived a system of non-linear equations that was later linearized and solved analytically by Straathof (2008). An alternative solution based on the first-order Taylor-series expansion to MRTs was proposed by Baier and Bergstrand (2009). Still another possible way to deal with this problem is to use country-time fixed effects to approximate the unobservable but potentially time-varying MRTs, along with ordered country-pair fixed effects to account for unobservable but time-constant heterogeneity (see e.g. Baltagi et al., 2003, Baldwin and Taglioni, 2006; Baier and Bergstrand (2007)). Adopting the popular log-linear specification of the gravity model, this latter empirical strategy leads to the following linear three-way error components model:

$$y_{ijt} = \mathbf{x}'_{ijt}\beta_0 + v_{ijt} \tag{1}$$

$$v_{ijt} = \alpha_{ij} + \gamma_{it} + \lambda_{jt} + u_{ijt} \tag{2}$$

for $i, j = 1, ..., N$, $i \neq j$ and $t = 1, ..., T_{ij}$ where T_{ij} denotes the number of time series observations for the (i, j) pair. y_{ijt} represents the volume of trade (typically either export or import flows) from country i to j at time t; $\mathbf{x}_{ijt}$ is the $K \times 1$ vector of structural explanatory variables (such as GDP and population), which may, or may not, vary in the complete ijt index space; β_0 is the $K \times 1$ true parameter vector, which is typically unknown and is the item of interest; N and T correspond to the number of countries and time periods, respectively; and u_{ijt} is the usual disturbance term. As indicated by the notation T_{ij}, there is no requirement for this panel to be balanced in any dimension. Importantly the α_{ij}, γ_{it} and λ_{jt} are the unobserved country-pair and country-time specific effects (such as unobserved supply, demand, and time effects), which have been consistently shown to be very important in the literature, see for example,Wall and Cheng (1999), Egger (2000), Egger (2002), Egger and Pfaffermayr (2003), Baier and Bergstrand (2007).

A seemingly straightforward way to estimate this model is by the Least Squares Dummy Variables (LSDV) estimator, whereby the unobserved individual characteristics are captured by sets of dummy variables for exporter-importer country pair and exporter-time, importer-time country-time specific effects. This brute force method, however, can be very cumbersome on many countries and time periods because of the excessively large number of dummy variables. In fact, it might be even unfeasible due to computer memory limitations and/or intrinsic restrictions of popular software packages on the number of variables and on the maximum allowable matrix size. It is not by chance that all studies that report LSDV estimation results for gravity models with MRTs and reasonable large

N and T (like e.g. Baier and Bergstrand, 2007; Subramanian and Wei, 2007; Eicher and Henn, 2011) work with 5-yearly data, effectively reducing the number of γ_{it} and λ_{jt} specific effects by 80 percent. This data reduction, however, has its own price tag. Namely, valuable information might be lost by reducing the data frequency and averaging yearly observations and there is no guarantee that the five-yearly estimation results faithfully describe all facets of the underlying data generating process.

If one is not particularly concerned with the estimation of the individual fixed effects, in certain situations these problems can be overcome by the application of some fixed effects (FE) data transformation prior to estimation.[2] However, there is not a single, relatively simple transformation for unbalanced data sets that could wipe out all α_{ij}, γ_{it} and λ_{jt} two-dimensional FEs simultaneously. Alternatively, one might combine FE data transformation and the LSDV estimator (FELSDV method). For example, one might eliminate the country-pair FEs by the appropriate transformation and then apply the LSDV estimator on the transformed data to take care of the country-time FEs. However, the FE transformation has to be performed on γ_{it} and λ_{jt} as well and the resulting transfomed variables are not simple 0/1 dummy variables any more, so the computational burden and computer memory requirement might still prove to be insurmountable obstacles. A further possibility is to eliminate γ_{it} and λ_{jt} by the *difference of difference* method (Head et al., 2010). The disadvantage of this approach is that while the estimation results might be sensitive to the choice of the partner countries, there is no simple rule for the selection of the partner countries, especially not when there are many country pairs in the data set. Moreover, in case of unbalanced panel data sets it might be necessary to consider different partner countries for different time periods.

3 The Nonlinear Generalised Method of Moments Estimation

In spite of the above mentioned practical difficulties, because of the theoretically straightforward estimation and/or concerns about endogeneity, the specific effects in model (1)-(2) are invariably treated as fixed constants. However, we suggest here that this is not an ideal solution to account for unobserved heterogeneity, and that a more fruitful approach might be to treat them all as random drawings from a unknown distributions, for several

[2]They are also know as *within* or *demeaning* transformations.

reasons:

1. Standard panel data literature; "With a large number of random draws from the cross section, it almost always makes sense to treat the unobserved effects as random..." (Wooldridge 2010)[3].

2. Very few, if any, of the key variables that are of interest to both researchers and policy-makers alike, vary in the ijt index. Therefore, including a full set of dummy variables as implied by equation (1) essentially means that it is impossible to identify the effects of most of our variables of interest.

3. As the data sets grow in any dimension (N or T, but especially the former) the number of required dummy variables, and the associated loss in degrees of freedom, is enormous. This is due to the fact for every N countries, there are $N(N-1)$ possible pairs of trade relationships, so the total number of observations for a balanced panel is $N(N-1)T$. Hence, it is effectively impossible to estimate a Gravity model of the form of equation (1) for a reasonable large set of countries over an extended period of time as the computational task is beyond the capabilities of standard software packages such as Stata, Limdep, Gauss etc. running on even very large and fast desktop computers.

Clearly an issue with treating these unobserved effects as random, is the risk of potential endogeneity arising from correlations between any of α_{ij}, γ_{it} and λ_{jt} and the observed covariates (Baier and Bergstrand 2007). Moreover, although an instrumental variable approach could be considered here (Serlenga and Shin 2007), it appears that, in general, finding appropriate instruments that are both strictly exogenous to all of the unobserved elements of the model and strongly related to the observed covariates, is somewhat of a search for the Holy Grail.

However, given the stochastic structure of v_{ijt} as defined in equation (2), it is possible to derive a set of second order moment conditions to yield a consistent estimate for β_0 by applying non-linear *generalised method of moment* (GMM). It is the aim of this section is to derive such moment conditions based on the variance-covariance structure of v_{ijt} and to show that nonlinear GMM under these moment conditions is consistent and asymptotically normal. An important contribution is that the nonlinear GMM estimator

[3] p. 286.

under the proposed moment conditions is consistent in the presence of endogeneity without the need to identify other variables as instruments. In other words, this approach utilises the second order moments to eliminate the effects of endogeneity.

To understand this, it is useful to point out that in his seminal work Hansen (1982) showed (Theorem 2.1) that under certain conditions GMM is still consistent in the presence of endogeneity. This paper argues that endogeneity generally affects only one type of conditions required for consistency, namely, the population moment conditions are satisfied only at β_0. Therefore, it is possible to obtain a consistent GMM estimator in the presence of endogeneity without instrumental variables, granted that there exists a set of suitable moment conditions that can only be satisfied at β_0.

The following notations will be used for the rest of the paper. Let $\mathbf{x}_{ijt} = \left(x'_{it}, x'_{jt}, x'_{ijt}\right)'$ where x_{it} and x_{jt} are $K_1 \times 1$ and $K_2 \times 1$ vectors denoting the variables specific to countries i and j, respectively, and x_{ijt} is a $K_3 \times 1$ vector denoting the variables specific to the (i, j) pair. Note that $K = K_1 + K_2 + K_3$. Let $\{\mathbf{Y}_{ij}, \mathbf{X}_{ij}\}$ denote the time series data for y_{ijt} and $\mathbf{x}_{ijt}$ which is an outcome of a sequence of random variable $\{\mathbf{W}_{ij}\}$ such that $\{\mathbf{Y}_{ij}, \mathbf{X}_{ij}\} \subset \{\mathbf{W}_{ij}\}$. Similarly, $\{\mathbf{Y}_t, \mathbf{X}_t\}$ denotes the cross section data for the time period, t. Let $\mathbf{Y}_{i\bullet t} = (y_{i1t}, ..., y_{iNt})'$ and $\mathbf{Y}_{\bullet jt} = (y_{1jt}, ..., y_{Njt})'$ be $(N-1) \times 1$ vectors [4] and $\{\mathbf{Y}, \mathbf{X}\}$ denotes the full dataset. $\otimes$ denotes the Kronecker product, i_N denotes a $N \times 1$ vector of 1's, $\mathbf{0}_K$ denotes a $K \times 1$ vector of 0's and $\mathbf{I}_K$ denotes a $K \times K$ identity matrix. The subscript may be omitted if the size of the matrix is clear from the context of the argument. $\mathbf{D}_i$ denotes a $(N-1) \times N$ selection matrix such that $\mathbf{D}_i A = i_{N-1} \otimes A_i$ where A_i denotes the i^{th} row of the matrix A. $\mathbf{E}_i$ is the $(N-1) \times N$ elimination matrix such that $\mathbf{E}_i A$ removes the i^{th} row of A. $||A||$ denotes the Euclidean norm of A, $\xrightarrow{p}$ denotes converges in probability and $\xrightarrow{d}$ denotes converges in distribution. If x is $m \times n$ matrix then $x < y$ denotes element-wise inequality if y is a $m \times n$ matrix and if y is a scalar, then all elements in x is less than y. The same definition extends to $\leq$, and $\geq$ in a natural way.

Consider the following assumptions:

A1. For every $i = 1, ..., N$, $\mathbb{E}(\alpha_{i\bullet}\alpha'_{i\bullet}) = \sigma_\alpha \mathbf{I}_{N-1}$ for all $i = 1, ..., N$.

[4] y_{iit} does not exist.

A2. Let $\gamma_t = (\gamma_{1t}, ..., \gamma_{Nt})'$,

$$\mathbb{E}(\gamma_t \gamma_s') = \begin{cases} \sigma_\gamma^2 \mathbf{I}_N & t = s \\ \mathbf{0}_N & t \neq s. \end{cases}$$

A3. Let $\lambda_t = (\lambda_{1t}, .., \lambda_{Nt})'$,

$$\mathbb{E}(\lambda_t \lambda_s) = \begin{cases} \sigma_\lambda^2 \mathbf{I}_N & t = s \\ \mathbf{0}_N & t \neq s. \end{cases}$$

A4. For every $i = 1, ..., N$,

$$\mathbb{E}(u_{i\bullet t} u_{i\bullet s}') = \mathbb{E}(u_{\bullet it} u_{\bullet is}') \begin{cases} \sigma_u^2 \mathbf{I}_{N-1} & t = s \\ \mathbf{0}_{N-1} & t \neq s. \end{cases}$$

A5. The four unobservable components, namely, α_{ij}, γ_{it}, λ_{jt} and u_{ijt} are independent from each other for all (i, j) pairs and for all $t = 1, ..., T$.

A6. For each (i, j) pair $\{\mathbf{W}_{ij}\}_{t=1}^T$ is stationary and ergodic. Moreover $\mathbb{E}\left(||\mathbf{X} \otimes \mathbf{X}||^2\right) < \infty$.

A7. For each (i, j) pair $\mathbb{E}(u_{ijt}|\mathbf{x}_{ijt}) = 0$, $\mathbb{E}(\alpha_{ij}|\mathbf{x}_{ijt}) = 0$, $\mathbb{E}(x_{it}\gamma_{it}) = \gamma < \infty$, $\mathbb{E}(x_{jt}\lambda_{jt}) = \lambda < \infty$ and $\mathbb{E}(\gamma_{it}|x_{jt}) = \mathbb{E}(\lambda_{jt}|x_{it}) = 0$ for $t = 1, ..., T$ and $i \neq j$.

Remark 1 *Assumptions 1 and 5 provide the variance-covariance structure of v_{ijt} which will be used to derive all the second order moment conditions. Assumptions 6-7 are required to show the consistency of the GMM estimator in the presence of endogeneity. The stationarity assumption may seem restrictive but it is speculated that stationarity can be replaced by other mixing conditions (see Davidson (2002)).*

Remark 2 *Assumption 7 specified the correlation structure between $\mathbf{x}_{ijt}$ and v_{ijt}. The moment conditions presented below will obviously change, if different correlation structure is assumed. However, the proof of Proposition 1 demonstrates that it is possible to derive a set of valid moment conditions subject to the correlation structure between $\mathbf{x}_{ijt}$ and v_{ijt}. An interesting question would be how general can one allow this correlation structure before it is no longer possible to eliminate the problem of endogeneity using second moment information. This is beyond the scope of current paper and will be left for further research.*

Let β be the $K \times 1$ parameter vector and Θ be a compact subset of $\mathbb{R}^K$ such that $\beta_0 \in \Theta$. Let $g(\beta)$ be the vector of M valid sample moment conditions, such that $\mathbb{E}\left[g(\beta; \mathbf{Y}, \mathbf{X})\right] =$

$\mathbf{0}_M$ with $M \geq K$. The GMM estimator is defined to be

$$\hat{\beta} = \arg\min_{\beta \in \Theta} g'(\beta; \mathbf{Y}, \mathbf{X}) \mathbf{\Sigma}^{-1} g(\beta; \mathbf{Y}, \mathbf{X}) \tag{3}$$

where $\mathbf{\Sigma}^{-1}$ is the optimal weight matrix. Following the standard approach in the literature, define β^* as the solution to the optimisation in equation (3) with $\mathbf{\Sigma} = \mathbf{I}_M$, then the optimal weight matrix can be estimated by

$$\hat{\mathbf{\Sigma}}(\beta^*) = T^{-1} \sum_{s=1}^{T} g(\beta^*; \mathbf{Y}_s, \mathbf{X}_s) g'(\beta^*; \mathbf{Y}_s, \mathbf{X}_s). \tag{4}$$

Hence, an efficient GMM estimator can be obtained as

$$\hat{\beta}_{\text{2-step}} = \arg\min_{\beta \in \Theta} g'(\beta; \mathbf{Y}, \mathbf{X}) \hat{\mathbf{\Sigma}}^{-1}(\beta^*) g(\beta; \mathbf{Y}, \mathbf{X}). \tag{5}$$

Under Assumptions 1-7, it is straightforward to show that the following hold:

C1. For all $i, j = 1, ..., N$, $i \neq j$, $\mathbb{E}\left(v_{i \bullet t} v'_{i \bullet t} - v_{j \bullet t} v'_{j \bullet t}\right) = \mathbf{0}_{(N-1)\times(N-1)}$, $t = 1, ..., T$.

C2. For all $i, j = 1, ..., N$, $i \neq j$, $\mathbb{E}\left(v_{\bullet it} v'_{\bullet it} - v_{\bullet jt} v'_{\bullet jt}\right) = \mathbf{0}_{(N-1)\times(N-1)}$, $t = 1, ..., T$.

C3. For all $i, j = 1, ..., N$, $i \neq j$, $\mathbb{E}\left(v_{i \bullet t} v'_{i \bullet t-1} - v_{j \bullet t} v'_{j \bullet t-1}\right) = \mathbf{0}_{(N-1)\times(N-1)}$, $t = 2, ..., T$.

C4. For all $i, j = 1, ..., N$, $i \neq j$, $\mathbb{E}\left(v_{i \bullet t} v'_{i \bullet t} - v_{j \bullet t-1} v'_{j \bullet t-1}\right) = \mathbf{0}_{(N-1)\times(N-1)}$, $t = 2, ..., T$.

C5. For all $i, j = 1, ..., N$, $\mathbb{E}\left(v_{i \bullet t} v'_{j \bullet t-1}\right) = \mathbf{0}_{(N-1)\times(N-1)}$, for $t = 2, ..., T$.

C6. For all $i, j = 1, ..., N$, $\mathbb{E}\left(v'_{\bullet it} v_{\bullet jt-1}\right) = \mathbf{0}_{(N-1)\times(N-1)}$, for $t = 2, ..., T$.

C1-C6 lead to a maximum of $\frac{5}{2} T N^2 (N-1)$ moment conditions. Their validity is implied by the following propositions:

Proposition 1 *If $\{\mathbf{Y}_{ij}, \mathbf{X}_{ij}\}$ follows the gravity model as defined in equations (1) - (2), then under Assumptions 1 -7, Assumptions 2.1-2.5 in Hansen (1982) and the moment conditions C1-C6, the GMM estimator as defined in equation (3) is consistent, that is $\hat{\beta} \xrightarrow{p} \beta_0$ as $T \to \infty$.*

Proof. See Appendix ■

Proposition 2 *Under the Assumptions in Proposition 1 and Assumptions 3.5 and 3.6 in Hansen (1982), $\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$ where*

$$V = \mathbb{E}\left[\frac{\partial g'(\beta)}{\partial \beta} \left(g(\beta) g'(\beta)\right)^{-1} \frac{\partial g(\beta)}{\partial \beta'}\right]^{-1} \Bigg|_{\beta = \hat{\beta}}. \tag{6}$$

Proof. See Appendix. ■

Given the results in Propositions 1 and 2, the test of over-identifying moment conditions can be conducted in the usual way. That is,

$$J = \sqrt{T}g'(\hat{\beta}; \mathbf{Y}, \mathbf{X})\mathbf{\Sigma}^{-1}g(\hat{\beta}; \mathbf{Y}, \mathbf{X}) \xrightarrow{d} \chi^2(M - K). \tag{7}$$

4 Empirical Application

For the sake of illustration we consider an unbalanced dataset of 1,056 trading partners and 33,514 observations of the current OECD countries over the years 1960-2005. The data was primarily sourced from: the IMF's *Direction of Trade Statistics*; the IMF's *International Financial Statistics*; the World Bank's *World Development Indicators*; the World Trade Organization's website (www.wto.org); and the CIA's *World Factbook*. [5]

The dependent variable is the (log of) real export flows. The explanatory variables are:

- $ONEIN$: dummy variable for one of the two countries being in GATT/WTO.
- $BOTHIN$: dummy variable for both countries being in GATT/WTO.
- $LNRGDP_{it}$, $LNRGDP_{jt}$: (log of) real GDP.
- $LNRGDPPOP_{it}$, $LNRGDPPOP_{jt}$: (log of) real per capita GDP.
- $LNDIST_{ij}$: (log of) distance between the trading partners.
- $LNLAND_i$, $LNLAND_j$: (log of) land area of country.
- $CLANG_{ij}$: countries share a common language.
- $CBORD_{ij}$: countries share a common land border.
- $LLOCK_i, LLOCK_i$: dummy variables for a landlocked country.
- $ISLAND_i$, $ISLAND_j$: dummy variables for an island nation.
- $EVCOL_{ij}$: dummy for whether country i ever colonised country j.

[5] This is a subset of the data used in Kónya *et al.* (2011). Further information about the variables and data sources can be found there.

- $COMCOL_{ij}$: dummy variable for a common coloniser.
- $MUNI_{ijt}$: dummy variable for monetary union membership.
- TA_{ijt}: dummy variable for a bilateral or regional trade agreement.

Given the definitions of the dependent and independent variables, the coefficients of $LNRGDP_{it}$, $LNRGDP_{jt}$, $LNRGDPPOP_{it}$ and $LNRGDPPOP_{jt}$ are elasticities, while the remaining coefficients are semi-elasticities. As regards their expected signs, if GATT/WTO membership has a positive effect on trade, one would expect $ONEIN_{ijt}$ and $BOTHIN_{ijt}$ to have positive coefficients and the coefficient of $BOTHIN_{ijt}$ to dominate that of $ONEIN_{ijt}$. However, if these variables represent trade diversion and trade creation effects of GATT/WTO membership, then one would expect a negative coeffient for $ONEIN_{ijt}$ and a positive coefficient for $BOTHIN_{ijt}$ (although the former one could still be positive due to the externalities of GATT/WTO on non-member countries). Moreover, one would expect richer countries; countries that share a language, land border, or colonial history; countries that belong to the same trading group, monetary union, or have some special bilateral agreement; all to trade more with each other. On the contrary, countries that are far apart, or geographically larger are likely to trade less. Finally, whether being landlocked or an island nation encourages trade or not seems to be ambiguous.

Let y_{ijt} denote the log of real export flow from country i to country j at time t and $\mathbf{x}_{ijt}$ be the vector of the corresponding explanatory variables. Given the model defined in equations (1) and (2) and $\nu(\beta) = \mathbf{Y} - \mathbf{X}\beta$, the sample counterparts of the moment restrictions as stated in equations (1)-(6) are:

$$
\begin{aligned}
g_{1t}(\beta) =& \text{vech}\left[\nu_{i\bullet t}(\beta)\nu'_{i\bullet t}(\beta) - N^{-1}\sum_{j=1}^{N}\nu_{j\bullet t}(\beta)\nu'_{j\bullet t}(\beta)\right] \quad i = 1,\dots,N \\
g_{2t}(\beta) =& \text{vech}\left[\nu_{\bullet it}(\beta)\nu'_{\bullet it}(\beta) - N^{-1}\sum_{j=1}^{N}\nu_{\bullet jt}(\beta)\nu'_{\bullet jt}(\beta)\right] \quad i = 1,\dots,N \\
g_{3t}(\beta) =& \text{vech}\left[\nu_{i\bullet t}(\beta)\nu'_{i\bullet t-1}(\beta) - N^{-1}\sum_{j=1}^{N}\nu_{j\bullet t}(\beta)\nu'_{j\bullet t-1}(\beta)\right] \quad i = 1,\dots,N \\
g_{4t}(\beta) =& \text{vech}\left[\nu_{\bullet it}(\beta)\nu'_{\bullet it-1}(\beta) - N^{-1}\sum_{j=1}^{N}\nu_{\bullet jt}(\beta)\nu'_{\bullet jt-1}(\beta)\right] \quad i = 1,\dots,N \\
g_{5t}(\beta) =& \text{vech}\left[N^{-1}\sum_{j=1}^{N}\nu_{i\bullet t}(\beta)\nu'_{j\bullet t-1}(\beta)\right] \quad i = 1,\dots,N \\
g_{6t}(\beta) =& \text{vech}\left[N^{-1}\sum_{j=1}^{N}\nu_{\bullet it}(\beta)\nu'_{\bullet jt-1}(\beta)\right] \quad i = 1,\dots,N
\end{aligned}
$$

The parameter vector β can be estimated by the non-linear GMM estimator as defined in equation (5) with $g(\beta) = (T-1)^{-1}\sum_{t=2}^{T} g_t(\beta)$ where $g_t(\beta) = (g'_{1t}(\beta),\dots,g'_{6t}(\beta))'$. Following the standard approach and introducing $\beta^* = \arg\max_{\beta} g'(\beta)g(\beta)$, the optimal weight matrix is constructed as

$$
\hat{\mathbf{\Sigma}}(\beta^*) = (T-1)^{-1}\sum_{t=2}^{T} g_t(\beta^*)g_t(\beta^*)'. \tag{8}
$$

The nonlinear GMM estimator, is then the solution to the optimisation problem: $\hat{\beta} = \arg\min_{\beta} g'(\beta)\mathbf{\Sigma}^{-1}(\beta^*)g(\beta)$. Moreover, the variance-covariance matrix of $\hat{\beta}$ can be estimated by

$$
\hat{V} = T^{-1}\sum_{t=2}^{T}\left.\frac{\partial g'_t(\beta)}{\partial\beta}\right|_{\beta=\hat{\beta}}\left[\sum_{t=2}^{T} g_t(\hat{\beta})g'_t(\hat{\beta})\right]^{-1}\sum_{t=2}^{T}\left.\frac{\partial g_t(\beta)}{\partial\beta'}\right|_{\beta=\hat{\beta}}.
$$

Note that for any $K\times K$ matrix, $\mathbf{A}$, vech$\mathbf{A} = \mathbf{S}$vec$\mathbf{A}$ where $\mathbf{S}$ is an $K(K+1)/2\times K^2$ selection matrix that will select the appropriate elements in vec $\mathbf{A}$. This implies

$$
\frac{\partial g(\beta)}{\partial\beta'} = (T-1)^{-1}\sum_{t=2}^{T}\begin{pmatrix}\mathbf{S}\dfrac{\partial\text{vec}g_{1t}(\beta)}{\partial\beta'} \\ \vdots \\ \mathbf{S}\dfrac{\partial\text{vec}g_{6t}(\beta)}{\partial\beta'}\end{pmatrix}, \tag{9}
$$

where

$$
\begin{aligned}
\frac{\partial \mathrm{vec} g_{1t}(\beta)}{\partial \beta'} =& - \left[\nu_{i\bullet t}(\beta) \otimes \mathbf{X}_{i\bullet t} + \mathbf{X}_{i\bullet t} \otimes \nu_{i\bullet t}(\beta)\right] \\
&- N^{-1} \sum_{j=1}^{N} \left[\nu_{j\bullet t}(\beta) \otimes \mathbf{X}_{j\bullet t} + \mathbf{X}_{j\bullet t} \otimes \nu_{j\bullet t}(\beta)\right] \quad i = 1, \ldots, N \\
\frac{\partial \mathrm{vec} g_{2t}(\beta)}{\partial \beta'} =& - \left[\nu_{\bullet it}(\beta) \otimes \mathbf{X}_{\bullet it} + \mathbf{X}_{\bullet it} \otimes \nu_{\bullet it}(\beta)\right] \\
&- N^{-1} \sum_{j=1}^{N} \left[\nu_{\bullet jt}(\beta) \otimes \mathbf{X}_{\bullet jt} + \mathbf{X}_{\bullet jt} \otimes \nu_{\bullet jt}(\beta)\right] \quad i = 1, \ldots, N \\
\frac{\partial \mathrm{vec} g_{3t}(\beta)}{\partial \beta'} =& - \left[\nu_{i\bullet t-1}(\beta) \otimes \mathbf{X}_{i\bullet t} + \mathbf{X}_{i\bullet t-1} \otimes \nu_{i\bullet t}(\beta)\right] \\
&- N^{-1} \sum_{j=1}^{N} \left[\nu_{j\bullet t-1}(\beta) \otimes \mathbf{X}_{j\bullet t} + \mathbf{X}_{j\bullet t-1} \otimes \nu_{j\bullet t}(\beta)\right] \quad i = 1, \ldots, N \\
\frac{\partial \mathrm{vec} g_{4t}(\beta)}{\partial \beta'} =& - \left[\nu_{\bullet it-1}(\beta) \otimes \mathbf{X}_{\bullet it} + \mathbf{X}_{\bullet it-1} \otimes \nu_{\bullet it}(\beta)\right] \\
&- N^{-1} \sum_{j=1}^{N} \left[\nu_{\bullet jt-1}(\beta) \otimes \mathbf{X}_{\bullet jt} + \mathbf{X}_{\bullet jt-1} \otimes \nu_{\bullet jt}(\beta)\right] \quad i = 1, \ldots, N \\
\frac{\partial \mathrm{vec} g_{5t}(\beta)}{\partial \beta'} =& - N^{-1} \sum_{j=1}^{N} \left[\nu_{j\bullet t-1}(\beta) \otimes \mathbf{X}_{i\bullet t} + \mathbf{X}_{j\bullet t-1} \otimes \nu_{i\bullet t}\right] \qquad i = 1, \ldots, N \\
\frac{\partial \mathrm{vec} g_{6t}(\beta)}{\partial \beta'} =& - N^{-1} \sum_{j=1}^{N} \left[\nu_{\bullet jt-1}(\beta) \otimes \mathbf{X}_{\bullet it} + \mathbf{X}_{\bullet jt-1} \otimes \nu_{\bullet it}\right] \qquad i = 1, \ldots, N.
\end{aligned}
$$

Table 1 presents two sets of estimation results. The first column labelled "Identity" contains the results for the case $\mathbf{\Sigma} = \mathbf{I}$ and the second column labelled "Optimal" contains the results from the two-step estimator. That is, $\mathbf{\Sigma} = \hat{\mathbf{\Sigma}}(\beta^*)$ where $\hat{\mathbf{\Sigma}}(\beta^*)$ is estimated optimal weight matrix as defined in equation (8). It is based on the parameter estimates from the first case where $\mathbf{\Sigma} = \mathbf{I}$.

The J test based on the estimated optimal weight matrix as defined in equation (7) gives 9.081 which does not provide sufficient evidence to suggest over-identifying restrictions. Considering the 'Identity' column, the slope coefficients are all strongly significant and have logical signs, expect those of $LNLAND_i$, $LNLAND_j$, $CBORD_{ij}$ and $EVCOL_{ij}$. The coefficients of the two most important independent variables, $ONEIN_{ijt}$ and $BOTHIN_{ijt}$, imply that bilateral trade between a GATT/WTO member country and a non-member country is expected to be about 77 percent and between two GATT/WTO members about 161 percent more than between two non-members. The results are consistent with the "Optimal" case and the two sets of results are not qualitatively different from each other. This is expected as the "Optimal" results should be more efficient than the "Identity" results but they are not expected to be significantly different in terms of the parameter estimates.

5 Conclusion

Gravity models with multiple indices (here i, j and t) are extremely popular in international trade models as they adequately account for the likely presence of several sets of unobserved heterogeneity and tend to fit the data very well. This is often undertaken using a *fixed-effects* approach; essentially including thousands of dummy variables in the regression model. Alternatively, one can adopt a *random-effects* approach, though it yields inconsistent estimates if some explanatory variables are correlated with the unobserved heterogeneity, which is often the case. As a possibly remedy, here we propose the use of a nonlinear GMM estimator based on moment conditions implied by the usual assumptions of the empirical gravity model in a panel data setting. This approach avoids well-known problems of the fixed effects approach, such as the huge loss of degrees of freedom, identification of the effects of all covariates, estimation on large data sets etc. Moreover, it provides consistent parameter estimates in the face of potential endogeneity, so long as the moment conditions used are valid.

Table 1: Estimation Results

	Identity	Optimal
$ONEIN_{ijt}$	0.569	0.632
	(6.979)	(7.336)
$BOTHIN_{ijt}$	0.958	1.038
	(12.220)	(12.711)
$LNRGDP_{it}$	0.038	0.090
	(1.872)	(1.831)
$LNRGDP_{jt}$	0.710	0.740
	(363.035)	(363.122)
$LNRGDPOP_{it}$	0.822	0.863
	(24.839)	(25.403)
$LNRGDPOP_{jt}$	0.210	0.144
	(67.373)	(67.705)
$LNDIST_{ij}$	-0.852	-0.834
	(-173.368)	(-174.025)
$LNLAND_{i}$	1.073	1.156
	(61.547)	(62.348)
$LNLAND_{j}$	0.027	-0.005
	(17.416)	(17.557)
$CLANG_{ij}$	0.748	0.771
	(57.952)	(58.176)
$CBORD_{ij}$	-0.387	-0.289
	(-9.832)	(-9.207)
$LLOCK_{i}$	0.771	0.711
	(9.710)	(9.157)
$LLOCK_{j}$	-0.426	-0.520
	(-47.048)	(-47.364)
$ISLAND_{i}$	-0.541	-0.591
	(-6.570)	(-6.988)
$ISLAND_{j}$	-0.075	-0.169
	(-13.266)	(-14.176)
$EVCOL_{ij}$	-3.640	-3.556
	(-63.628)	(-64.032)
$COMCOL_{ij}$	1.150	1.165
	(23.303)	(24.139)
$MUNI_{ijt}$	0.906	1.004
	(40.433)	(41.021)
TA_{ijt}	0.076	0.147
	(6.889)	(7.680)
C	-25.138	-25.117
	(-88.646)	(-89.537)
σ^2_γ	36.660	44.820
	(3.797)	(9.517)
σ^2_λ	0.247	0.154
	(3.284)	(3.479)
σ^2_u	6.656	6.732
	(2.970)	(4.099)
σ^2_α	366.924	349.281
	(2.277)	(6.916)
σ^2_v	410.488	400.988
	(3.379)	(20.578)

*t-statistics are in the parentheses.

References

Anderson, J. (2011): "The Gravity Model," *Annual Review of Economics*, 3, 133–160.

Anderson, J., and E. V. Wincoop (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93, 170–92.

Anderson, J. E. (2009): "Gravity, Productivity, and the Pattern of Production and Trade," *NBER Working Paper 14642*.

Baier, S., and J. Bergstrand (2007): "Do Free Trade Agreements Actually Increase Members International Trade?," *Journal of International Economics*, 71, 72–95.

Davidson, J. (2002): *Stochastic Limit Theory*. Oxford University Press.

Egger, P. (2000): "A Note on the Proper Specification of the Gravity Model," *Economics Letters*, 66, 25–31.

——— (2002): "An Econometric View on the Estimation of Gravity Models and the Calculations of Trade Potentials," *The World Economy*, 25, 297–312.

Egger, P., and M. Pfaffermayr (2003): "The Proper Econometric Specification of the Gravity Model: A Three-Way Model with Bilateral Interaction Effects," *Empirical Economics*, 28, 571–580.

Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Liu, X. (2009): "GATT/WTO Promotes Trade Strongly: Sample Selection and Model Specification," *Review of International Economics*, 17, 428–446.

Magee, C. S. P. (2008): "New Measures of Trade Creation and Trade Diversion," *Journal of International Economics*, 75, 349–62.

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

Serlenga, L., and Y. Shin (2007): "Gravity Models of Intra-EU Trade: Application of the CCEP-HT Estimation in Heterogeneous Panels with Unobserved Common Time-Specific Factors," *Journal of Applied Econometrics*, 22(2), 361–381.

Silva, S., and S. Tenreyro (2006): "The Log of Gravity," *Review of Economics and Statistics*, 88, 151–175.

Subramanian, A., and S.-J. Wei (2007): "The WTO Promotes Trade, Strongly but Unevenly," *Journal of International Economics*, 72, 151–175.

Wall, H., and I. Cheng (1999): "Controlling for Heterogeneity in Gravity Models of Trade," Working Paper 99-010, Federal Reserve Bank of St. Louis.

Wooldridge, J. (2010): *Econometric Analysis of Cross Section and Panel Data, 2e.* MIT Press, Cambridge, Massachusetts.

Appendix

The following lemma is useful for the proof of Proposition 1.

Lemma 1 *For* $i, j = 1, ...N$, $\mathbf{D}_i\mathbf{D}_i' = \mathbf{D}_j\mathbf{D}_j'$ *and* $\mathbf{E}_i\mathbf{E}_i' = \mathbf{E}_j\mathbf{E}_j'$.

Proof. Note that $\mathbf{D}_i = (\mathbf{0}_{N-1}, \ldots, i_{N-1}, \ldots, \mathbf{0}_{N-1})$. This implies $\mathbf{D}_i\mathbf{D}_i' = i_{N-1}i_{N-1}'$ for all i. Let $\mathbf{e}_i$ be the $(N-1) \times 1$ vector such that all components except the i^{th} component are zeros and equals to 1 at the i^{th} component. $\mathbf{E}_i$ can then be expressed as $\mathbf{E}_i = (\mathbf{e}_1, \ldots, \mathbf{0}_{N-1}, \ldots, \mathbf{e}_{N-1})$ where $\mathbf{0}_{N-1}$ is the i^{th} column in $\mathbf{E}_i$. Direct calculations then gives $\mathbf{E}_i\mathbf{E}_i' = \sum_{k=1}^{N-1} \mathbf{e}_k\mathbf{e}_k' = \mathbf{I}_{N-1}$, for all i. This completes the proof. ■

Proof of Proposition 1. The proposition is an application of Theorem 2.1 in Hansen (1982), so it is sufficient to verify its assumptions under C1 - C6 and Assumptions 1-7.

It is straightforward to show that Assumptions 1-7 and C1-C6 satisfy Assumptions 2.1-2.5 in Hansen (1982). Moreover, the linearity of the model as defined in equation (1) and the distributional assumptions on v_{ijt} and its components ensure that the sample moment conditions as implied by C1-C6 are first moment continuous for all $\beta \in \Theta$ where Θ is a compact subset of $\mathbb{R}^K$. Thus, it is sufficient to verify that $\mathbb{E}\left[g(\beta; \mathbf{Y}, \mathbf{X})\right] = 0$ if and only if $\beta = \beta_0$.

To show this, rewrite the gravity model as defined in equations (1) and (2) in terms of the observed data for given i and t, this gives:

$$\mathbf{Y}_{i\bullet t} = \mathbf{X}_{i\bullet t}\beta_0 + v_{i\bullet t}$$

and hence, for any $\beta \in \Theta$, define

$$\nu_{i \bullet t}(\beta) = \mathbf{X}_{i \bullet t}(\beta_0 - \beta) + v_{i \bullet t} \tag{10}$$

$$= \mathbf{X}_{i \bullet t}(\beta_0 - \beta) + \alpha_{i \bullet} + \mathbf{D}_i \gamma_t + \mathbf{E}_i \lambda_t + u_{i \bullet t}. \tag{11}$$

Similarly,

$$\nu_{\bullet it}(\beta) = \mathbf{X}_{\bullet it}(\beta_0 - \beta) + \alpha_{\bullet i} + \mathbf{E}_i \gamma_t \mathbf{D}_i \lambda_t + u_{\bullet it}. \tag{12}$$

Note that $\nu_{i \bullet t}(\beta)$ is used to denote the residual function which depends on β whereas $v_{i \bullet t}$ denotes the unobserved random variables. Under Assumptions 1-7, direct calculation gives:

$$\mathbb{E}\left[\nu_{i \bullet t}(\beta)\nu'_{i \bullet t}(\beta)\right] = \mathbb{E}\left[\mathbf{X}_{i \bullet t}(\beta_0 - \beta)(\beta_0 - \beta)' \mathbf{X}'_{i \bullet t} + v_{i \bullet t} v'_{i \bullet t} + \mathbf{X}_{i \bullet t}(\beta_0 - \beta) v'_{i \bullet t} + v_{i \bullet t}(\beta_0 - \beta)' \mathbf{X}'_{i \bullet t}\right]$$

with

$$\begin{aligned}\mathbb{E}\left(v_{i \bullet t} v'_{i \bullet t}\right) &= \mathbb{E}\left[\alpha_{i \bullet}\alpha'_{i \bullet} + \mathbf{D}_i \gamma_t \gamma'_t \mathbf{D}'_i + \mathbf{E}_i \lambda_t \lambda'_t \mathbf{E}_i + u_{i \bullet t} u'_{i \bullet t}\right] \\ &= \left(\sigma^2_\alpha + \sigma^2_\lambda + \sigma^2_u\right)\mathbf{I} + \sigma^2_\gamma i i'\end{aligned}$$

The last line follows from Lemma 1 and Assumptions 1 - 5. For C1, this implies

$$\begin{aligned}\text{vec}\mathbb{E}\left[\nu_{i \bullet t}(\beta)\nu'_{i \bullet t}(\beta) - \nu_{j \bullet t}(\beta)\nu'_{j \bullet t}(\beta)\right] &= \mathbb{E}\left[\mathbf{X}_{i \bullet t} \otimes \mathbf{X}_{i \bullet t} - \mathbf{X}_{j \bullet t} \otimes \mathbf{X}_{j \bullet t}\right] \text{vec}\left(\beta_0 - \beta\right)\left(\beta_0 - \beta\right)' \\ &\quad + \mathbb{E}\left(v_{i \bullet t} \otimes \mathbf{X}_{i \bullet t} - v_{j \bullet t} \otimes \mathbf{X}_{j \bullet t}\right) \text{vec}\left(\beta_0 - \beta\right) \\ &\quad + \mathbb{E}\left(\mathbf{X}_{i \bullet t} \otimes v_{i \bullet t} - \mathbf{X}_{j \bullet t} \otimes v_{j \bullet t}\right) \text{vec}\left(\beta_0 - \beta\right) \\ &= \mathbb{E}\left[\mathbf{X}_{i \bullet t} \otimes \mathbf{X}_{i \bullet t} - \mathbf{X}_{j \bullet t} \otimes \mathbf{X}_{j \bullet t}\right] \text{vec}\left(\beta_0 - \beta\right)\left(\beta_0 - \beta\right)'.\end{aligned}$$

The last line follows from Assumption 7 and the fact that $\mathbb{E}\left(v_{i \bullet t} v'_{i \bullet t}\right) = (\sigma^2_\alpha + \sigma^2_\lambda + \sigma^2_u)\mathbf{I} + \sigma^2_\gamma i i'$ for all $i = 1, ..., N$. Under Assumption 6, $\mathbb{E}\left(\mathbf{X}_{i \bullet t} \otimes \mathbf{X}_{i \bullet t} - \mathbf{X}_{j \bullet t} \otimes \mathbf{X}_{j \bullet t}\right)$ exists and has full rank and thus

$$\mathbb{E}\left[\mathbf{X}_{i \bullet t} \otimes \mathbf{X}_{i \bullet t} - \mathbf{X}_{j \bullet t} \otimes \mathbf{X}_{j \bullet t}\right] \text{vec}\left(\beta_0 - \beta\right)\left(\beta_0 - \beta\right)' = 0$$

if and only $\beta = \beta_0$. The same arguments apply to C2. For C3, first note that

$$\begin{aligned}\mathbb{E}\left(\nu_{i \bullet t}(\beta)\nu'_{i \bullet t}(\beta)\right) &= \mathbb{E}\left[\mathbf{X}'_{i \bullet t}\left(\beta_0 - \beta\right)\left(\beta_0 - \beta\right)' \mathbf{X}_{i \bullet t-1}\right] \\ &\quad + \mathbb{E}\left[v_{i \bullet t} v'_{i \bullet t-1} + \mathbf{X}'_{i \bullet t}\left(\beta_0 - \beta\right) v'_{i \bullet t-1}\right] \\ &\quad + \mathbb{E}\left[v_{i \bullet t}\left(\beta_0 - \beta\right)' \mathbf{X}_{i \bullet t-1}\right]\end{aligned}$$

Under Assumptions 1 - 5, it is straightforward to show that $\mathbb{E}\left(v_{i\bullet t}v'_{i\bullet t-1}\right) = \sigma_\alpha^2\mathbf{I}$ and hence

$$\text{vec}\mathbb{E}\left[v_{i\bullet t}v'_{i\bullet t-1} - v_{j\bullet t}v'_{j\bullet t-1}\right] = \mathbb{E}\left(\mathbf{X}_{i\bullet t-1}\otimes\mathbf{X}_{i\bullet t} - \mathbf{X}_{j\bullet t-1}\otimes\mathbf{X}_{j\bullet t}\right)\text{vec}\left(\beta_0-\beta\right)\left(\beta_0-\beta\right)'$$

under Assumption 6. Since $\mathbb{E}\left(\mathbf{X}_{i\bullet t-1}\otimes\mathbf{X}_{i\bullet t} - \mathbf{X}_{j\bullet t-1}\otimes\mathbf{X}_{j\bullet t}\right)$ has full rank, it implies that

$$\mathbb{E}\left(\mathbf{X}_{i\bullet t-1}\otimes\mathbf{X}_{i\bullet t} - \mathbf{X}_{j\bullet t-1}\otimes\mathbf{X}_{j\bullet t}\right)\text{vec}\left(\beta_0-\beta\right)\left(\beta_0-\beta\right)' = 0$$

if and only $\beta = \beta_0$. The same arguments apply to C4. For C5, note that

$$\begin{aligned}\mathbb{E}\left[\nu_{i\bullet t}(\beta)\nu'_{j\bullet t-1}(\beta)\right] =& \mathbb{E}\left[\mathbf{X}_{i\bullet t}\left(\beta_0-\beta\right)\left(\beta_0-\beta\right)'\mathbf{X}'_{j\bullet t-1}\right]\\ &+ \mathbb{E}\left[\mathbf{X}_{i\bullet t}\left(\beta_0-\beta\right)v'_{j\bullet t-1}\right] + \mathbb{E}\left[v_{i\bullet t}\left(\beta_0-\beta\right)'\mathbf{X}'_{j\bullet t-1}\right]\\ &+ \mathbb{E}\left(v_{i\bullet}v'_{j\bullet t-1}\right)\\ =& \mathbb{E}\left[\mathbf{X}_{i\bullet t}\left(\beta_0-\beta\right)\left(\beta_0-\beta\right)'\mathbf{X}_{j\bullet t-1}\right]\end{aligned}$$

The last line follows from the fact that all terms on the right hand side except the first identically equal to zeros under Assumptions 1 - 5 and 7. This implies:

$$\begin{aligned}\text{vec}\mathbb{E}\left[v_{i\bullet t}v'_{j\bullet t-1}\right] =& \mathbb{E}\left[\mathbf{X}_{j\bullet t-1}\otimes\mathbf{X}_{i\bullet t}\right]\text{vec}\left(\beta_0-\beta\right)\left(\beta_0-\beta\right)'\\ =& 0\end{aligned}$$

if and only if $\beta = \beta_0$. The same arguments apply to C6. This completes the proof. ■

Proof of Proposition 2. Under Assumptions 1-7 and the result as presented in Proposition 1, Proposition 2 is a straightforward application of Theorem 3.1 in Hansen (1982). This completes the proof. ■