A Panel Data Approach for Spatial and Network Selection Models

–PRELIMINARY VERSION–

Sophia Ding* and Peter Egger**

May 2, 2018

Abstract

Many data in the social sciences at large and in economics in particular feature some form of spatial or network interdependence. For instance, firms that share the same geography or the same input-output network relations have been shown to adopt certain strategies (market entry, exporting, foreign plant set-up, etc.) in a way that suggests spatial or network interdependence. Often data on these firms are incomplete (in the sample- or treatment-selection sense). However, there are no models to date that would allow researchers to account both for selection on unobservables as well as spatial or network patterns with panel-data. The present paper contributes to the literature by proposing a two-step approach towards selection on unobservables in the spirit of Heckman (1976, 1979) and Wooldridge (1995) but for panel-data with spatial or network interdependencies in both the selection and the outcome equation. Apart from outlining the econometric model and the associated estimation procedure for parameter point estimates as well as their variance-covariance matrix, the paper illustrates the suitability of the proposed approach in finite samples by way of Monte Carlo simulations. Moreover, we intend to apply the model to illustrate the relevance of self-selection of firms into exporting when analyzing the size of the exporter-wage premium in the city of Shenzhen, P.R. China.

Keywords: Spatial and network interdependence; Sample selection; Treatment selection; Panel data

JEL-codes: C23; C33; C34.

*Sophia Ding, ETH Zurich, LEE G104, Leonhardstrasse 21, 8092 Zurich, Switzerland, E-mail: ding@kof.ethz.ch.;

**Peter Egger: ETH Zurich and CEPR, LEE G101, Leonhardstrasse 21, 8092 Zurich, Switzerland, E-mail: egger@kof.ethz.ch. The authors gratefully acknowledge financial support by the Swiss National Science Foundation grant CRSII1_154446 "Inequality and Globalization: Demand versus Supply Forces and Local Outcomes".

1 Introduction

Cross-sectional interdependence through spatial or network relations among economic units or agents is a salient feature of many data in the social sciences in general and in economics in particular. For instance, firms that share the same geography, the same product market, the same factor market, or vertical links through input-output relations have been shown to adopt certain strategies (market entry, exporting, foreign plant set-up, etc.) in a way that suggests spatial or network interdependence. The latter means that economic outcomes of firms – conditional on observable characteristics that may account for part of the unconditional interdependence in outcome – feature a pattern of geographical or network interdependence. There is also evidence that individuals adapt their behaviour (learning effort, sports performance, etc.) depending on their peers or friends in a social network. Also with individuals, it appears that conditioning on their characteristics alone is not sufficient to explain the interdependence in outcome. Similarly, jurisdictions such as countries, subnational macro regions, or municipalities adopt policies (subsidization of certain firm activities, investments of specific types, implementation of particular rules, etc.) in an interdependent way. And also there, the feature of interdependence tends to prevail after conditioning on observables. There is evidence on the salience of interdependence of economic outcome both in cross sections as well as in panels of data at all levels, micro, meso, and macro.[1]

[1]For example, Conley and Ligon (2002) and Ertur and Koch (2007) find evidence that economic growth is correlated between countries, e.g. due to technological interdependencies. There is also empirical evidence for the existence of productivity spillovers on the region-level (Holtz-Eakin, 1994), the industry-state level (Audretsch and Feldmann, 1996) and on the plant-level (Martin, Mayer and Mayneris, 2011). Furthermore, spatial interdependence seems to play a role for bilateral trade flows (Behrens, Ertur, and Koch, 2012; Egger, and Pfaffermayr, 2016) as well as bilateral FDI (Blonigen, Davies, Waddell, and Naughton, 2007; Baltagi, Egger, and Pfaffermayr, 2007, 2008). Several studies focus on the spatial interdependence of (local) government policy-making regarding value-added taxes (Egger, Pfaffermayr, and Winner, 2005), corporate taxes (Devereux, Lockwood, and Redoano, 2008), and fiscal policy (Case, Hines, and Rosen, 1993). The spatial correlation of firm-level decision-making has also been studied extensively: for instance, Pinkse, Slade, and Brett (2002) examine spatial price competition among gasoline wholesalers. More recently, the role of export spillovers has been scrutinized both for the composition of the export basket of countries (Bahar, Hausmann, and Hidalgo, 2014) and the formation of importer-exporter links between firms (Kamal and Sundaram, 2016). Finally, there is a large body of literature concerned with the study of networks in social interactions, which has brought forth both theoretical (Ballester, Calvó-Armegol, and Zenou, 2006; Lee, 2007b; Blume, Bock, Durlauf, and Jayaraman, 2015) as well as applied contributions, e.g., on the role of peer effects among students in a class room (Calvó-Armengol, Patacchini, and Zenou, 2009; Cohen-Cole, Liu, and Zenou, forthcoming), for labor market participation (Ioannides, and Loury, 2004), and for criminal activities of individuals

In practice, cross-sectional interdependence poses a problem whenever the data at hand are incomplete. As incompleteness rarely occurs at random, interesting cases are the censoring (e.g., through top- or bottom-coding) or truncation of data, where selection into sample versus treatment can be distinguished. With sample selection, outcome is observed only for a non-random sample. With treatment selection, counterfactual outcome is unobserved for any (binary) treatment state. In the latter case, the average treatment effect (ATE) as a parameter on a binary treatment indicator is of interest beyond the other model parameters. However, to date there are no suitable econometric models to generically cover the case of, e.g., truncation in the presence of interdependence, in particular not with panel data.[2]

The present paper contributes to closing this gap in the literature by outlining panel-data approaches for sample and treatment selection as two related cases in the presence of cross-sectional spatial or network interdependence. Specifically, the paper proposes a two-step approach towards selection on unobservables in the spirit of Heckman (1976, 1979) for panel-data as in Wooldridge (1995) but with spatial or network interdependencies in both the selection and the outcome equation. Specifically, we adapt the selection-correction approach to the case of selection into the sample or treatment under spatial or network interdependence.

The estimation procedure of the spatial or network sample-selection model (SNSS) and the spatial or network treatment-selection model (SNTS) follows three steps: first, we estimate the selection equation using a pooled Bayesian Spatial/Network Error Probit Model to obtain consistent estimates of slope parameters and the spatial autoregressive parameter; second, we use these estimated parameters to construct a spatial/network adjustment factor and a (generalized) Inverse Mills' Ratio; finally, we add the estimated spatially/network adjusted (generalized) Inverse Mills'

(Patacchini, and Zenou, 2012).

[2]Clearly, there are cross-section models (see, e.g., Kelejian and Prucha, 1998, 1999, 2010; Lee, 2004, 2007a; and the survey article by Anselin and Bera, 1988) and panel-data models (see Kapoor, Kelejian, and Prucha, 2007; Lee, 2007b) which feature interdependence with complete data. Moreover, there are cross-section models (see Heckman, 1976, 1978; Greene, 1995) and panel-data models (see Nijman and Verbeek, 1992; Wooldridge, 1995; Vella and Verbeek, 1999; Semykina and Wooldridge, forthcoming) for the analysis of data-incompleteness in the form of truncation, mostly focusing on sample selection. Also, there are a few models to tackle data truncation in the form of sample selection with cross-section data (see McMillen, 1995; Flores-Lagunes and Schnier, 2012; Doğan and Taşpinar, 2017). However, to date there is no approach which would support the analysis of data-incompleteness in the form of truncation generically – covering both sample and treatment selection – with cross-sectional interdependence in panel-data.

Ratio in a control function in the respective outcome equation to obtain consistent estimates of the slope parameters as well as the spatial-/network-interdependence parameter using pooled Non-Linear Least Squares. Apart from point estimates, we derive an estimator of the variance-covariance matrix of the model parameters which builds on the insights of Heckman (1979) as well as Murphy and Topel (1985, 2002) but is adapted to accommodate the inherent interdependence among the cross-sectional units.

We illustrate the suitability of the proposed approach in finite samples by way of Monte Carlo simulations and compare it to other approaches which ignore selection and/or ignore cross-sectional interdependence. There are three main insights from these experiments: first, as expected, the SNSS/SNTS models outperform other ones in the presence of spatial/network interdependence. The bias of these estimators tends to increase with the magnitude of absolute deviations of the spatial-/network-interdependence parameter from zero. Second, even in the absence of interdependence, the proposed estimator is only marginally outperformed by its more efficient competitors. Third, overall, the performance of the proposed SNSS/SNTS estimators increases with the number of cross-sectional units.

We intend to apply the proposed SNTS estimator to study the exporter-wage premium on average wages paid in Chinese firms when considering the spatial clustering of exporters as well as a spatial element in wages. Specifically, doing so we will use data from the Chinese Annual Survey of Industrial Firms Database (CASIF) for the city of Shenzhen in the Peoples' Republic of China. The exporter-wage premium, i.e., the fact that exporting firms pay higher wages per worker than non-exporters on average, has been documented in many data-sets for various countries, always assuming that firms' selection into exporting was independent of other firms (and often even random), and that wages were set independently as well (see Klein, Moser, and Urban, 2013; Egger, Egger, and Kreickemeier, 2013; Egger, Egger, Kreickemeier, and Moser, 2017). Indeed, we are expecting the presence of interdependence in both participating in export markets as well as in wage payments and that ignoring them will lead to a bias in the data at hand.[3]

The remainder of this paper is organized as follows: Section 2 sets up the econometric model

[3]There is both theoretical and empirical evidence suggesting that shocks to the profitability of exporting dissipate across firms due to spatial and network interdependence (see Antràs, Fort, and, Tintelnot, 2017; Baltagi, Egger, and Kesina, 2017; Tintelnot, 2017; Chaney, 2014). Moreover, there is theoretical and empirical work suggesting that the wages paid have a spatial pattern and might therefore be correlated across firms due to local labor markets and worker flows (see Moretti, 2011).

for the spatial/network sample selection and the spatial/network treatment selection case. Section 3 outlines an estimation procedure for these models. To examine the finite sample properties of our estimators we conduct Monte Carlo experiments and report the associated design as well as results in Section 4. We then derive the analytical variance-covariance matrix of the parameters and provide an estimation strategy in Section 5. In Section 6 we give a brief overview of the empirical application that we intend to do next. Section 7 concludes.

2 Econometric Model

In this section, we outline panel-data approaches for sample selection versus treatment selection as two closely-related cases in the presence of spatial or network interdependence among the cross-sectional units. With sample selection, outcome is observed only for a non-random sample. With treatment selection, we still have truncation due to a lack of observability of counterfactual outcome for the treated and the untreated and a non-random selection into treatment. In the latter case, the average treatment effect (ATE) as a parameter on a binary treatment indicator is of additional interest. Overall, what we will discuss and consider is an approach of selection on unobservables in the spirit of Heckman (1976, 1979) and Wooldridge (1995) but for panel-data with spatial or network interdependencies along the lines of Kapoor, Kalejian, and Prucha (2007).

2.1 Some General Notation

Let us use indices $i = 1, ..., N$ and $t = 1, ..., T$ to refer to a unit, e.g. a firm or an individual, and time period, respectively. For example, we could think of N as those firms among all potential producers that actually, owed to a lucky productivity draw, produce during the period of investigation of length T. We could then consider the emergence and disappearance of firms as a random process conditional on the aforementioned observables. In any case, the number of firms present in year t, N_t, may vary with t, and so may the number of years in which firm i is observed, T_i, vary with i. We denote the number of observations in the data by $n = \sum_{t=1}^{T} N_t = \sum_{i=1}^{N} T_i$.

Models of the kind we are interested in involve two equations. We will generally use superscript $\ell = \{A, B\}$ to refer to the selection equation (A) and the outcome equations (B), respectively, of the model. Moreover, we will refer to latent dependent variables as $y_{ti}^{\ell *}$ and to their observed counterparts as y_{ti}^{ℓ}. While outcome y_{ti}^{B} will be continuous, y_{ti}^{A} will be a binary selection indicator.

We will refer to the non-stochastic, exogenous, time-variant explanatory variables in equation ℓ by x_{ti}^{ℓ}, to the parameters on them by β^{ℓ}, and to the disturbances which feature spatial or network interdependence by e_{ti}^{ℓ}. We will refer to the rank of x_{ti}^{ℓ} by k^{ℓ} and assume that x_{ti}^{A} contains x_{ti}^{B}, whereby $k^A \geq k^B$.[4]

Finally, the subsequent analysis will involve different types of disturbances: e_{ti}^{ℓ} will be ones that feature spatial or network interdependence among the cross-sectional units ($E\left[e_{ti}^{\ell} e_{tj}^{\ell}\right] \neq 0$ at least for some $i \neq j$); an error component μ_i^{ℓ} will capture time-invariant shocks which are specific to individual i but independent between i and any $j \neq i$; and an error component ε_{ti}^{ℓ} are idiosyncratic shocks which vary by t as well as i and which are independent between i and any $j \neq i$ as well as t and any $s \neq t$. We will refer to the later two error components jointly as $\xi_{ti}^{\ell} \equiv \mu_i^{\ell} + \varepsilon_{ti}^{\ell}$.

We will see that an introduction of some further notation below will help simplifying the subsequent model outline.

2.2 Selection Equation

Let the latent outcome underlying y_{ti}^{A} linearly depend on the $k^A \times 1$ vector x_{ti}^{A} through

$$\begin{aligned} y_{ti}^{A*} &= x_{ti}^{A\prime}\beta^A + e_{ti}^{A}, \quad \text{for each} \quad t = 1, ..., T; i = 1, ..., N, \\ y_{ti}^{A} &= 1[y_{ti}^{A*} > 0]. \end{aligned}$$

Consider that e_{ti}^{A} features spatial or network interdependence along the lines of Kapoor, Kalejian, and Prucha (2007). Then, we can specify

$$e_{ti}^{A} = \rho^A \sum_{j \in \mathfrak{N}_t} w_{tij} e_{tj}^{A} + \bar{x}_i^{A\prime}\delta^A + \xi_{ti}^{A} \tag{1}$$

$$\xi_{ti}^{A} \equiv \mu_i^{A} + \varepsilon_{ti}^{A}, \tag{2}$$

where $\mathfrak{N}_t$ is the set of N_t cross-sectional units which exist at time t, w_{tij} is a known scalar which parametrizes the neighborliness in some space or the network at time t between two cross-sectional units i and j, ρ^A is an unknown parameter which scales the strength of interdependence between any two cross-sectional units at time t, $\bar{x}_i^{A\prime}\delta^A$ is a parametrization of the individual-specific fixed

[4] A well-known disadvantage of the situation with $k^A = k^B$ under sample selection is that identification of the parameters relies exclusively on the functional form of the model. This problem pertains to small samples. With treatment selection, the problem is even absent (see Vella, 1998).

effect of i along the lines of Mundlak (1978), Chamberlain (1982), and Wooldridge (1995) with $\bar{x}_i^A = T^{-1}\sum_{t=1}^T x_{ti}^A$ and parameter vector δ^A.

2.3 Outcome Equation

In general terms, we assume that outcome follows a similar generic model structure as selection. However, here we must distinguish between the cases of sample selection with

$$\begin{aligned} y_{ti}^{B*} &= x_{ti}^{B\prime}\beta^B + e_{ti}^B \\ y_{ti}^B &= \begin{cases} y_{ti}^{B*} & \text{if} \quad y_{ti}^A = 1 \\ - & \text{if} \quad y_{ti}^A = 0 \end{cases} \end{aligned}$$

and treatment selection with

$$y_{ti}^B = y_{ti}^{B*} = \alpha y_{ti}^B + x_{ti}^{B\prime}\gamma + e_{ti}^B.$$

Regarding disturbances, we assume

$$e_{ti}^B = \rho^B \sum_{j\in\mathfrak{N}_t} w_{tij} e_{tj}^B + \bar{x}_i^{B\prime}\delta^B + \xi_{ti}^B, \quad (3)$$

$$\xi_{ti}^B \equiv \mu_i^B + \varepsilon_{ti}^B. \quad (4)$$

The interpretation of parameters is similar to the one on latent outcome in the selection equation so that we can suppress a detailed discussion here. However, what is important to note with regard to sample selection, is that the number of units underlying latent outcomes ($\{N_t, T_i, n\}$ for time period, individual, and the overall sample) and observed outcomes ($\{\underline{N}_t \leq N_t, \underline{T}_i \leq T_i, \underline{n} < n\}$) differ, and the aforementioned processes for e_{ti}^B and ξ_{ti}^B are generated on the full sample.

2.4 Variance-covariance Matrix for the System of Selection and Outcome Equations

Before stating the variance-covariance matrices of different concepts of disturbances in the above model, let us state some assumptions.

Assumption 1: $\{\mu_i^\ell, \varepsilon_{ti}^\ell\}$

We assume that μ_i^ℓ and ε_{ti}^ℓ are distributed independently each, whereby $E\left[\left(\mu^\ell\right)_i^2\right] = \sigma_{\mu^\ell}^2$ and

$E\left[\mu_i^\ell \mu_j^\ell\right] = 0$ for $i \neq j$ as well as $E\left[\left(\varepsilon^\ell\right)_{ti}^2\right] = \sigma_{\varepsilon^\ell}^2$ and $E\left[\varepsilon_{ti}^\ell \varepsilon_{sj}^\ell\right] = 0$ for $i \neq j$ and/or $t \neq s$. Moreover, we assume that $E\left[\mu_i^\ell \varepsilon_{tj}^\ell\right] = 0$ for any tuple $\{tij\}$.

Assumption 2: $E\left[\xi_{ti}^A \xi_{ti}^B\right]$

We assume that the elements of ξ_{ti}^B are linear in the respective elements of ξ_{ti}^A, whereby $\xi_{ti}^B = \tau\xi_{ti}^A + \nu_{ti}^B$ and ν_{ti}^B is a residual term that does not depend in any way on the selection equation and whose i-specific and ti-specific components are identically and independently distributed each. Of this linear relationship, joint normality is a special case, and we will outline the analysis for this case, here, as the derivation of truncated-variance expressions is transparent, even though most of the results would apply also under the milder assumption of ξ_{ti}^B being linear in ξ_{ti}^A (see Wooldridge, 1995, for the case without spatial/network dependence). Under joint normality and after defining $E\left[\mu_i^\ell \mu_i^\ell\right] = \sigma_{\mu^\ell}^2$ and $E\left[\mu_i^A \mu_i^B\right] = \sigma_{\mu^{AB}}^2$, $E\left[\varepsilon_{ti}^\ell \varepsilon_{ti}^\ell\right] = \sigma_{\varepsilon^\ell}^2$ and $E\left[\varepsilon_{ti}^A \varepsilon_{ti}^B\right] = \sigma_{\varepsilon^{AB}}^2$, as well as $E\left[\xi_{ti}^\ell \xi_{ti}^\ell\right] = \sigma_{\xi^\ell}^2 = \sigma_{\mu^\ell}^2 + \sigma_{\varepsilon^\ell}^2$ and $E\left[\xi_{ti}^A \xi_{ti}^B\right] = \sigma_{\xi^{AB}}^2 = \sigma_{\xi^{BA}} = \sigma_{\mu^{AB}} + \sigma_{\varepsilon^{AB}}$, we have for observation ti

$$\begin{pmatrix} \xi_{ti}^A \\ \xi_{ti}^B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\xi^A}^2 & \sigma_{\xi^{AB}} \\ \sigma_{\xi^{AB}} & \sigma_{\xi^B}^2 \end{pmatrix}\right), \tag{5}$$

and, after stacking the ξ_{ti}^ℓ first within a year into the $N_t \times 1$ vectors ξ_t^ℓ and then into the $TN \times 1$ vectors $\xi^\ell = (\xi_1^{\ell\prime}, ..., \xi_T^{\ell\prime})'$, and after defining the $T\times$T matrix of ones J_T and the size-T and size-N identity matrices I_T and I_N, respectively, we obtain

$$\begin{pmatrix} \xi^A \\ \xi^B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{\xi^A} & \Omega_{\xi^{AB}} \\ \Omega_{\xi^{BA}} & \Omega_{\xi^B} \end{pmatrix}\right), \quad \text{with}$$

$$\begin{aligned} \Omega_{\xi^A} &= (J_T \otimes \sigma_{\mu^A}^2 I_N) + \sigma_{\varepsilon^A}^2 (I_T \otimes I_N) \\ \Omega_{\xi^B} &= (J_T \otimes \sigma_{\mu^B}^2 I_N) + \sigma_{\varepsilon^B}^2 (I_T \otimes I_N), \\ \Omega_{\xi^{BA}} &= \Omega_{\xi^{AB}}' = (J_T \otimes \sigma_{\mu^{BA}} I_N) + \sigma_{\varepsilon^{BA}} (I_T \otimes I_N), \end{aligned}$$

where $\Omega_{\xi^{BA}} = \Omega_{\xi^{AB}}'$.

Let w_{tij} be a typical element of the $N_t \times N_t$ matrix W_t, I_{N_t} be an N_t-size identity matrix, ι_{N_t} be an N_t-size vector of ones, and $R_t^\ell = (r_{tij}^\ell) \equiv (I_{N_t} - \rho^\ell W_t)^{-1}$. When stacking the vector $\bar{x}_i^\ell$ into the matrix $\bar{x}^\ell = (\bar{x}_1^{\ell\prime}, ..., \bar{x}_N^{\ell\prime})'$ and stacking e_{ti}^ℓ in the same way as ξ_{ti}^ℓ, we can state the reduced forms $e_t^\ell = R_t^\ell(\bar{x}^{\ell\prime}\delta^\ell + \xi_t^\ell)$ and $e^\ell = R^\ell((\iota_T \otimes \bar{x}^{\ell\prime})\delta^\ell + \xi^\ell)$, where $R = diag_t(R_t)$.

Assumption 3: ρ^ℓ and w_{tij}

Following Kelejian and Prucha (2010) we assume that the elements w_{tij} are normalized so that either $W_t\iota_{N_t}$ is an $N_t \times 1$ vector of ones (row normalization) or the minimum of the maximum entries of the row-sum vector $W_t\iota_{N_t}$ and the column-sum vector $W_t'\iota_{N_t}$ is bounded from above by unity. Moreover, assuming that $|\rho^\ell| < 1$ is a sufficient condition for the inverses R_t^ℓ to exist and to be finite.

When using $u_{ti}^\ell \equiv \sum_{j=1}^N r_{tij}^\ell \xi_{tj}^\ell$ as well as $u_t^\ell \equiv R_t^\ell \xi_t^\ell$ and $u^\ell \equiv R^\ell \xi^\ell$, we can state the variance-covariance matrix

$$\begin{pmatrix} u^A \\ u^B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{u^A} & \Omega_{u^{AB}} \\ \Omega_{u^{AB}} & \Omega_{u^B} \end{pmatrix}\right), \quad \text{with}$$

$$\Omega_{u^\ell} = R^\ell \Omega_{\xi^\ell} R^{\ell\prime}$$

and,

$$\Omega_{u^{AB}} = R^B \Omega_{\xi^{AB}} R^{A\prime}.$$

For observation ti, the latter can be stated as

$$\begin{pmatrix} u_{ti}^A \\ u_{ti}^B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\xi^A}^2 \sum_{j=1}^N \left(r_{tij}^A\right)^2 & \sigma_{\xi^{AB}} \sum_{j=1}^N r_{tij}^B r_{tij}^A \\ \sigma_{\xi^{BA}} \sum_{j=1}^N r_{tij}^B r_{tij}^A & \sigma_{\xi^B}^2 \sum_{j=1}^N \left(r_{tij}^B\right)^2 \end{pmatrix}\right),$$

2.5 Control-function-augmented Outcome Equation

In order to correct for (sample/treatment) selection bias, we first compute the expectation of the outcome equation, conditional on being selected into the sample/into treatment. Note, that as we do not observe the latent variable y_{ti}^{A*}, the disturbances of the selection equation u_{ti}^A are unknown. Therefore, we condition on the selection indicator y_{ti}^A instead of u_{ti}^A. Since we condition on the set of covariates $\left\{x^{A0}, x^B\right\}$ this problem reduces to finding the truncated expectations of u_{ti}^B:

Spatial/Network Sample Selection

$$\begin{aligned} E[y_{ti}^B | y_{ti}^A = 1, x^{A0}, x^B] &= x_{ti}^{B\prime}\beta^B + \sum_{j=1}^N r_{tij}^B \bar{x}_j^{B\prime}\delta^B + E[u_{ti}^B | y_{ti}^A = 1, x^{A0}, x^B] \\ &= x_{ti}^{B\prime}\beta^B + \sum_{j=1}^N r_{tij}^B \bar{x}_j^{B\prime}\delta^B + \tau\psi_{ti}\lambda_{ti} \qquad (6) \end{aligned}$$

Spatial/Network Treatment Selection

$$\begin{aligned} E[y_{ti}^B|y_{ti}^A, x^{A0}, x^B] &= \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \sum_{j=1}^N r_{tij}^B \bar{x}_j^{B\prime}\delta^B + E[u_{ti}^B|y_{ti}^A, x^{A0}, x^B], \\ &= \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \sum_{j=1}^N r_{tij}^B \bar{x}_j^{B\prime}\delta^B + \tau\psi_{ti}\lambda_{ti}^g \end{aligned} \tag{7}$$

where x^{A0} are the elements of matrix $x^A = \left(x_1^{A\prime}, ..., x_N^{A\prime}\right)$ that are not contained in $x^B = \left(x_1^{B\prime}, ..., x_N^{B\prime}\right)$ and where we have defined

$$\begin{aligned} \tau &= \frac{\sigma_{\xi^{BA}}}{\sqrt{\sigma_{\xi^A}^2}} \quad , \quad \psi_{ti} = \frac{\sum_{j=1}^N r_{tij}^B r_{tij}^A}{\sum_j^N \left(r_{tij}^A\right)^2} \quad , \quad z_{ti} = \frac{x_{ti}^A\beta^A + \sum_{j=1}^N r_{tij}^A \bar{x}_j^A \delta^A}{\sqrt{\sigma_{\xi^A}^2 \sum_j^N \left(r_{tij}^A\right)^2}}, \\ \lambda_{ti} &= \frac{\phi\left(z_{ti}\right)}{\Phi\left(z_{ti}\right)}, \quad \lambda_{ti}^g = \phi\left(z_{ti}\right)\frac{y_{ti}^A - \Phi\left(z_{ti}\right)}{\Phi\left(z_{ti}\right)\left[1-\Phi\left(z_{ti}\right)\right]}. \end{aligned}$$

$\phi(\cdot)$ and $\Phi(\cdot)$ denote the Standard Normal PDF and Standard Normal CDF respectively.

We have replaced the truncated expectations of u_{ti}^B, i.e. $E[u_{ti}^B|y_{ti}^A = 1, x^{A0}, x^B]$ (sample selection) and $E[u_{ti}^B|y_{ti}^A, x^{A0}, x^B]$ (treatment selection), with the Inverse Mills' Ratio λ_{ti} and generalized Inverse Mills' Ratio λ_{ti}^g, which are multiplied by the spatial/network adjustment factor ψ_{ti} (see Appendix 1 for proofs).

When estimating the regression equation, ignoring these truncated expectations would lead to biased parameter estimates stemming from omitted variable bias. Therefore, in order to correct for sample selection (treatment selection) bias, we can include the spatially/network adjusted Inverse Mills' Ratio $\psi_{ti}\lambda_{ti}$ ($\psi_{ti}\lambda_{ti}^g$) as additional regressor in the outcome equation akin to Heckman (1979). The spatially/network adjusted Inverse Mill's Ratio has previously been used as correction function to correct for sample selection in McMillen (1995) and Flores-Lagunes & Schnier (2012) for cross-sectional data.[5]

[5]Once having conditioned on u_{ti}^A, u_{ti}^B is mean independent of x_{ti}^A and hence $\bar{x}_i^A$ (this is essentially the approach taken in McMillen (1995) and Flores-Lagunes and Schnier (2012) for the cross-sectional case).The main drawback of this approach (which essentially is a special case of the assumptions made in Wooldridge (1995)) compared to e.g. Nijman and Verbeek (1992), Vella and Verbeek (1999) is that by conditioning on only u_{ti}^A and not u_i^A, the selection cannot depend on past values of the selection indicator y_{ti}^A. At the same time, this allows for the serial correlation in u_{ti}^A to be fully unrestricted. If we wanted to include a spatial lag of y_{ti}^A in the selection equation, we would have to condition on u_t^A.

3 Estimation Strategy

The estimation of the parameters of the outcome equation follows three steps: First, we estimate the selection equation using a **Pooled Bayesian Spatial/ Network Error Probit Model** to obtain consistent estimates of parameters $\hat{\theta}_A = \{\hat{\tilde{\beta}}^A, \hat{\tilde{\delta}}^B, \hat{\rho}^A\}$, where $\tilde{\beta}^A = \frac{\beta^A}{\sigma_{\xi^A}}$ and $\tilde{\delta}^A = \frac{\delta^A}{\sigma_{\xi^A}}$. We then use these estimated parameters to construct the Spatial/Network Adjustment/Network Factor $\hat{\psi}_i$ and the Inverse Mills' Ratio $\hat{\lambda}_{ti}$ or the Generalized Inverse Mills' Ratio $\hat{\lambda}^g_{ti}$. Finally, we add $\hat{\psi}_i\hat{\lambda}_{ti}$ or $\hat{\psi}_i\hat{\lambda}^g_{ti}$ as additional regressor to the respective outcome equation. We then estimate the outcome equation using **Pooled Non-Linear Least Squares** to obtain the vector of parameters $\hat{\theta}^B = \{\hat{\beta}^B, \hat{\delta}^B, \hat{\tau}, \hat{\rho}^B\}$ (spatial/network sample selection) or $\hat{\theta}^B = \{\hat{\alpha}, \hat{\beta^B}, \hat{\delta^B}, \hat{\tau}, \hat{\rho}^B\}$ (spatial/network treatment selection). The remainder of this section gives more detail on each of the estimation steps.

Step 1: Pooled Bayesian Spatial/Network Error Probit Model

Recall the selection equation:

$$y^{A*}_{ti} = x^{A\prime}_{ti}\beta^A + \sum_{j=1}^{N} r^A_{tij}\bar{x}^{A\prime}_j\delta^A + u^A_{ti} \quad \text{with} \quad u^A_{ti} = \sum_{j=1}^{N} r^A_{tij}\xi_{tj}$$

This equation cannot be estimated directly, as r^A_{tij} is a function of the unknown spatial autocorrelation parameter ρ^A. However, $R^A_t = (I_{N_t} - \rho^A W_t)^{-1} = \sum_{l=0}^{\infty} \rho^{Al} W^l_t$ is a geometric progression, which we can approximate by $R^A_t \approx I_{N_t} + \rho^A W_t + (\rho^A)^2 W^2_t$ (Kelejian & Prucha, 1998). The selection equation then becomes:

$$y^{A*}_{ti} = x^A_{ti}\beta + \bar{x}^A_i\delta^A + \sum_{j=1}^{N} w_{tij}\bar{x}^A_j(\rho^A\delta^A) + \sum_{j=1}^{N} w^2_{tij}\bar{x}^A_j(\rho^{A2}\delta^A) + u^A_{ti} \qquad (8)$$

We estimate (8) using a **Pooled Bayesian Spatial/Network Error Probit Model** with binary y^A_{ti} as dependent variable (LeSage & Pace, 2009).[6,7] Pooling the data results in consistent estimation of the regression parameters. Note, however, that (as in standard Pooled Probit Models,

[6] We chose to estimate the first stage using a Bayesian Spatial Error Probit Model as alternatives currently available to estimate non-linear spatial models seemed unattractive: the Pinkse & Slade (1998) GMM estimator performs poorly in Monte Carlo studies (Calabrese and Elkink, 2014). Maximum Likelihood estimation faces computational issues as taking into account the spatial structure of the variance-covariance matrix of u^A_{ti} requires multidimensional integration in addition to computing the log determinant of a large matrix (Fleming, 2004).

[7] Since it is tedious to find the joint posterior distribution of the first-stage parameters analytically, we can sample

see, e.g., Arulampalam, 1999) estimates of coefficients β^A and δ^A are scaled by σ_{ξ^A}.[8] We therefore obtain the vector of estimated parameters $\hat{\theta}^B = \{\hat{\beta}^B, \hat{\delta}^B, \hat{\tau}, \hat{\rho}^B\}$ from our first-stage regression, where $\tilde{\beta}^A = \frac{\beta^A}{\sigma_{\xi^A}}$ and $\tilde{\delta}^A = \frac{\delta^A}{\sigma_{\xi^A}}$.

Step 2: Compute Correction Terms

We use the estimated first-stage parameters to construct the spatial adjustment/network factor $\hat{\psi}_i$ and the Inverse Mills' Ratio $\hat{\lambda}_{ti}$ or the Generalized Inverse Mills' Ratio $\hat{\lambda}^g_{ti}$:

$$\hat{\psi}_i = \frac{\sum_{j=1}^N r^B_{tij}\hat{r}^A_{tij}}{\sqrt{\sum_{j=1}^N (\hat{r}^A_{tij})^2}}, \quad \hat{\lambda}_{ti} = \frac{\phi(\tilde{z}_{ti})}{\Phi(\tilde{z}_{ti})}, \quad \hat{\lambda}^g_{ti} = \phi(\tilde{z}_{ti}) \frac{y^A_{ti} - \Phi(\tilde{z}_{ti})}{\Phi(\tilde{z}_{ti})\left[1-\Phi(\tilde{z}_{ti})\right]},$$
$$\text{where} \quad \tilde{z}_{ti} = \frac{x^A_{ti}\hat{\tilde{\beta}}^A + \sum_{j=1}^N \hat{r}^A_{tij}\bar{x}^A_j\hat{\tilde{\delta}}^A}{\sqrt{\sum_{j=1}^N (\hat{r}^A_{tij})^2}},$$

and $\hat{r}^A_{tij}$ is the ij-th element of the matrix $(I_{N_t} - \hat{\rho}^A W_t)^{-1}$.

Step 3: Pooled Non-Linear Least Squares

Akin to Heckman (1976, 1979) we add the correction terms $\hat{\psi}_i\hat{\lambda}_{ti}$ or $\hat{\psi}_i\hat{\lambda}^g_{ti}$ as additional regressor to the respective outcome equation.

Outcome Equation (Spatial/Network Sample Selection)

$$\begin{aligned} E[y^B_{ti}|y^A_{ti}=1, x^{A0}, x^B] &= x^{B\prime}_{ti}\beta^B + \sum_{j=1}^N r^B_{tij}\bar{x}^{B\prime}_j\delta^B + \tau\hat{\psi}_i\hat{\lambda}^g_{ti} \\ &= x^{B\prime}_{ti}\beta^B + \sum_{j=1}^N r^B_{tij}\bar{x}^{B\prime}_j\delta^B + \tau\frac{\sum_{j=1}^N r^B_{tij}\hat{r}^A_{tij}}{\sqrt{\sum_{j=1}^N (\hat{r}^A_{tij})^2}}\hat{\lambda}_{ti} \end{aligned}$$

them using a Monte Carlo Markov Chain (MCMC). Essentially, the estimation procedure involves sequentially sampling the model parameters from their respective conditional distributions. Doing this for a large number of repetitions results in a sequence of draws for the first-stage parameters that converge to the unconditional joint posterior distribution. We sample the $T\times N$ parameters in the latent variable vector y^{A*}, which follows a multivariate truncated normal distribution, using an m-step Gibbs-Sampler suggested in Geweke (1991) as discussed in LeSage and Pace (2009). Note that we cannot simply sample a sequence of $T \times N$ conditional univariate truncated normal distributions as initially proposed in LeSage (2000) and which is implemented in the LeSage Spatial Econometrics Toolbox for the SEM Probit Model.

[8]As in the non-spatial/non-network Probit model, we face an identification problem, since the several parameter values result in the same value of the likelihood function. Therefore, only $\frac{\beta^A}{\sigma_{\xi^A}}$ and $\frac{\delta^A}{\sigma_{\xi^A}}$ are identified.

Outcome Equation (Spatial/Network Treatment Selection)

$$\begin{aligned} E[y_{ti}^B | y_{ti}^A, x^{A0}, x^B] &= \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \sum_{j=1}^N r_{tij}^B \bar{x}_j^{B\prime}\delta^B + \tau\hat{\psi}_i\hat{\lambda}_{ti}^g \\ &= \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \sum_{j=1}^N r_{tij}^B \bar{x}_j^{B\prime}\delta^B + \tau\frac{\sum_{j=1}^N r_{tij}^B \hat{r}_{tij}^A}{\sqrt{\sum_{j=1}^N (\hat{r}_{tij}^A)^2}}\hat{\lambda}_{ti}^g \end{aligned}$$

Since r_{tij}^B is a non-linear function of the unknown spatial autocorrelation parameter ρ^B, we estimate the outcome equation using **Pooled Non-Linear Least Squares** to obtain the vector of parameters $\hat{\theta}_B = \{\hat{\beta}^B, \hat{\delta}^B, \hat{\tau}, \hat{\rho}^B\}$ (spatial/network sample selection) or $\hat{\theta}^B = \{\hat{\alpha}, \hat{\beta}^B, \hat{\delta}^B, \hat{\tau}, \hat{\rho}^B\}$ (spatial/network treatment selection).

4 Monte Carlo Evidence

In order to explore the finite-sample performance of the Spatial/Network Sample Selection Estimator and the Spatial/Network Treatment Selection Estimator, we conduct a set of Monte Carlo simulations. In this section, we outline the design of the respective experiments and summarize the associated results.

4.1 Monte Carlo Design

For the model outline in this subsection, it is useful to stack the data within a time period t. Moreover, the matrix of exogenous regressors in the selection equation will generally have two columns and contain one exogenous variable from the outcome equation, x_t^B, and an additional regressor which exclusively appears in the selection equation, x_t^{A0}, so that $x_t^{A\prime}=\left[x_t^B, x_t^{A0}\right]$ and $\bar{x}^{A\prime} = \left[\bar{x}^B, \bar{x}^{A0}\right]$, where bars indicate individual-specific averages over time t as above. Moreover, we will use the reuced-form expressions for residuals of $u_t^A = R_t^A(\mu^A + \varepsilon_t^A)$, $u_t^B = R_t^B(\mu^B + \varepsilon_t^B)$. Then, the data-generating processes for the latent selection variable y_t^{A*}, the selection indicator y_t^A, and outcome y_t^B can be written as follows:

Selection Equation

$$y_t^{A*} = x_t^{A\prime}\beta^A + R_t^A\bar{x}^{A\prime}\delta^A + u_t^A$$
$$y_t^A = 1[y_t^{A*} > 0]$$

Outcome Equation (Spatial/Network Sample Selection)

$$
\begin{aligned}
y_t^{B*} &= \beta_1^B x_t^B + \delta_1^B R_t^B \bar{x}^B + u_t^B \\
y_t^B &= \begin{cases} y_t^{B*} & \text{if} \quad y_t^A = 1 \\ - & \text{if} \quad y_t^A = 0 \end{cases}
\end{aligned}
$$

Outcome Equation (Spatial/Network Treatment Selection)

$$
y_t^B = \alpha y_t^A + \beta_1^B x_t^B + \delta_1^B R_t^B \bar{x}^B + u_t^B .
$$

Hence, the outcome equation contains only one time-variant exogenous regressor, x_t^B, while the selection equation contains two time-variant exogenous regressors, $\left[x_t^B, x_t^{A0}\right]$. Each of the elements in $\left[x_t^B, x_t^{A0}\right]$ will be generated independently by randomly drawing from a univariate standard normal distribution, $\mathcal{N}(0,1)$. The $N_t \times N_t$ spatial/network weights matrix W_t underlying $R_t^\ell = (I_N - \rho^\ell W_t)^{-1}$ will be based on a row-normalized 5-before-5-behind wrap-around structure, akin to the design in Kapoor, Kelejian, and Prucha (2007) and Baltagi, Egger, and Kesina (2017).[9] Hence, the elements of W_t are either 0 or 0.1.

The error components μ_i^A, μ_i^B, ε_{ti}^A, and ε_{ti}^B are drawn from the following joint normal distributions (except for one configuration for each model, where $\mu_i^{AB} = \mu_i^{BA} = 0$, and $\varepsilon_{ti}^{AB} = \varepsilon_{ti}^{BA} = 0$, i.e. $\tau = 0$).

$$
\begin{pmatrix} \mu_i^A \\ \mu_i^B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)
$$
$$
\begin{pmatrix} \varepsilon_{ti}^A \\ \varepsilon_{ti}^B \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)
$$

The following parameter configurations stay the same throughout all experiments:

Selection Equation:
The parameters on x_t^B and x_t^{A0} in the selection equation are β_1^A=1, β_2^A=1, and the parameters on $\bar{x}^B$ and $\bar{x}^{A0}$ in the selection equation are δ_1^A=1, δ_2^A=1. Accordingly, the actually estimated

[9] In the case of 20 ordered individuals, such a structure entails that individual 1 has the neighbors $\{16; 17; 18; 19; 20\}$ "before" it and the neighbors $\{2; 3; 4; 5; 6\}$ "behind" it, individual 2 has the neighbors $\{17; 18; 19; 20; 1\}$ "before" it and the neighbors $\{3; 4; 5; 6; 7\}$ "behind" it, etc.

parameters in the selection equation are $\tilde{\beta}_1^A$=0.707, $\tilde{\beta}_2^A$=0.707, $\tilde{\delta}_1^A$=0.707, $\tilde{\delta}_2^A$=0.707

Outcome Equation (Spatial/Network Sample Selection):
With spatial/network sample selection, The parameters on x_t^B and The parameters on $\bar{x}^B$ in the outcome equation are β_1^B=1 and δ_1^B=3, respectively.

Outcome Equation (Spatial/Network Treatment Selection):
Finally, with spatial/network treatment selection, the average-treatment-effect parameter on the binary treatment indicator y_t^A in the outcome equation is α=1 and the other parameters are identical to the case of sample selection: β_1^B=1, and δ_1^B=3.

In any case, the spatial/network autocorrelation parameters take on different values, namely $\rho^A = \{0; 0.5; 0.75\}$ and $\rho^B = \{0; 0.5; 0.75\}$. We run experiments for two cross-sectional sample sizes of $N = \{250; 500\}$, we generally set $T = 3$, and conduct $M = 1000$ draws.[10] Hence, we end up with 36 different simulations which are based on 9 configurations of (ρ^A, ρ^B) at $\tau = \frac{\sigma_{\xi^B A}}{\sqrt{\sigma^2_{\xi^A}}} = \frac{1}{\sqrt{2}} = 0.707$ for each one of the two cases of sample and treatment selection. Moreover, there are 2 simulations which are based on 1 configuration with $(\rho^A = 0, \rho^B = 0, \tau = 0)$ for each one of the two cases of sample and treatment selection. The corresponding results are presented in the subsequent section.

4.2 Results

We report the results of our Monte Carlo experiment in the following two tables, where Table 1 is devoted to the case of sample selection, while Table 2 is devoted to the case of treatment selection. In each table, we report on the cross-sectional sample size underlying the selection equation in the first column and on the employed estimation procedure in the second column. We generally compare the parameters of three alternative estimators. The Spatial/Network Sample Selection Estimator (SNSS) and the Spatial/Network Treatment Selection Estimator (SNTS) at the top. This estimator is based on the model outline and routines presented above and involves the selection equation in (8) and the outcome equations in (6) and (7). Wooldridge's (1995) estimator for sample selection

[10]For each one of these 1000 draws, we generate a chain of 2500 draws for the selection equation in the first step, of which we discard the first 500 ones (burn-in) and use all of the remaining 2000 ones (i.e., use a thinning of zero) to compute moments of the parameters of the selection equation before proceeding to the second step.

with panel-data (WSS) and an adapted version thereof for treatment selection (WTS) either of which ignores the spatial/network interdependence in the data. This estimator is based on the following selection and outcome equations:

Selection Equation:

$$\begin{aligned} y_{ti}^{A*} &= x_{ti}^{A\prime}\beta^A + \bar{x}_i^{A\prime}\delta^A + \mu_i^A + \varepsilon_{ti}^A, \\ y_{ti}^A &= 1[y_{ti}^{A*} > 0] \end{aligned} \tag{9}$$

Outcome Equation (WSS)

$$E[y_{ti}^B | y_{ti}^A = 1, x^{A0}, x^B] = x_{ti}^{B\prime}\beta^B + \bar{x}_i^{B\prime}\delta^B + \tau\lambda_{ti}^W \tag{10}$$

Outcome Equation (WTS)

$$E[y_{ti}^B | y_{ti}^A, x^{A0}, x^B] = \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \bar{x}_i^{B\prime}\delta^B + \tau\lambda_{ti}^{gW}, \tag{11}$$

with $\lambda_{ti}^W = \frac{\phi\left(z_{ti}^W\right)}{\Phi\left(z_{ti}^W\right)}$, $\lambda_{ti}^{gW} = \phi\left(z_{ti}^W\right)\frac{y_{ti}^A - \Phi\left(z_{ti}^W\right)}{\Phi\left(z_{ti}^W\right)[1-\Phi\left(z_{ti}^W\right)]}$, and $z_{ti}^W = \frac{x_{ti}^A\beta^A + \bar{x}_i^A\delta^A}{\sqrt{\sigma_{\xi^A}^2}}$.

Finally, we present the results for a pooled non-linear least-squares (NLLS) model of the form

$$\begin{aligned} E[y_{ti}^B | x^B] &= x_{ti}^{B\prime}\beta^B + \sum_{j=1}^{N} r_{tij}^B \bar{x}_j^{B\prime}\delta^B, \\ E[y_{ti}^B | x^B] &= \alpha y_{ti}^A + x_{ti}^{B\prime}\beta^B + \sum_{j=1}^{N} r_{tij}^B \bar{x}_j^{B\prime}\delta^B, \end{aligned} \tag{12}$$

which ignores the issue of endogenous selection (and, hence, the spatial/network adjusted Inverse Mills' Ratio/generalized Inverse Mills' Ratio) in Step 2.

In the third column, we indicate which statistic the cells to the right reflect, namely the Mean (parameter point estimate), the Bias, and the RMSE (root mean-squared error). The remaining columns pertain to statistics of the individual model parameters of interest which are mentioned at the top of the table.

In a vertical dimension, each one of the two tables is organized in 10 panels (dubbed Panel a to Panel j), each of which contains the results for one configuration of (ρ^A, ρ^B, τ). We assume that $\tau = 0.707$ for any configuration of (ρ^A, ρ^B) in Panels a-i. Hence, for one of the configurations in those panels, namely where $(\rho^A = 0, \rho^B = 0)$, the WSS (WTS) estimator is the right one to use, while SNSS (SNTS) is inefficient and NLLS is biased. In Panel j, ($\rho^A = 0, \rho^B = 0, \tau = 0$), so that NLLS, SNSS (SNTS) and the WSS (WTS) estimators are inefficient as each if them neglects to restrict at least one of the above parameters to zero.

Table 1: Spatial/Network Sample Selection: Results for N={250;500}, T=3, M=1000

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	β_1^B	δ_1^B	τ	ρ^B
Panel a		True	0.707	0.707	0.707	0.707	0	1	3	0.707	0
N=250	SNSS	Mean	0.739	0.740	0.743	0.744	-0.114	1.001	3.003	0.699	-0.024
		Bias	0.032	0.033	0.036	0.037	-0.114	0.001	0.003	-0.008	-0.024
		RMSE	0.092	0.085	0.193	0.177	0.245	0.082	0.185	0.126	0.189
	WSS	Mean	0.715	0.716	0.719	0.718		0.999	3.011	0.705	
		Bias	0.008	0.009	0.011	0.010		-0.001	0.011	-0.002	
		RMSE	0.082	0.075	0.143	0.140		0.082	0.180	0.127	
	NLLS	Mean						0.996	2.955		-0.119
		Bias						-0.004	-0.045		-0.119
		RMSE						0.082	0.191		0.227
N=500	SNSS	Mean	0.721	0.723	0.730	0.718	-0.066	0.999	3.000	0.703	-0.006
		Bias	0.014	0.016	0.023	0.011	-0.066	-0.001	0.000	-0.004	-0.006
		RMSE	0.055	0.059	0.133	0.133	0.176	0.050	0.127	0.087	0.107
	WSS	Mean	0.709	0.711	0.714	0.706		0.998	3.002	0.707	
		Bias	0.002	0.004	0.007	-0.001		-0.002	0.002	0.000	
		RMSE	0.052	0.056	0.102	0.102		0.050	0.125	0.087	
	NLLS	Mean						0.988	3.006		0.012
		Bias						-0.012	0.006		0.012
		RMSE						0.052	0.126		0.109
Panel b		True	0.707	0.707	0.707	0.707	0.5	1	3	0.707	0
N=250	SNSS	Mean	0.732	0.735	0.736	0.734	0.449	1.000	3.004	0.700	-0.024
		Bias	0.025	0.028	0.029	0.027	-0.051	0.000	0.004	-0.007	-0.024
		RMSE	0.091	0.084	0.198	0.185	0.131	0.081	0.183	0.125	0.187
	WSS	Mean	0.653	0.670	0.681	0.708		0.997	3.012	0.676	
		Bias	-0.054	-0.037	-0.026	0.001		-0.003	0.012	-0.031	
		RMSE	0.096	0.081	0.140	0.134		0.081	0.178	0.125	
	NLLS	Mean						0.995	2.953		-0.114
		Bias						-0.005	-0.047		-0.114
		RMSE						0.081	0.189		0.222
N=500	SNSS	Mean	0.716	0.719	0.721	0.711	0.482	1.000	2.998	0.702	-0.001
		Bias	0.009	0.012	0.014	0.004	-0.018	0.000	-0.002	-0.005	-0.001
		RMSE	0.055	0.060	0.134	0.140	0.074	0.050	0.127	0.090	0.102
	WSS	Mean	0.660	0.657	0.686	0.671		0.999	2.998	0.671	
		Bias	-0.047	-0.050	-0.022	-0.036		-0.001	-0.002	-0.036	
		RMSE	0.069	0.074	0.098	0.109		0.050	0.125	0.093	
	NLLS	Mean						0.990	2.997		0.014
		Bias						-0.010	-0.003		0.014
		RMSE						0.051	0.126		0.107

Table 1: Spatial/Network Sample Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	β_1^B	δ_1^B	τ	ρ^B
Panel c		True	0.707	0.707	0.707	0.707	0.75	1	3	0.707	0
N=250	SNSS	Mean	0.736	0.739	0.738	0.728	0.735	1.000	3.005	0.705	-0.025
		Bias	0.029	0.032	0.030	0.021	-0.015	0.000	0.005	-0.002	-0.025
		RMSE	0.102	0.095	0.206	0.203	0.056	0.080	0.180	0.149	0.191
	WSS	Mean	0.509	0.533	0.587	0.643		0.997	3.011	0.581	
		Bias	-0.198	-0.174	-0.121	-0.065		-0.003	0.011	-0.126	
		RMSE	0.211	0.188	0.174	0.140		0.079	0.173	0.175	
	NLLS	Mean						0.997	2.952		-0.107
		Bias						-0.003	-0.048		-0.107
		RMSE						0.080	0.185		0.221
N=500	SNSS	Mean	0.721	0.722	0.714	0.707	0.746	0.999	2.999	0.703	-0.002
		Bias	0.014	0.015	0.007	0.000	-0.004	-0.001	-0.001	-0.004	-0.002
		RMSE	0.061	0.067	0.147	0.150	0.036	0.049	0.127	0.103	0.103
	WSS	Mean	0.528	0.521	0.603	0.580		0.997	2.997	0.572	
		Bias	-0.179	-0.186	-0.104	-0.127		-0.003	-0.003	-0.135	
		RMSE	0.186	0.193	0.137	0.159		0.049	0.124	0.158	
	NLLS	Mean						0.991	2.993		0.009
		Bias						-0.009	-0.007		0.009
		RMSE						0.050	0.125		0.107
Panel d		True	0.707	0.707	0.707	0.707	0	1	3	0.707	0.5
N=250	SNSS	Mean	0.739	0.740	0.743	0.744	-0.114	1.000	3.010	0.701	0.488
		Bias	0.032	0.033	0.036	0.037	-0.114	0.000	0.010	-0.007	-0.012
		RMSE	0.092	0.085	0.193	0.177	0.245	0.098	0.196	0.183	0.111
	WSS	Mean	0.715	0.716	0.719	0.718		0.963	3.043	0.628	
		Bias	0.008	0.009	0.011	0.011		-0.037	0.043	-0.079	
		RMSE	0.082	0.075	0.143	0.140		0.104	0.196	0.209	
	NLLS	Mean						0.988	2.993		0.423
		Bias						-0.012	-0.007		-0.077
		RMSE						0.097	0.192		0.149
N=500	SNSS	Mean	0.721	0.723	0.730	0.718	-0.066	0.999	3.003	0.705	0.495
		Bias	0.014	0.016	0.023	0.011	-0.066	-0.001	0.003	-0.002	-0.005
		RMSE	0.055	0.059	0.133	0.133	0.176	0.063	0.143	0.131	0.063
	WSS	Mean	0.709	0.711	0.714	0.706		0.982	3.226	0.758	
		Bias	0.002	0.004	0.007	-0.001		-0.018	0.226	0.051	
		RMSE	0.052	0.056	0.102	0.102		0.065	0.268	0.146	
	NLLS	Mean						0.988	2.997		0.506
		Bias						-0.012	-0.003		0.006
		RMSE						0.063	0.143		0.064

Table 1: Spatial/Network Sample Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	β_1^B	δ_1^B	τ	ρ^B
Panel e		True	0.707	0.707	0.707	0.707	0.5	1	3	0.707	0.5
N=250	SNSS	Mean	0.732	0.735	0.736	0.734	0.449	0.999	3.011	0.699	0.487
		Bias	0.025	0.028	0.029	0.027	-0.051	-0.001	0.011	-0.009	-0.013
		RMSE	0.091	0.084	0.198	0.185	0.131	0.095	0.192	0.173	0.109
	WSS	Mean	0.653	0.670	0.681	0.708		0.964	3.038	0.679	
		Bias	-0.054	-0.037	-0.026	0.001		-0.036	0.038	-0.028	
		RMSE	0.096	0.081	0.140	0.134		0.102	0.192	0.189	
	NLLS	Mean						0.987	2.988		0.421
		Bias						-0.013	-0.012		-0.079
		RMSE						0.095	0.189		0.149
N=500	SNSS	Mean	0.716	0.719	0.721	0.711	0.482	0.999	3.001	0.702	0.497
		Bias	0.009	0.012	0.014	0.004	-0.018	-0.001	0.001	-0.005	-0.003
		RMSE	0.055	0.060	0.134	0.140	0.074	0.061	0.141	0.127	0.061
	WSS	Mean	0.660	0.657	0.686	0.671		0.982	3.222	0.808	
		Bias	-0.047	-0.050	-0.022	-0.036		-0.018	0.222	0.101	
		RMSE	0.069	0.074	0.098	0.109		0.064	0.264	0.170	
	NLLS	Mean						0.990	2.987		0.508
		Bias						-0.010	-0.013		0.008
		RMSE						0.061	0.141		0.064
Panel f		True	0.707	0.707	0.707	0.707	0.75	1	3	0.707	0.5
N=250	SNSS	Mean	0.736	0.739	0.738	0.728	0.735	0.999	3.012	0.699	0.487
		Bias	0.029	0.032	0.030	0.021	-0.015	-0.001	0.012	-0.008	-0.013
		RMSE	0.102	0.095	0.206	0.203	0.056	0.093	0.189	0.189	0.112
	WSS	Mean	0.509	0.533	0.587	0.643		0.968	3.028	0.659	
		Bias	-0.198	-0.174	-0.121	-0.065		-0.032	0.028	-0.048	
		RMSE	0.211	0.188	0.174	0.140		0.098	0.187	0.195	
	NLLS	Mean						0.989	2.981		0.420
		Bias						-0.011	-0.019		-0.080
		RMSE						0.092	0.184		0.154
N=500	SNSS	Mean	0.721	0.722	0.714	0.707	0.746	0.998	3.001	0.700	0.497
		Bias	0.014	0.015	0.007	0.000	-0.004	-0.002	0.001	-0.007	-0.003
		RMSE	0.061	0.067	0.147	0.150	0.036	0.058	0.140	0.136	0.061
	WSS	Mean	0.528	0.521	0.603	0.580		0.982	3.215	0.789	
		Bias	-0.179	-0.186	-0.104	-0.127		-0.018	0.215	0.082	
		RMSE	0.186	0.193	0.137	0.159		0.061	0.258	0.160	
	NLLS	Mean						0.990	2.982		0.507
		Bias						-0.010	-0.018		0.007
		RMSE						0.059	0.140		0.065

Table 1: Spatial/Network Sample Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	β_1^B	δ_1^B	τ	ρ^B
Panel g		True	0.707	0.707	0.707	0.707	0	1	3	0.707	0.75
N=250	SNSS	Mean	0.739	0.740	0.743	0.744	-0.114	1.000	3.015	0.701	0.740
		Bias	0.032	0.033	0.036	0.037	-0.114	0.000	0.015	-0.006	-0.010
		RMSE	0.092	0.085	0.193	0.177	0.245	0.149	0.247	0.306	0.071
	WSS	Mean	0.715	0.716	0.719	0.718		0.890	3.215	0.485	
		Bias	0.008	0.009	0.011	0.011		-0.110	0.215	-0.222	
		RMSE	0.082	0.075	0.143	0.140		0.183	0.321	0.422	
	NLLS	Mean						0.972	3.035		0.707
		Bias						-0.028	0.035		-0.043
		RMSE						0.143	0.239		0.091
N=500	SNSS	Mean	0.721	0.723	0.730	0.718	-0.066	0.999	3.005	0.706	0.746
		Bias	0.014	0.016	0.023	0.011	-0.066	-0.001	0.005	-0.001	-0.004
		RMSE	0.055	0.059	0.133	0.133	0.176	0.098	0.194	0.225	0.040
	WSS	Mean	0.709	0.711	0.714	0.706		0.963	3.702	0.886	
		Bias	0.002	0.004	0.007	-0.001		-0.037	0.702	0.179	
		RMSE	0.052	0.056	0.102	0.102		0.105	0.731	0.314	
	NLLS	Mean						0.990	2.983		0.753
		Bias						-0.010	-0.017		0.003
		RMSE						0.097	0.196	0.040	
Panel h		True	0.707	0.707	0.707	0.707	0.5	1	3	0.707	0.75
N=250	SNSS	Mean	0.732	0.735	0.736	0.734	0.449	0.999	3.017	0.695	0.740
		Bias	0.025	0.028	0.029	0.027	-0.051	-0.001	0.017	-0.012	-0.010
		RMSE	0.091	0.084	0.198	0.185	0.131	0.145	0.241	0.275	0.071
	WSS	Mean	0.653	0.670	0.681	0.708		0.894	3.201	0.656	
		Bias	-0.054	-0.037	-0.026	0.001		-0.106	0.201	-0.051	
		RMSE	0.096	0.081	0.140	0.134		0.176	0.307	0.359	
	NLLS	Mean						0.967	3.033		0.703
		Bias						-0.033	0.033		-0.047
		RMSE						0.140	0.232		0.095
N=500	SNSS	Mean	0.716	0.719	0.721	0.711	0.482	0.999	3.002	0.700	0.747
		Bias	0.009	0.012	0.014	0.004	-0.018	-0.001	0.002	-0.007	-0.003
		RMSE	0.055	0.060	0.134	0.140	0.074	0.095	0.191	0.205	0.039
	WSS	Mean	0.660	0.657	0.686	0.671		0.962	3.696	1.063	
		Bias	-0.047	-0.050	-0.022	-0.036		-0.038	0.696	0.356	
		RMSE	0.069	0.074	0.098	0.109		0.103	0.724	0.443	
	NLLS	Mean						0.991	2.968		0.755
		Bias						-0.009	-0.032		0.005
		RMSE						0.094	0.196		0.040

Table 1: Spatial/Network Sample Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	β_1^B	δ_1^B	τ	ρ^B
Panel i		True	0.707	0.707	0.707	0.707	0.75	1	3	0.707	0.75
N=250	SNSS	Mean	0.736	0.739	0.738	0.728	0.735	0.998	3.019	0.691	0.739
		Bias	0.029	0.032	0.030	0.021	-0.015	-0.002	0.019	-0.016	-0.011
		RMSE	0.102	0.095	0.206	0.203	0.056	0.138	0.237	0.276	0.072
	WSS	Mean	0.509	0.533	0.587	0.643		0.907	3.179	0.755	
		Bias	-0.198	-0.174	-0.121	-0.065		-0.093	0.179	0.048	
		RMSE	0.211	0.188	0.174	0.140		0.163	0.290	0.367	
	NLLS	Mean						0.967	3.026		0.697
		Bias						-0.033	0.026		-0.053
		RMSE						0.133	0.224		0.101
N=500	SNSS	Mean	0.721	0.722	0.714	0.707	0.746	0.998	3.002	0.694	0.747
		Bias	0.014	0.015	0.007	0.000	-0.004	-0.002	0.002	-0.013	-0.003
		RMSE	0.061	0.067	0.147	0.150	0.036	0.088	0.187	0.203	0.038
	WSS	Mean	0.528	0.521	0.603	0.580		0.962	3.680	1.168	
		Bias	-0.179	-0.186	-0.104	-0.127		-0.038	0.680	0.461	
		RMSE	0.186	0.193	0.137	0.159		0.097	0.708	0.536	
	NLLS	Mean						0.991	2.957		0.756
		Bias						-0.009	-0.043		0.006
		RMSE						0.088	0.196		0.041
Panel j		True	0.707	0.707	0.707	0.707	0.5	1	3	0	0
N=250	SNSS	Mean	0.739	0.741	0.744	0.752	-0.123	0.998	3.001	0.004	-0.014
		Bias	0.032	0.034	0.037	0.045	-0.123	-0.002	0.001	0.004	-0.014
		RMSE	0.093	0.084	0.197	0.184	0.248	0.085	0.189	0.140	0.199
	WSS	Mean	0.715	0.717	0.719	0.721		0.997	3.011	0.001	
		Bias	0.007	0.010	0.012	0.014		-0.003	0.011	0.001	
		RMSE	0.083	0.073	0.147	0.144		0.085	0.186	0.141	
	NLLS	Mean						0.998	3.001		-0.014
		Bias						-0.002	0.001		-0.014
		RMSE						0.085	0.189		0.199
N=500	SNSS	Mean	0.721	0.720	0.726	0.723	-0.064	0.998	2.997	0.002	-0.001
		Bias	0.014	0.013	0.019	0.016	-0.064	-0.002	-0.003	0.002	-0.001
		RMSE	0.056	0.061	0.133	0.135	0.175	0.051	0.131	0.100	0.111
	WSS	Mean	0.708	0.709	0.711	0.709		0.998	3.002	0.002	
		Bias	0.001	0.001	0.004	0.002		-0.002	0.002	0.002	
		RMSE	0.053	0.057	0.103	0.104		0.052	0.131	0.101	
	NLLS	Mean						0.998	2.997		-0.001
		Bias						-0.002	-0.003		-0.001
		RMSE						0.051	0.131		0.111

Table 2: Spatial/Network Treatment Selection: Results for N={250;500}, T=3, M=1000

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	α	β_1^B	δ_1^B	τ	ρ^B
Panel a		True	0.707	0.707	0.707	0.707	0	1	1	3	0.707	0
N=250	SNTS	Mean	0.739	0.740	0.743	0.744	-0.114	1.005	1.000	3.006	0.697	-0.009
		Bias	0.032	0.033	0.036	0.037	-0.114	0.005	0.000	0.006	-0.010	-0.009
		RMSE	0.092	0.085	0.193	0.177	0.245	0.112	0.051	0.132	0.106	0.137
	WTS	Mean	0.715	0.716	0.719	0.718		1.000	0.999	3.009	0.706	
		Bias	0.008	0.009	0.011	0.011		0.000	-0.001	0.009	-0.001	
		RMSE	0.082	0.075	0.143	0.140		0.112	0.050	0.130	0.106	
	NLLS	Mean						1.350	0.942	2.973		0.067
		Bias						0.350	-0.058	-0.027		0.067
		RMSE						0.362	0.076	0.130		0.148
N=500	SNTS	Mean	0.721	0.723	0.730	0.718	-0.066	1.003	0.999	3.002	0.701	-0.006
		Bias	0.014	0.016	0.023	0.011	-0.066	0.003	-0.001	0.002	-0.006	-0.006
		RMSE	0.055	0.059	0.133	0.133	0.176	0.086	0.033	0.099	0.076	0.087
	WTS	Mean	0.709	0.711	0.714	0.706		1.001	0.999	3.002	0.705	
		Bias	0.002	0.004	0.007	-0.001		0.001	-0.001	0.002	-0.002	
		RMSE	0.052	0.056	0.102	0.102		0.086	0.033	0.096	0.076	
	NLLS	Mean						1.358	0.942	2.935		-0.016
		Bias						0.358	-0.058	-0.065		-0.016
		RMSE						0.365	0.066	0.114		0.093
Panel b		True	0.707	0.707	0.707	0.707	0.5	1	1	3	0.707	0
N=250	SNTS	Mean	0.732	0.735	0.736	0.734	0.449	1.006	1.001	3.004	0.694	-0.008
		Bias	0.025	0.028	0.029	0.027	-0.051	0.006	0.001	0.004	-0.013	-0.008
		RMSE	0.091	0.084	0.198	0.185	0.131	0.108	0.050	0.132	0.107	0.135
	WTS	Mean	0.653	0.670	0.681	0.708		0.995	1.000	3.010	0.684	
		Bias	-0.054	-0.037	-0.026	0.001		-0.005	0.000	0.010	-0.023	
		RMSE	0.096	0.081	0.140	0.134		0.108	0.050	0.130	0.107	
	NLLS	Mean						1.343	0.945	2.974		0.052
		Bias						0.343	-0.055	-0.026		0.052
		RMSE						0.355	0.073	0.130		0.143
N=500	SNTS	Mean	0.716	0.719	0.721	0.711	0.482	1.003	0.999	3.000	0.699	-0.004
		Bias	0.009	0.012	0.014	0.004	-0.018	0.003	-0.001	0.000	-0.008	-0.004
		RMSE	0.055	0.060	0.134	0.140	0.074	0.083	0.033	0.099	0.085	0.077
	WTS	Mean	0.660	0.657	0.686	0.671		0.991	1.000	3.004	0.685	
		Bias	-0.047	-0.050	-0.022	-0.036		-0.009	0.000	0.004	-0.022	
		RMSE	0.069	0.074	0.098	0.109		0.084	0.033	0.096	0.078	
	NLLS	Mean						1.355	0.944	2.936		-0.031
		Bias						0.355	-0.056	-0.064		-0.031
		RMSE						0.361	0.064	0.114		0.097

Table 2: Spatial/Network Treatment Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	α	β_1^B	δ_1^B	τ	ρ^B
Panel c		True	0.707	0.707	0.707	0.707	0.75	1	1	3	0.707	0
N=250	SNTS	Mean	0.736	0.739	0.738	0.728	0.735	1.007	1.002	3.004	0.692	-0.007
		Bias	0.029	0.032	0.030	0.021	-0.015	0.007	0.002	0.004	-0.016	-0.007
		RMSE	0.102	0.095	0.206	0.203	0.056	0.110	0.050	0.133	0.122	0.135
	WTS	Mean	0.509	0.533	0.587	0.643		0.988	1.001	3.010	0.598	
		Bias	-0.198	-0.174	-0.121	-0.065		-0.012	0.001	0.010	-0.109	
		RMSE	0.211	0.188	0.174	0.140		0.111	0.049	0.130	0.153	
	NLLS	Mean						1.323	0.954	2.976		0.034
		Bias						0.323	-0.046	-0.024		0.034
		RMSE						0.334	0.066	0.129		0.138
N=500	SNTS	Mean	0.721	0.722	0.714	0.707	0.746	1.003	1.000	2.999	0.699	-0.004
		Bias	0.014	0.015	0.007	0.000	-0.004	0.003	0.000	-0.001	-0.008	-0.004
		RMSE	0.061	0.067	0.147	0.150	0.036	0.084	0.032	0.100	0.086	0.084
	WTS	Mean	0.528	0.521	0.603	0.580		0.976	1.002	3.007	0.606	
		Bias	-0.179	-0.186	-0.104	-0.127		-0.024	0.002	0.007	-0.101	
		RMSE	0.186	0.193	0.137	0.159		0.088	0.032	0.097	0.125	
	NLLS	Mean						1.336	0.952	2.941		-0.046
		Bias						0.336	-0.048	-0.059		-0.046
		RMSE						0.342	0.057	0.111		0.103
Panel d		True	0.707	0.707	0.707	0.707	0	1	1	3	0.707	0.5
N=250	SNTS	Mean	0.739	0.740	0.743	0.744	-0.114	1.006	1.000	3.006	0.699	0.493
		Bias	0.032	0.033	0.036	0.037	-0.114	0.006	0.000	0.006	-0.008	-0.007
		RMSE	0.092	0.085	0.193	0.177	0.245	0.215	0.062	0.139	0.149	0.095
	WTS	Mean	0.715	0.716	0.719	0.718		0.856	1.023	3.075	0.816	
		Bias	0.008	0.009	0.011	0.011		-0.144	0.023	0.075	0.108	
		RMSE	0.082	0.075	0.143	0.140		0.259	0.066	0.159	0.190	
	NLLS	Mean						1.367	0.940	2.949		0.532
		Bias						0.367	-0.060	-0.051		0.032
		RMSE						0.402	0.083	0.143		0.095
N=500	SNTS	Mean	0.721	0.723	0.730	0.718	-0.066	1.004	0.999	3.003	0.702	0.495
		Bias	0.014	0.016	0.023	0.011	-0.066	0.004	-0.001	0.003	-0.005	-0.005
		RMSE	0.055	0.059	0.133	0.133	0.176	0.165	0.041	0.106	0.110	0.057
	WTS	Mean	0.709	0.711	0.714	0.706		1.038	0.993	3.158	0.711	
		Bias	0.002	0.004	0.007	-0.001		0.038	-0.007	0.158	0.004	
		RMSE	0.052	0.056	0.102	0.102		0.170	0.041	0.191	0.118	
	NLLS	Mean						1.372	0.940	2.936		0.496
		Bias						0.372	-0.060	-0.064		-0.004
		RMSE						0.391	0.070	0.118		0.059

Table 2: Spatial/Network Treatment Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	α	β_1^B	δ_1^B	τ	ρ^B
Panel e		True	0.707	0.707	0.707	0.707	0.5	1	1	3	0.707	0.5
N=250	SNTS	Mean	0.732	0.735	0.736	0.734	0.449	1.006	1.002	3.005	0.694	0.494
		Bias	0.025	0.028	0.029	0.027	-0.051	0.006	0.002	0.005	-0.013	-0.006
		RMSE	0.091	0.084	0.198	0.185	0.131	0.207	0.060	0.138	0.146	0.091
	WTS	Mean	0.653	0.670	0.681	0.708		0.849	1.023	3.076	0.873	
		Bias	-0.054	-0.037	-0.026	0.001		-0.151	0.023	0.076	0.166	
		RMSE	0.096	0.081	0.140	0.134		0.257	0.065	0.159	0.231	
	NLLS	Mean						1.379	0.940	2.951		0.528
		Bias						0.379	-0.059	-0.049		0.028
		RMSE						0.411	0.082	0.142		0.094
N=500	SNTS	Mean	0.716	0.719	0.721	0.711	0.482	1.004	0.999	3.002	0.698	0.495
		Bias	0.009	0.012	0.014	0.004	-0.018	0.004	-0.001	0.002	-0.009	-0.005
		RMSE	0.055	0.060	0.134	0.140	0.074	0.160	0.040	0.105	0.108	0.055
	WTS	Mean	0.660	0.657	0.686	0.671		1.027	0.995	3.160	0.779	
		Bias	-0.047	-0.050	-0.022	-0.036		0.027	-0.005	0.160	0.072	
		RMSE	0.069	0.074	0.098	0.109		0.163	0.040	0.193	0.140	
	NLLS	Mean						1.390	0.939	2.938		0.490
		Bias						0.390	-0.061	-0.062		-0.010
		RMSE						0.408	0.071	0.117		0.060
Panel f		True	0.707	0.707	0.707	0.707	0.75	1	1	3	0.707	0.5
N=250	SNTS	Mean	0.736	0.739	0.738	0.728	0.735	1.008	1.002	3.005	0.688	0.495
		Bias	0.029	0.032	0.030	0.021	-0.015	0.008	0.002	0.005	-0.019	-0.005
		RMSE	0.102	0.095	0.206	0.203	0.056	0.204	0.059	0.138	0.162	0.088
	WTS	Mean	0.509	0.533	0.587	0.643		0.834	1.022	3.077	0.867	
		Bias	-0.198	-0.174	-0.121	-0.065		-0.166	0.022	0.077	0.160	
		RMSE	0.211	0.188	0.174	0.140		0.265	0.062	0.160	0.233	
	NLLS	Mean						1.378	0.948	2.956		0.521
		Bias						0.378	-0.052	-0.044		0.021
		RMSE						0.409	0.075	0.139		0.092
N=500	SNTS	Mean	0.721	0.722	0.714	0.707	0.746	1.004	1.000	3.000	0.696	0.496
		Bias	0.014	0.015	0.007	0.000	-0.004	0.004	0.000	0.000	-0.011	-0.004
		RMSE	0.061	0.067	0.147	0.150	0.036	0.157	0.038	0.104	0.114	0.054
	WTS	Mean	0.528	0.521	0.603	0.580		1.007	0.998	3.164	0.787	
		Bias	-0.179	-0.186	-0.104	-0.127		0.007	-0.002	0.164	0.080	
		RMSE	0.186	0.193	0.137	0.159		0.159	0.038	0.196	0.148	
	NLLS	Mean						1.395	0.944	2.944		0.482
		Bias						0.395	-0.056	-0.056		-0.018
		RMSE						0.412	0.066	0.113		0.062

Table 2: Spatial/Network Treatment Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	α	β_1^B	δ_1^B	τ	ρ^B
Panel g		True	0.707	0.707	0.707	0.707	0	1	1	3	0.707	0.75
N=250	SNTS	Mean	0.739	0.740	0.743	0.744	-0.114	1.010	1.001	3.002	0.699	0.743
		Bias	0.032	0.033	0.036	0.037	-0.114	0.010	0.001	0.002	-0.009	-0.007
		RMSE	0.092	0.085	0.193	0.177	0.245	0.432	0.097	0.176	0.245	0.065
	WTS	Mean	0.715	0.716	0.719	0.718		0.593	1.066	3.311	1.036	
		Bias	0.008	0.009	0.011	0.011		-0.407	0.066	0.311	0.330	
		RMSE	0.082	0.075	0.143	0.140		0.592	0.115	0.358	0.434	
	NLLS	Mean						1.412	0.938	2.919		0.764
		Bias						0.412	-0.062	-0.081		0.014
		RMSE						0.520	0.105	0.194		0.063
N=500	SNTS	Mean	0.721	0.723	0.730	0.718	-0.066	1.005	0.999	3.002	0.702	0.747
		Bias	0.014	0.016	0.023	0.011	-0.066	0.005	-0.001	0.002	-0.005	-0.003
		RMSE	0.055	0.059	0.133	0.133	0.176	0.330	0.066	0.144	0.185	0.037
	WTS	Mean	0.709	0.711	0.714	0.706		1.109	0.982	3.532	0.741	
		Bias	0.002	0.004	0.007	-0.001		0.109	-0.018	0.532	0.033	
		RMSE	0.052	0.056	0.102	0.102		0.349	0.066	0.551	0.222	
	NLLS	Mean						1.407	0.933	2.936		0.749
		Bias						0.407	-0.067	-0.064		-0.001
		RMSE						0.471	0.086	0.149		0.037
Panel h		True	0.707	0.707	0.707	0.707	0.5	1	1	3	0.707	0.75
N=250	SNTS	Mean	0.732	0.735	0.736	0.734	0.449	1.010	1.003	3.001	0.692	0.744
		Bias	0.025	0.028	0.029	0.027	-0.051	0.010	0.003	0.001	-0.015	-0.006
		RMSE	0.091	0.084	0.198	0.185	0.131	0.419	0.092	0.174	0.062	0.228
	WTS	Mean	0.653	0.670	0.681	0.708		0.584	1.064	3.312	1.211	
		Bias	-0.054	-0.037	-0.026	0.001		-0.416	0.064	0.312	0.504	
		RMSE	0.096	0.081	0.140	0.134		0.591	0.112	0.358	0.586	
	NLLS	Mean						1.452	0.936	2.914		0.764
		Bias						0.452	-0.064	-0.086		0.014
		RMSE						0.550	0.105	0.195		0.063
N=500	SNTS	Mean	0.716	0.719	0.721	0.711	0.482	1.006	0.999	3.001	0.696	0.747
		Bias	0.009	0.012	0.014	0.004	-0.018	0.006	-0.001	0.001	-0.011	-0.003
		RMSE	0.055	0.060	0.134	0.140	0.074	0.324	0.063	0.141	0.170	0.035
	WTS	Mean	0.660	0.657	0.686	0.671		1.096	0.984	3.534	0.940	
		Bias	-0.047	-0.050	-0.022	-0.036		0.096	-0.016	0.534	0.232	
		RMSE	0.069	0.074	0.098	0.109		0.339	0.064	0.553	0.327	
	NLLS	Mean						1.456	0.928	2.932		0.747
		Bias						0.456	-0.072	-0.068		-0.003
		RMSE						0.514	0.090	0.149		0.037

Table 2: Spatial/Network Treatment Selection: Results for N={250;500}, T=3, M=1000 (*cont.*)

			$\tilde{\beta}_1^A$	$\tilde{\beta}_2^A$	$\tilde{\delta}_1^A$	$\tilde{\delta}_2^A$	ρ^A	α	β_1^B	δ_1^B	τ	ρ^B
Panel i		True	0.707	0.707	0.707	0.707	0.75	1	1	3	0.707	0.75
N=250	SNTS	Mean	0.736	0.739	0.738	0.728	0.735	1.013	1.003	3.002	0.683	0.745
		Bias	0.029	0.032	0.030	0.021	-0.015	0.013	0.003	0.002	-0.024	-0.005
		RMSE	0.102	0.095	0.206	0.203	0.056	0.411	0.086	0.171	0.235	0.060
	WTS	Mean	0.509	0.533	0.587	0.643		0.555	1.060	3.314	1.334	
		Bias	-0.198	-0.174	-0.121	-0.065		-0.445	0.060	0.314	0.627	
		RMSE	0.211	0.188	0.174	0.140		0.609	0.105	0.360	0.706	
	NLLS	Mean						1.485	0.941	2.914		0.763
		Bias						0.485	-0.059	-0.086		0.013
		RMSE						0.574	0.099	0.193		0.062
N=500	SNTS	Mean	0.721	0.722	0.714	0.707	0.746	1.007	1.000	3.000	0.690	0.747
		Bias	0.014	0.015	0.007	0.000	-0.004	0.007	0.000	0.000	-0.017	-0.003
		RMSE	0.061	0.067	0.147	0.150	0.036	0.319	0.059	0.136	0.169	0.034
	WTS	Mean	0.528	0.521	0.603	0.580		1.069	0.989	3.540	1.087	
		Bias	-0.179	-0.186	-0.104	-0.127		0.069	-0.011	0.540	0.379	
		RMSE	0.186	0.193	0.137	0.159		0.328	0.059	0.558	0.453	
	NLLS	Mean						1.499	0.929	2.934		0.744
		Bias						0.499	-0.071	-0.066		-0.006
		RMSE						0.551	0.087	0.145		0.038
Panel j		True	0.707	0.707	0.707	0.707	0	1	1	3	0	0
N=250	SNTS	Mean	0.739	0.741	0.744	0.752	-0.123	0.998	0.999	3.004	-0.005	0.003
		Bias	0.032	0.034	0.037	0.045	-0.123	-0.003	-0.001	0.004	-0.005	0.003
		RMSE	0.093	0.084	0.197	0.184	0.248	0.120	0.052	0.130	0.145	0.112
	WTS	Mean	0.715	0.717	0.719	0.721		0.996	0.999	3.009	0.004	
		Bias	0.007	0.010	0.012	0.014		-0.004	-0.001	0.009	0.004	
		RMSE	0.083	0.073	0.147	0.144		0.120	0.052	0.130	0.114	
	NLLS	Mean						0.999	0.998	3.004		-0.005
		Bias						-0.001	-0.002	0.004		-0.005
		RMSE						0.099	0.051	0.130		0.145
N=500	SNTS	Mean	0.721	0.720	0.726	0.723	-0.064	0.999	0.999	2.998	-0.002	0.001
		Bias	0.014	0.013	0.019	0.016	-0.064	-0.001	-0.001	-0.002	-0.002	0.001
		RMSE	0.056	0.061	0.133	0.135	0.175	0.091	0.034	0.098	0.091	0.079
	WTS	Mean	0.708	0.709	0.711	0.709		1.000	0.999	3.001	0.001	
		Bias	0.001	0.001	0.004	0.002		0.000	-0.001	0.001	0.001	
		RMSE	0.053	0.057	0.103	0.104		0.091	0.034	0.097	0.079	
	NLLS	Mean						1.000	0.999	2.998		-0.002
		Bias						0.000	-0.001	-0.002		-0.002
		RMSE						0.073	0.032	0.097		0.091

The main insights from an inspection of Tables 1 and 2 can be summarized as follows. First of all, whenever $\rho^A \neq 0$ and/or $\rho^B \neq 0$, neither WSS (WTS) nor NLLS should be used. The bias of these estimators tends to increase with the magnitude of absolute deviations of $(\rho^A, \rho^B))$ from zero. To see this, e.g., compare the results in Panel c with those of Panel b of Table 1, or those of Panel i with the ones of Panel g in the same table.

Second, as the sample from which the parameters are estimated are bigger for treatment selection than for sample selection by the chosen design, these biases tend to be more pronounced for the case of treatment selection in Table 2 than for the case of sample selection in Table 1 when comparing identically-labelled panels.

In Panel j, all three estimators are inefficient, as parameters are estimated that could have been restricted to be zero: in the SNSS (SNTS) estimators, $(\rho^A \neq 0, \rho^B \neq 0)$ and $\tau \neq 0$, in the WSS (WTS) model, $(\rho^A = 0, \rho^B = 0)$ and $\tau \neq 0$, and in the NLLS, $(\rho^A \neq 0, \rho^B \neq 0), \tau = 0$.

None of the estimators clearly outperforms the others for this configuration. Note however, the SNSS (SNTS) model also performs at an acceptable rate in the cases where it is inefficient.

Third, overall, the performance of the proposed SNSS (SNTS) estimator increases as we increase the sample size. To see this, compare the lower block in a panel with the corresponding upper one in each one of the tables. I.e., the bias in $\hat{\tau}$ drops by almost one-quarter in Panel i of Table 1 for the SNSS estimator, and it drops by more than one-quarter in Panel i of Table 2 for the SNTS estimator.

5 Variance-covariance Matrix of Parameters

In this section we derive the variance-covariance matrix of the two parameter vectors θ^A and θ^B. We pay special attention to the derivation of the variance-covariance matrix of the parameters of the outcome equation. Two issues have to be taken into account:

(1) We use estimated parameters $\hat{\theta}_A$ for the estimation of $\hat{\theta}_B$, as both Spatial Adjustment/Network Factor and Inverse Mills' Ratio are functions of parameters of the first stage. In order to address this problem, we use a Murphy & Topel (1985, 2002) type of correction, which we adapt to our estimation procedure (pooled Bayesian Spatial Probit and pooled Non-Linear Least Squares) following steps outlined in Greene (2008). Note that for the Murphy & Topel correction ψ_{ti} and λ_{ti} (λ_{ti}^g) have to be twice continuously differentiable in θ^A. It further requires a consistently

estimated vector of first-stage parameters $\hat{\theta}^A$, which is asymptotically normal and for which a consistent Variance-Covariance matrix estimate $\hat{\Omega}_{\theta^A}$ exists. We therefore make use of the fact that the Bayesian posterior distribution of $\hat{\theta}^A$ is asymptotically equivalent to the MLE $\hat{\theta}^A_{MLE}$ (LeSage & Pace, 2009).[11]

(2) The corrected variance-covariance matrix is a function of the truncated variance and truncated covariance of the spatial error components of the outcome equation. We derive these and outline an estimation procedure along the lines of Heckman (1979).

5.1 Variance-covariance Matrix of θ^A

5.1.1 Analytical Variance-covariance Matrix

Using the familiar result regarding the asymptotic distribution of the MLE, the asymptotic distribution of first-stage parameters θ^A is given by:

$$\begin{aligned} \sqrt{TN}(\hat{\theta}^A_{MLE} - \theta^A) &\overset{d}{\to} N\left(0, \Omega_{\theta^A}\right) \\ \text{with} \quad \Omega_{\theta^A} &= TN\mathcal{I}(\theta^A)^{-1}. \end{aligned}$$

$\mathcal{I}(\theta^A) = -E\left[\frac{\partial^2 l}{\partial\theta^A \partial\theta^{A\prime}}\right]$ denotes the Information Matrix. To derive the Information Matrix, we need to compute the Hessian of l, which is the log-likelihood function of the first stage:

$$\begin{aligned} l(\theta^A, \Omega_{\xi^A}) &= K - \frac{1}{2}ln|\Omega_{\xi^A}| + ln|(R^A)^{-1}| - \frac{1}{2}\nu^{A\prime}\nu^A, \\ \text{where} \quad \nu^A &= \Omega_{\xi^A}^{-\frac{1}{2}}(R^A)^{-1}\left[y^{A*} - x^A\beta^A - R^A(\iota_T \otimes \bar{x}^A)\delta^A\right] \end{aligned}$$

and where K is a constant independent of parameters.[12] The elements of the Information Matrix and their derivation can be found in Appendix 2.

[11] In order to perform inference, the Bayesian approach simply uses the standard deviation of the MCMC-draws of parameters as an estimate of the standard error. This is why, in order to derive the analytical second-stage Variance-covariance matrix, we use the asymptotic variance-covariance matrix of the MLE as they are asymptotically equivalent.

[12] The Bayesian Spatial/Network Error Probit Model first samples the vector of parameters y^{A*} before sampling θ^A from the same conditional marginal distributions as in the linear Bayesian spatial/network error model. To find the analytical variance-covariance matrix of θ^A we therefore derive the Information Matrix of this linear model to simplify the calculations. Derivations of the Information Matrix follow steps outlined in Anselin (1988) and can be found in Appendix 2.

5.1.2 Estimation

Since we pool the data to estimate the first-stage parameters, we cannot simply use the standard deviation of MCMC draws (minus Burn-In) to obtain a consistent estimate of the standard error of θ^A. However, we can replace true parameters θ^A by their estimated counterparts $\hat{\theta}^A$ in the information matrix, which we have derived above.

To compute the estimated matrix of error components $\widehat{\Omega}_{\xi^A} = (J_T \otimes \hat{\sigma}^2_{\mu^A} I_N) + \hat{\sigma}^2_{\varepsilon^A}(I_T \otimes I_N)$, we estimate the variances of the error components $\hat{\sigma}^2_{\mu^A}$ and $\hat{\sigma}^2_{\varepsilon^A}$ as follows:

First we derive vector of estimated error components $\hat{\xi}^A$ by pre-multiplying the vector of residuals from the first stage $\hat{u}^A = \hat{y}^{A*} - (x^{A\prime}\beta^A + R^A\bar{x}^{A\prime}\delta^A)$ by the inverse of the estimated matrix of weights $(I_T \otimes \hat{R}^A_t)^{-1}$.

$$\hat{\xi}^A = (I_T \otimes \hat{R}^A_t)^{-1}\hat{u}^A$$

Then, following Baltagi (2005):

$$\hat{\sigma}^2_{\varepsilon^A} = \frac{\hat{\xi}^{A\prime}Q\hat{\xi}^A}{tr(Q)}, \quad \hat{\sigma}^2_1 = \frac{\hat{\xi}^{A\prime}P\hat{\xi}^A}{tr(P)}, \quad \hat{\sigma}^2_{\mu^A} = \frac{\hat{\sigma}^2_1 - \hat{\sigma}^2_{\varepsilon^A}}{T}$$

with $Q = \left(I_T - \frac{J_T}{T}\right) \otimes I_N$ and $P = \frac{J_T}{T} \otimes I_N$.

5.2 Variance-covariance Matrix of θ^B

Before we derive the analytical variance-covariance matrix and outline the estimation procedure, we need to introduce some further notation. First, we define the score vector of the log-likelihood function of the first stage as

$$\begin{aligned} \frac{\partial l}{\partial \theta^A} &= g^A(\theta^A), \\ \bar{g}^B(\theta^A) &= \frac{1}{TN} g^A(\theta^A). \end{aligned}$$

Second, let ν^B_{ti} denote the residual of the augmented outcome equation.

Spatial/Network Sample Selection

$$\nu^B_{ti} = y^B_{ti} - (x^{B\prime}_{ti}\beta^B + \sum_{j=1}^{N} r^B_{tij}\bar{x}^{B\prime}_j\delta^B + \tau\psi_{ti}\lambda_{ti})$$

Spatial/Network Treatment Selection

$$\nu^B_{ti} = y^B_{ti} - (\alpha y^A_{ti} - x^{B\prime}_{ti}\beta^B + \sum_{j=1}^{N} r^B_{tij}\bar{x}^{B\prime}_j\delta^B + \tau\psi_{ti}\lambda^g_{ti})$$

Further, define the sum of squared residuals of the augmented outcome equation as

$$q = \nu^{B\prime}\nu^{B} = \sum_{t=1}^{T}\sum_{i=1}^{N}(\nu_{ti}^{B})^2.$$

Abbreviating $q = q(x^{A0}, x^B, \theta^A, \theta^B)$, the score vector (gradient) of the sum of squared residuals is given by

$$\begin{aligned}
\frac{\partial q}{\partial \theta^B} &= \sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial(\nu_{ti}^{B})^2}{\partial\theta^B} = g^B(\theta^A, \theta^B), \\
\bar{g}^B(\theta^A, \theta^B) &= \frac{1}{TN} g^B(\theta^A, \theta^B)
\end{aligned}$$

We define the Hessian of the sum of squared residuals with respect to θ^B and θ^B as

$$\begin{aligned}
\frac{\partial^2 q}{\partial \theta^B \partial\theta^{B\prime}} &= \sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial^2(\nu_{ti}^{B})^2}{\partial\theta^B\partial\theta^{B\prime}} = H^B(\theta^A, \theta^B), \\
\bar{H}^B(\theta^A, \theta^B) &= \frac{1}{TN} H^B(\theta^A, \theta^B)
\end{aligned}$$

and the Hessian of the sum of squared residuals with respect to θ^B and θ^A as

$$\begin{aligned}
\frac{\partial^2 q}{\partial \theta^B \partial\theta^{A\prime}} &= \sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial^2(\nu_{ti}^{B})^2}{\partial\theta^B\partial\theta^{A\prime}} = H^A(\theta^A, \theta^B), \\
\bar{H}^A(\theta^A, \theta^B) &= \frac{1}{TN} H^A(\theta^A, \theta^B)
\end{aligned}$$

5.2.1 Murphy-Topel Correction

To account for the fact that estimated first-stage parameters θ^A are being used in the estimation of second-stage parameters θ^B, we adjust the variance-covariance matrix using well-known results regarding two-step estimators (Murphy & Topel, 1985, 2002). The asymptotic distribution of second-stage parameters is given by (see Appendix 3.1 for proof):

$$\sqrt{TN}(\hat{\theta}^B - \theta^B) \xrightarrow{d} N\left(0, \Omega_{\theta^B}\right)$$

where

$$\begin{aligned}
\Omega_{\theta^B} = {} & \underbrace{E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1}\Omega_{g^B} E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1}}_{\text{Standard NLLS VC Matrix}} \\
&+ E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} E\left[\bar{H}^A(\theta^A, \theta^B)\right]\Omega_{\theta^A} E\left[\bar{H}^A(\theta^A, \theta^B)\right]' E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} \\
&+ E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1}\Omega_{g^{BA}}\Omega_{\theta^A} E\left[\bar{H}^A(\theta^A, \theta^B)\right]' E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} \\
&+ E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} E\left[\bar{H}^A(\theta^A, \theta^B)\right]\Omega_{\theta^A}\Omega_{g^{AB}} E\left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1}
\end{aligned}$$

using $\Omega_{g^B} = Var\left[\sqrt{TN}\bar{g}^B(\theta^A, \theta^B)\right]$ and $\Omega_{g^{BA}} = Cov\left[\sqrt{TN}\bar{g}^B(\theta^A, \theta^B), \sqrt{TN}\bar{g}^A(\theta^A)\right]$.

5.2.2 Spatial/Network Truncated Variance-covariance Matrix

The elements of matrix Ω_{g^B} are functions of the following conditional expectations (see Appendix 3.2.2 for the specific elements of this matrix and Appendix 3.3.1–3.3.3 for proofs):

$$\begin{aligned} E\left[(\nu^B_{ti})^2|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2-\tau^2\psi^2_{ti}\zeta_{ti} \\ E\left[\nu^B_{ti}\nu^B_{tj}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma^2_{\xi^B}\sum_{m=1}^{N}r^B_{tim}r^B_{tjm}-\tau^2\psi_{ti}\frac{\sum_{m=1}^N r^A_{tim}r^B_{tjm}}{\sqrt{\sum_{j=1}^N\left(r^A_{tij}\right)^2}}\zeta_{ti} \end{aligned} \quad (13)$$

$$\begin{aligned} E\left[\nu^B_{ti}\nu^B_{si}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma^2_{\mu^B}\sum_{j=1}^{N}r^B_{tij}r^B_{sij}-\tau\psi_{ti}\frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sum_{j=1}^N r^A_{tij}r^B_{sij}}{\sqrt{\sum_{j=1}^N\left(r^A_{tij}\right)^2}}\zeta_{ti} \\ E\left[\nu^B_{ti}\nu^B_{sj}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma^2_{\mu^B}\sum_{m=1}^{N}r^B_{tim}r^B_{sjm}-\tau\psi_{ti}\frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sum_{m=1}^N r^A_{tim}r^B_{sjm}}{\sqrt{\sum_{j=1}^N\left(r^A_{tij}\right)^2}}\zeta_{ti} \end{aligned} \quad (14)$$

with $\zeta_{ti}=\lambda^2_{ti}+\frac{x^{A\prime}_{ti}\beta^A+\sum_{j=1}^N r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N(r^A_{tij})^2}}\lambda_{ti}$ and $\varrho^{AB}=\frac{Cov\left[u^A_{ti},u^B_{sj}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}}$.

Similarly, the elements of matrix $\Omega_{g^{AB}}$ are functions of the following conditional expectations:

$$\begin{aligned} E\left[u^A_{ti}\nu^B_{ti}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma_{\xi^{AB}}\sum_{j=1}^{N}r^A_{tij}r^B_{tij}\left[1-\zeta_{ti}\right] \\ E\left[u^A_{ti}\nu^B_{tj}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma_{\xi^{AB}}\sum_{m=1}^{N}r^A_{tim}r^B_{tjm}\left[1-\zeta_{ti}\right] \\ E\left[u^A_{ti}\nu^B_{si}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma_{\mu^{AB}}\sum_{j=1}^{N}r^A_{tij}r^B_{sij}\left[1-\zeta_{ti}\right] \\ E\left[u^A_{ti}\nu^B_{sj}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= \sigma_{\mu^{AB}}\sum_{m=1}^{N}r^A_{tim}r^B_{sjm}\left[1-\zeta_{ti}\right] \end{aligned} \quad (15)$$

5.2.3 Estimation

In order to compute $\hat{\Omega}_{\theta^B}$, we replace Ω_{θ^A} with $\hat{\Omega}_{\theta^A}$. The remaining elements can be obtained by first deriving the Hessian matrices $H^A(\theta^A,\theta^B)$ and $H^B(\theta^A,\theta^B)$. Then we substitute the estimated parameters $\hat{\theta}^A$ and $\hat{\theta}^B$ for the true parameters θ^A and θ^B.

However, the variances and covariances of the error components $\sigma^2_{\xi^B}$, and $\sigma_{\xi^{AB}}$ as well as the variances and covariances of the individual-specific time-invariant component $\sigma^2_{\mu^B}$, and $\sigma_{\mu^{AB}}$ are unknown to us. We can estimate them using the following formulas (see Appendix 3.3.4 for proofs, which closely follow Heckman (1979)):

$$\begin{aligned}
\hat{\sigma}^2_{\xi^B} &= \frac{1}{\widehat{\overline{\sum_{j=1}^N (r_j^B)^2}}} \left(\frac{\sum_{t=1}^T \sum_{j=1}^N (\hat{\nu}_{ti}^B)^2}{TN} + \hat{\tau}^2 \widehat{\overline{\psi^2 \zeta}} \right) \\
\hat{\sigma}_{\xi^{BA}} &= \frac{\hat{\tau}}{\sqrt{\hat{\sigma}^2_{\xi^B}}} \\
\hat{\sigma}_{\mu^{AB}} &= \frac{\sum_{s<t} \sum_{i=1}^N \hat{u}_{ti}^A \hat{\nu}_{si}^B}{TN} \left[\widehat{\overline{\sum_{j=1}^N r_j^A r_j^B}} - \widehat{\overline{\sum_{j=1}^N r_j^A r_j^B \zeta}} \right]^{-1} \\
\hat{\sigma}^2_{\mu^B} &= \frac{1}{\widehat{\overline{\sum_{j=1}^N r_j^B r_j^B}}} \left[\frac{\sum_{s<t} \sum_{i=1}^N \hat{\nu}_{ti}^B \hat{\nu}_{si}^B}{TN} + \hat{\tau} \frac{\hat{\sigma}_{\mu^{AB}}}{\sqrt{\hat{\sigma}^2_{\xi^A}}} \widehat{\overline{\psi \frac{\sum_{j=1}^N r_j^A r_j^B}{\sqrt{\sum_{j=1}^N \left(r_j^A\right)^2}} \zeta}} \right],
\end{aligned}$$

where bars indicate averages across both time and individuals.

6 Application: Export-wage Premium Among Firms in Shenzhen, P.R. China (*Future Work*)

We intend to apply the spatial/network treatment selection model to study the relevance of self-selection of firms into exporting (treatment) when analyzing the size of the exporter-wage premium. In order to do so we use Chinese firm-level panel-data from the Chinese Annual Survey of Industrial Firms Database (CASIF) on firms in the city of Shenzhen, P.R. China, for the years 1999-2009. The data contains accounting information of all state-owned enterprises (SOEs) as well as all non-SOEs with sales above 500 Mio RMB per year. Moreover, they include information on firm addresses and industry affiliation, permitting a spatial/network analysis. The number of firms vary by year due to entry/exit (i.e. the spatial/network weights change across years).

The exporter-wage premium, i.e. the fact that exporting firms pay higher wages per worker than non-exporters, has been documented in many data-sets for various countries, always assuming that firms' selection into exporting was independent of other firms (and often even random), and that wages were set independently as well (see Klein, Moser, and Urban, 2013; Egger, Egger, and Kreickemeier, 2013; Egger, Egger, Kreickemeier, and Moser, 2017).

However, random assignment of exporter status seems unlikely, given that it appears that only firms above a certain profitability threshold export and that more profitable firms are more likely to pay higher wages. We will compute log wages per employee y_{ti}^{B} in firm i at time t in the city of Shenzhen, P.R. China. These are likely to depend on covariates x_{ti}^{B}, such as firm productivity (e.g., captured by relative domestic sales of total domestic sales in a Melitz-style model) as well as binary exporting y_{ti}^{A} (treatment indicator).

Some firms are exporters and others are not. The exporting decision is assumed to be determined by latent export profitability y_{ti}^{A*}. The latter depends on theoretically motivated factors x_{ti}^{B}: factor costs, market potential, location, etc. of firm i at t. Shocks to export profitability have been found to spatially dissipate across firms due to, e.g. industry networks or input-ouput linkages (see Antràs, Fort, and, Tintelnot, 2017; Baltagi, Egger, and Kesina, 2017; Tintelnot, 2017; Chaney, 2014). Shocks to the average wage at a firm might follow a spatial pattern and might therefore be correlated across firms due to local labor markets and worker flows (see Moretti, 2011).

7 Conclusions

Missing data generate difficulties in case that the units of observation are not missing at random, in particular, if the data are not independent of each other. A rapidly growing literature in the social sciences at large and in economics in particular is concerned with understanding the effects of networks on outcomes, often on outcomes which are repeatedly observed over time. However, often the data are generated from surveys, or they are incomplete due to confidentiality clauses, legal (e.g., size) thresholds imposed for data delivery requirements, etc. Similarly, self-selection of households or firms into certain states such as unemployment and exporting, respectively, entails that units are systematically unobserved in the counterfactual state. Associated panel-data situations in the presence of networks cannot be analyzed with existing methods.

The purpose of the present paper was to outline econometric models which are suited for problems of selection (and truncation) with panel data, where the units of observation depend on each other through potentially time-variant network structures. We derived two-step, control-function procedures for the cases of both sample selection as well as treatment selection. We reported on a set of Monte Carlo simulations to show that the proposed models work well for both point as well as variance-covariance-matrix estimation of the parameters of interest.

Finally, we have described an application, for which we expect our model to be crucial: the analysis of the role of interdependent export decisions for wage premia in Chinese firms in the city of Shenzhen. In those data, we anticipate that the decision to export or not is interdependently made among firms in close-by neighborhoods, and wages appear to have a spatial component as well. No competing model could be used for the analysis of the respective data, as such models do either not feature interdependence, or they cannot be used for panel data.

8 References

Anselin, L., Bera, A.K. , 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economics Statistics. Marcel Dekker, New York, pp. 237–289.

Antràs, P., Fort, T.C., Tintelnot, F., 2017. The Margins of Global Sourcing: Theory and Evidence from US Firms. American Economic Review, 107 (9), 2514–2564.

Arulampalam, W., 1999. A Note on Estimated Coefficients in Random Effects Probit Models. Oxford Bulletin of Economics and Statistics 61 (4), 597-602.

Audretsch, D., Feldman, M.P., 1996. R&D Spillovers and the Geography of Innovation and Production. American Economic Review 86 (3), 630–40.

Bahar, D., Hausmann, R., Hidalgo, C.A., 2014. Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion?. Journal of International Economics 92, 111-123.

Ballester, C., Calvó-Armengol, A., Zenou, Y., 2006. Who's Who in Networks. Wanted: The Key Player. Econometrica 74 (5), 1403–1417.]

Baltagi, B.H., 2005. Econometric Analysis of Panel Data. John Wiley & Sons Ltd, New York.

Baltagi, B.H., Egger, P.H., Kesina, M., 2017. Determinants of firm-level domestic sales and exports with spillovers: Evidence from China. Journal of Econometrics 199 (2), 184–201, 2017.

Baltagi, B.H., Egger, P., Pfaffermayr, M., 2007. Estimating models of complex FDI: Are there third-country effects?. Journal of Econometrics 140, 260–281.

——, 2008. Estimating regional trade agreement effects on FDI in an interdependent world. Journal of Econometrics 145 (1-2), 194-208.

Behrens, K., Ertur, C., Koch, W., 2012. Dual gravity: Using spatial econometrics to control for multilateral resistance. Journal of Applied Econometrics 27 (5), 773–794.

Blonigen, B., Davies, R., Waddell, G., Naughton, H., 2007. FDI in space: Spatial autoregressive relationships in foreign direct investment. European Economic Review 51 (5), 1303–1325.

Blume, L.E., Brock, W.A., Durlauf, S.N., Jayaraman, R., 2015. Linear Social Interaction Models. Journal of Political Economy 123 (2), 444–496.

Calabrese, R., Elkink, J.A., 2014. Estimators of binary spatial autoregressive models: a Monte Carlo study. Journal of Regional Science 54 (4), 664-687.

Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer Effects and Social Networks in Education. Review of Economic Studies 76 (4), 1239–1267.

Case, A., Hines, J., Rosen, H., 1993. Budget spillovers and fiscal policy interdependence: Evidence from the states. Journal of Public Economics 52 (3), 285–307.

Chamberlain, G., 1982. Multivariate regression models for panel data. Journal of Econometrics 18 (1), 5–46.

Chaney, T. 2014. The Network Structure of International Trade. American Economic Review 104 (11), 36003634.

Cohen-Cole, E., Liu, X., Zenou, Y., 2014. Multivariate choices and identification of social interactions. Journal of Applied Econometrics, forthcoming.

Conley, T., Ligon, E., 2002. Economic Distance and Cross-Country Spillovers. Journal of Economic Growth 7 (2), 157–187.

Devereux, M.P., Lockwood, B., Redoano, M., 2008. Do countries compete over corporate tax rates?. Journal of Public Economics 92, 1210–1235.

Doğan, O., Taşpinar, S., 2017. Bayesian Inference in Spatial Sample Selection Models. Oxford Bulleting of Economics and Statistics 80, 90–121.

Egger, H., Egger, P.H., Kreickemeier, U., 2013. Trade, wages, and profits. European Economic Review 64 (C), 332–350.

Egger, H., Egger, P., Kreickemeier, U., Moser, C., 2017. The exporter wage premium when firms and workers are heterogeneous, CEPIE Working Paper, No. 12/17, http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-227402

Egger, P.H., Pfaffermayr, M., 2016. A Generalized Spatial Error Components Model for Gravity Equations. Empirical Economics 50 (1), 177–195.

Egger, P.H., Pfaffermayr, M., Winner, H., 2005. An unbalanced spatial panel data approach to US state tax competition. Economics Letters 88 (3), 329–335.

Ertur, C., Koch, W., 2007. Growth, technological interdependence and spatial externalities: theory and evidence. Journal of Applied Econometrics 22 (6), 1033–1062.

Fleming, M.M., 2004. Techniques for Estimating Spatially Dependent Discrete Choice Models. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics. Methodology, Tools and Applications. Springer, Berlin, pp. 145-167.

Flores-Lagunes, A., Schnier, K.E., 2012. Estimation of Sample Selection Models with Spatial Dependence. Journal of Applied Econometrics 27, 173–204.

Geweke, J., 1991. Efficient simulation from the multivariate normal and Student t distributions subject to linear constraints. In: Keramidas, E. (Ed.), Computer Science and Statistics Proceedings of the Twenty-Third Symposium on the Interface. pp. 571-578.

Greene, W.H., 1995. Count Data. Manuscript, Department of Economics, Stern School of Business, New York University.

—— 2008. Econometric Analysis. Pearson Prentice Hall, New Jersey.

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

——, 1979. Sample Selection Bias as a Specification Error. Econometrica 47 (1), 153–61.

Ioannides, Y. M., Loury, D. L., 2004. Job information networks, neighborhood effects, and inequality. Journal of Economic Literature 42, 1056–1093.

Kamal, F., Sundaram, A., 2016. Buyer-seller relationships in international trade: Do your neighbors matter?. Journal of International Economics 102, 128–140.

Kapoor, M, Kelejian, H.H. Prucha, I.R., 2007. Panel Data Models with Spatially Correlated Error Components. Journal of Econometrics 140, 97–130.

Kelejian, H.H., Prucha, I.R., 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. Journal of Real Estate Finance and Economics 17, 99–121.

——, 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

———, 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157 (1), 53–67.

Klein, M.W., Moser, C., Urban, D.M., 2013. Exporting, skills and wage inequality. Labour Economics 25, 76–85.

Kotz, S., Balakrishnan, N., Johnson, N.L., 2005. Continuous Multivariate Distributions: Models and Applications. John Wiley & Sons Ltd, New York.

Holtz-Eakin, D., 1994. Public-Sector Capital and the Productivity Puzzle. The Review of Economics and Statistics 76 (1), 12–21.

Lee, L.-F., 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica 72 (6), 1899–1925.

——, 2002. Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models. Econometric Theory 18, 252277.

——, 2007a. GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models. Journal of Econometrics 137. 489–514.

——, 2007b. Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects, Journal or Econometrics 140, 333–374.]

LeSage, J.P., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models. Geographical Analysis 32 (1), 19-35.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman and Hall/CRC, Boca Raton.

Martin, P., T. Mayer, Mayneris, F., 2011. Spatial Concentration and Plant-Level Productivity in France. Journal of Urban Economics, 69 (2), 182–195.

McMillen, D.P., 1995. Selection Models in Spatial Econometric Models. Journal of Regional Science 35 (3), 417–436.

Moretti, E., 2011. Local Labor Markets. In: Ashenfelter, O. and Card, D. (Eds.), Handbook of Labor Economics. Elsevier, Amsterdam, pp. 1237-1313.

Murphy, K., Topel, R., 1985. Estimation and Inference in Two Step Econometric Models. Journal of Business and Economic Statistics 3, 370–379. Reprinted, 2002, Journal of Business and Economic Statistics 20, 88–97.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69–85.

Nijman, T., Verbeek, M., 1992. Nonresponse in panel data: the impact on estimates of a life cycle consumption function. Journal of Applied Econometrics 7, 243–257.

Patacchini, E., Zenou, Y., 2012. Juvenile delinquency and conformism. Journal of Law, Economics, and Organization 28, 1–31.

Pinkse, J., Slade, M., 1998. Contracting in space: An application of spatial statistics to discrete-choice models. Journal of Econometrics 85 (1), 125–154.

Pinkse, J., Slade, M.E., Brett, C., 2002. Spatial price competition: A semiparametric approach. Econometrica 70 (3), 1111–1153.

Semykina, A., Wooldridge, J.M., forthcoming. Binary response panel data models with sample selectionand self-selection. Journal of Applied Econometrics.

Vella, F., 1998. Estimating Models with Sample Selection Bias: A Survey. Journal of Human Resources 33 (1), 127–169.

Vella F., Verbeek, M., 1999. Two-step estimation of panel data models with censored endogenous variables and selection bias. Journal of Econometrics 90 (2), 239–263.

Wooldridge, J., 1995. Selection corrections for panel data models under conditional mean independence assumptions. Journal of Econometrics 68 (1), 115–132.

Tintelnot, F., 2017. Global Production with Export Platforms. Quarterly Journal of Economics, 132 (1), 157–209.

Appendix

1 Deriving the Correction Functions

Spatial Sample Selection: The Spatially Adjusted Inverse Mills' Ratio

$$
\begin{aligned}
& E[u^B_{ti}|y^A_{ti}=1,x^{A0},x^B] \\
= \quad & E[u^B_{ti}|y^{A*}_{ti}>0,x^{A0},x^B] \\
= \quad & E[u^B_{ti}|u^A_{ti}>-x^{A\prime}_{ti}\beta^A-\sum_{j=1}^{N} r^A_{tij}\bar{x}^{A\prime}_j\delta^A,x^{A0},x^B] \\
= \quad & \frac{Cov(u^B_{ti},u^A_{ti})}{V(u^A_{ti})}E\left[u^A_{ti}|u^A_{ti}>-x^{A\prime}_{ti}\beta^A-\sum_{j=1}^{N} r^A_{tij}\bar{x}^{A\prime}_j\delta^A,x^{A0},x^B\right] \\
= \quad & \frac{Cov(u^B_{ti},u^A_{ti})}{\sqrt{V(u^A_{ti})}}E\left[\frac{u^A_{ti}}{\sqrt{V(u^A_{ti})}}|\frac{u^A_{ti}}{\sqrt{V(u^A_{ti})}}>\frac{-x^{A\prime}_{ti}\beta^A-\sum_{j=1}^{N} r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{V(u^A_{ti})}},x^{A0},x^B\right] \\
= \quad & \frac{\sigma_{\xi^{BA}}\sum_{j=1}^{N} r^B_{tij}r^A_{tij}}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\frac{\phi\left(\frac{x^A_{ti}\beta^A+\sum_{j=1}^{N} r^A_{tij}\bar{x}^A_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)}{\Phi\left(\frac{x^{A\prime}_{ti}\beta^A+\sum_{j=1}^{N} r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)} \\
= \quad & \tau\psi_{ti}\lambda_{ti}, \quad \text{where we have defined}
\end{aligned}
$$

$$
\tau = \frac{\sigma_{\xi^{BA}}}{\sqrt{\sigma^2_{\xi^A}}} \quad , \quad \psi_{ti}=\frac{\sum_{j=1}^{N} r^B_{tij}r^A_{tij}}{\sum_j^N (r^A_{tij})^2} \quad , \quad \lambda_{ti}=\frac{\phi\left(\frac{x^A_{ti}\beta^A+\sum_{j=1}^{N} r^A_{tij}\bar{x}^A_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)}{\Phi\left(\frac{x^{A\prime}_{ti}\beta^A+\sum_{j=1}^{N} r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)},
$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the Standard Normal PDF and Standard Normal CDF respectively.

Treatment selection: The generalized inverse Mills ratio

$$\begin{aligned}
& E[u^B_{ti}|y^A_{ti}, x^{A0}, x^B] \\
= \;& y^A_{ti} E[u^B_{ti}|y^{A*}_{ti} > 0, x^{A0}, x^B] + (1-y^A_{ti})E[u^B_{ti}|y^{A*}_{ti} \leq 0, x^{A0}, x^B] \\
= \;& y^A E[u^B_{ti}|u^A_{ti} > -x^A_{ti}\beta^A - \sum_{j=1}^{N} r^A_{tij}\bar{x}^A_j \delta^A, x^{A0}, x^B] + \\
+ \;& (1-y^A_{ti})E[u^B_{ti}|u^A_{ti} \leq -x^A_{ti}\beta^A - \sum_{j=1}^{N} r^A_{tij}\bar{x}^A_j \delta^A, x^{A0}, x^B] \\
= \;& y^A_{ti}\frac{Cov(u^B_{ti}, u^A_{ti})}{V(u^A_{ti})} E\left[u^A_{ti}|u^A_{ti} > -x^{A\prime}_{ti}\beta^A - \sum_{j=1}^{N} r^A_{tij}\bar{x}^{A\prime}_j \delta^A, x^{A0}, x^B\right] + \\
+ \;& (1-y^A_{ti})\frac{Cov(u^B_{ti}, u^A_{ti})}{V(u^A_{ti})} E\left[u^A_{ti}|u^A_{ti} \leq -x^{A\prime}_{ti}\beta^A - \sum_{j=1}^{N} r^A_{tij}\bar{x}^{\prime}_j \delta^A, x^{A0}, x^B\right] \\
= \;& \frac{\sigma_{\xi^{BA}}\sum_{j=1}^{N} r^B_{tij} r^A_{tij}}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}} \frac{y^A_{ti} - \Phi\left(\frac{x^{A\prime}_{ti}\beta^A + \sum_{j=1}^N r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)}{\Phi\left(\frac{x^{A\prime}_{ti}\beta^A + \sum_{j=1}^N r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)[1-\Phi\left(\frac{x^{A\prime}_{ti}\beta^A + \sum_{j=1}^N r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)]} \\
= \;& \tau\psi_{ti}\lambda^g_{ti}, \quad \text{where}
\end{aligned}$$

$$\lambda^g_{ti} = \frac{y^A_{ti} - \Phi\left(\frac{x^{A\prime}_{ti}\beta^A + \sum_{j=1}^N r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)}{\Phi\left(\frac{x^{A\prime}_{ti}\beta^A + \sum_{j=1}^N r^A_{tij}\bar{x}^{A\prime}_j\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)[1-\Phi\left(\frac{x^{A\prime}_{ti}\beta^A + \sum_{j=1}^N r^A_{tij}\bar{x}^{A\prime}_j}{\sqrt{\sigma^2_{\xi^A}\sum_j^N (r^A_{tij})^2}}\right)]}$$

2 Deriving the Asymptotic Distribution of θ^A

The likelihood of the RE SEM model is:

$$\begin{aligned} & L(\theta^A, \Omega_{\xi^A}) \\ = \quad & \frac{1}{2\pi^{TN/2}} |\Omega_{u^A}|^{-\frac{1}{2}} exp \left\{ -\frac{1}{2} \underbrace{\left[y^{A*} - x^A\beta^A - (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right]' \Omega_{u^A}^{-1} \left[y^{A*} - x^A\beta^A - (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right]}_{=\nu^{A\prime}\nu^A} \right\}, \end{aligned}$$

with $\underset{TN\times TN}{\Omega_{u^A}} = (\underset{T\times T}{I_T} \otimes \underset{N\times N}{R_t^A}) \underset{TN\times TN}{\Omega_{\xi^A}} (I_T \otimes R_t^A)' = (I_T \otimes R_t^A)[(J_T \otimes \sigma^2_{\mu^A} I_N) + \sigma^2_{\varepsilon^A}(I_T \otimes I_N)](I_T \otimes R_t^A)'$

and $\underset{TN\times 1}{\nu}{}^A = \underset{TN\times TN}{\Omega_{\xi^A}^{-\frac{1}{2}}} (I_T \otimes (R_t^A)^{-1}) \left[\underset{TN\times 1}{y^{A*}} - \underset{TN\times K}{x^A}\underset{K\times 1}{\beta^A} - (I_T \otimes R_t^A)(\underset{T\times 1}{\iota_T} \otimes \underset{N\times K}{\bar{x}^A})\underset{K\times 1}{\delta^A}\right]$[13]

The log-likelihood then is (with K a constant independent of parameters):

$$\begin{aligned} & l(\theta^A, \Omega_{\xi^A}) \\ = \quad & K - \frac{1}{2} ln|\Omega_{u^A}| - \frac{1}{2}\left[y^{A*} - x^A\beta^A - (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right]' \Omega_{u^A}^{-1} \left[y^{A*} - x^A\beta^A - (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right] \\ = \quad & K - \frac{1}{2} ln|\Omega_{\xi^A}| + ln|I_T \otimes (R_t^A)^{-1}| - \frac{1}{2}\left[y^{A*} - x^A\beta^A - (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right]' (I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1}) \left[y^{A*} - x^A\beta^A - (I_T \otimes R_t^A)(\iota_T \otimes \right. \\ = \quad & K - \frac{1}{2} ln|\Omega_{\xi^A}| + ln|I_T \otimes (R_t^A)^{-1}| - \frac{1}{2}\nu^{A\prime}\nu^A, \end{aligned}$$

[13] $\Omega_{\xi^A}^{\frac{1}{2}}$ can be obtained using Cholesky decomposition, i.e. $\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})' = \Omega_{\xi^A}$, a symmetric matrix.

2.1 Asymptotic Normality of MLE

To proof the asymptotic normality of the MLE we follow closely Greene (2011) both with regard to the notation as well as the steps taken in the proof.[14]

We define

$$\frac{\partial l}{\partial \theta^A} = \sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial l_{ti}}{\partial \theta^A} = \sum_{t=1}^{T}\sum_{i=1}^{N} g_{ti}(\theta^A) = g^A(\theta^A),$$

the gradiant or score vector of the log-likelihood function

$$\frac{\partial^2 l}{\partial \theta^A \partial \theta^{A\prime}} = \sum_{t=1}^{T}\sum_{i=1}^{N}\frac{\partial^2 l_{ti}}{\partial \theta^A \partial \theta^{A\prime}} = \sum_{t=1}^{T}\sum_{i=1}^{N} H_{ti}(\theta^A) = H(\theta^A),$$

the Hessian of the log-likelihood function.

By definition of the MLE we know that

$$g^A(\hat{\theta}^A) = 0,$$

i.e. the gradient of the log-likelihood function evaluated at $\hat{\theta}^A$ is equal to zero, since $\hat{\theta}^A$ maximizes the log-likelihood function.

Using a Taylor series expansion around the true parameter θ^A (neglecting all higher order terms except the second, which we can do by the mean value theorem), this set of derivatives becomes:

$$g^A(\hat{\theta}^A) = g^A(\theta^A) + H(\theta^A)(\hat{\theta}^A - \theta^A) = 0,$$
$$\text{with} \quad \theta^A = k\hat{\theta}^A + (1-k)\theta^A \quad \text{for some} \quad 0 < k < 1.$$

Solving for $(\hat{\theta}^A - \theta^A)$ and multiplying both sides of the equation with $\sqrt{TN}$ yields:

$$\sqrt{TN}(\hat{\theta}^A - \theta^A) = \left[-H(\theta^A)\right]^{-1}\left[\sqrt{TN} g^A(\theta^A)\right]$$

Since we have assumed consistency of $\hat{\theta}^A$

$$\hat{\theta}^A \xrightarrow{p} \theta^A,$$

it follows

$$\theta^A \xrightarrow{p} \theta^A.$$

[14] We refrain from proving consistency of the MLE and proceed directly with the proof for asymptotic normality since it is instructive for the derivation of the asymptotic distribution of the second-stage estimator, i.e. the Murphy-Topel correction. All proofs that follow assume implicitly that we can estimate all parameters consistently.

By the Continuous Mapping Theorem:

$$\sqrt{TN}(\hat{\theta}^A - \theta^A) \xrightarrow{p} \left[-H(\theta^A)\right]^{-1} \left[\sqrt{TN} g^A(\theta^A)\right]$$

Expanding the right-hand-side by $1 = \frac{TN}{TN}$:

$$\begin{aligned} \sqrt{TN}(\hat{\theta}^A - \theta^A) &\xrightarrow{p} \left[-\frac{1}{TN} H(\theta^A)\right]^{-1} \left[\sqrt{TN} \frac{1}{TN} g^A(\theta^A)\right] \\ &= \left[-\bar{H}(\theta^A)\right]^{-1} \sqrt{TN} \bar{g}^A(\theta^A) \end{aligned}$$

Finally:

$$\begin{aligned} \sqrt{TN} \bar{g}^A(\theta^A) &\xrightarrow{d} N\left(0, -E\left[\bar{H}(\theta^A)\right]\right) \\ -\bar{H}(\theta^A) &\xrightarrow{p} -E\left[\bar{H}(\theta^A)\right] \end{aligned}$$

Then by Slutzky's Theorem:

$$\begin{aligned} & \left[-\bar{H}(\theta^A)\right]^{-1} \sqrt{TN} \bar{g}^A(\theta^A) \\ \xrightarrow{d} \quad & N\left(0, \left\{-E\left[\bar{H}(\theta^A)\right]\right\}^{-1} \left\{-E\left[\bar{H}(\theta^A)\right]\right\} \left\{-E\left[\bar{H}(\theta^A)\right]\right\}^{-1}\right) \end{aligned}$$

Or since convergence in probability implies convergence in distribution:

$$\begin{aligned} \sqrt{TN}(\hat{\theta}^A - \theta^A) &\xrightarrow{d} N\left(0, \left\{-E\left[\bar{H}(\theta^A)\right]\right\}^{-1}\right) \\ &\xrightarrow{d} N\left(0, TN \mathcal{I}(\theta^A)^{-1}\right), \end{aligned}$$

where $\mathcal{I}(\theta^A)$ denotes the Information Matrix.

2.2 Derivation of the Information Matrix of the Random Effects Spatial Error Model

The following derivations closely follow Anselin (1988, pp. 74-77).[15]

[15] We use the following properties in our derivations of the score vector and the information matrix:

1. $(A \otimes B)^{-1} = (R_t^A \otimes R_t^B)$
2. $(A \otimes B)' = (A' \otimes B')$
3. $(A \otimes B)(C \otimes D) = (AC \otimes BD)$
4. $\frac{\partial X(a)'}{\partial a} = \left[\frac{\partial X(a)}{\partial a}\right]'$
5. $\frac{\partial ln|X(a)|}{\partial a} = Tr\left(X(a)^{-1}\frac{\partial X(a)}{\partial a}\right)$
6. $\frac{\partial X(a)^{-1}}{\partial a} = -X(a)^{-1}\frac{\partial X(a)}{\partial a}X(a)^{-1}$

2.2.1 Derivation of First-Order Derivatives

The score vector is defined as the vector of first-order derivatives of the log-likelihood function.

$$\underset{1\times 1}{\frac{\partial l}{\partial \beta_k^A}} = -\frac{1}{2}\frac{\partial \nu^{A\prime}\nu^A}{\partial \beta_k^A} = -\frac{1}{2}\underbrace{\left[\frac{\partial \nu^{A\prime}}{\partial \beta_k^A}\nu^A + \nu^{A\prime}\frac{\partial \nu^A}{\partial \beta_k^A}\right]}_{=2\nu^{A\prime}\frac{\partial \nu^A}{\partial \beta_k^A}} = -\nu^{A\prime}\frac{\partial \nu^A}{\partial \beta_k^A} = \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})x_k^A$$

$$\begin{aligned}
\underset{1\times 1}{\frac{\partial l}{\partial \delta_k^A}} &= -\frac{1}{2}\frac{\partial \nu^{A\prime}\nu^A}{\partial \delta_k^A} = -\frac{1}{2}\left[\frac{\partial \nu^{A\prime}}{\partial \delta_k^A}\nu + \nu'\frac{\partial \nu^A}{\partial \delta_k^A}\right] = -\nu^{A\prime}\frac{\partial \nu^A}{\partial \delta_k^A} \\
&= \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}\underbrace{(I_T \otimes (R_t^A)^{-1})(I_T \otimes R_t^A)}_{=(I_T\otimes I_N)=I_{TN}}(\iota_T \otimes \bar{x}_k^A) = \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A)
\end{aligned}$$

$$\begin{aligned}
\underset{1\times 1}{\frac{\partial l}{\partial \rho^A}} &= \frac{\partial ln|I_T \otimes A|}{\partial \rho^A} - \frac{1}{2}\frac{\partial \nu^{A\prime}\nu^A}{\partial \rho^A} = Tr\left[(I_T \otimes R_t^A)\frac{\partial I_T \otimes A}{\partial \rho^A}\right] - \frac{1}{2}\left[\frac{\partial \nu^{A\prime}}{\partial \rho^A}\nu^A + \nu^{A\prime}\frac{\partial \nu^A}{\partial \rho^A}\right] \\
&= Tr\left[(I_T \otimes R_t^A)\left(\underbrace{\frac{\partial I_N}{\partial \rho^A}}_{=0}\otimes A + I_N \otimes \frac{\partial (R_t^A)^{-1}}{\partial \rho^A}\right)\right] - \nu'\frac{\partial \nu}{\partial \rho^A} \\
&= Tr\left[(I_T \otimes R_t^A)\left(I_N \otimes -W_t\right)\right] - \\
&- \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}\left\{(I_T \otimes \frac{\partial (R_t^A)^{-1}}{\partial \rho^A})\left[y^{A*} - x^A\beta^A - (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right] - (I_T \otimes (R_t^A)^{-1})(I_T \otimes \frac{\partial R_t^A}{\partial \rho^A})(\iota_T \otimes \bar{x}^A)\delta^A\right\} \\
&= Tr\left(I_N \otimes -R_t^A W_t\right) - \\
&- \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}\left\{(I_T \otimes -W_t)(y^{A*} - x^A\beta^A) - (I_T \otimes -W_t)(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A - (I_T \otimes (R_t^A)^{-1})(I_T \otimes (-R_t^A)(-W_t)R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right\} \\
&= Tr\left(I_N \otimes -R_t^A W_t\right) - \\
&- \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}\left\{(I_T \otimes -W_t)(y^{A*} - x^A\beta^A) + (I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A - (I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\right\} \\
&= -Tr\left(I_T \otimes R_t^A W_t\right) + \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)
\end{aligned}$$

2.2.2 Derivation of Second-Order Derivatives

The information matrix is defined as the Hessian matrix of the log-likelihood function.

$$\begin{aligned}
\underset{1\times1}{\frac{\partial^2 l}{\partial\beta_k^A\partial\beta_l^A}} &= \frac{\nu^{A\prime}}{\partial\beta_l^A}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A = -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_l^A\right]'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A = -x_l^{A\prime}(I_T\otimes(R_t^A)^{-1})'(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A = \\
&= -x_l^{A\prime}((I_T\otimes(R_t^A)^{-1\prime}))[\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})']^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A = -x_l^{A\prime}((I_T\otimes(R_t^A)^{-1\prime}))\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A \quad \text{for all} \quad k=1,...,K; l=1,...,K
\end{aligned}$$

$$\begin{aligned}
\underset{1\times1}{\frac{\partial^2 l}{\partial\beta_k^A\partial\delta_l^A}} &= \frac{\nu^{A\prime}}{\partial\delta_l^A}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A = -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T\otimes\bar{x}_l^A)\right]'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A = -(\iota_T'\otimes\bar{x}_l^{A\prime})(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A \\
&= -(\iota_T'\otimes\bar{x}_l^{A\prime})[\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})']^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A = -(\iota_T'\otimes\bar{x}_l^{A\prime})\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A \quad \text{for all} \quad k=1,...,K; l=1,...,K
\end{aligned}$$

$$\begin{aligned}
\underset{1\times1}{\frac{\partial^2 l}{\partial\beta_k^A\partial\rho^A}} &= \frac{\nu^{A\prime}}{\partial\rho^A}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A + \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes\frac{\partial(R_t^A)^{-1}}{\partial\rho^A})x_k^A \\
&= -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes W_t)\left(y^{A*}-x^A\beta^A\right)\right]'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A + \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes -W_t)x_k^A \\
&= -\left(y^{A*}-x^A\beta^A\right)'(I_T\otimes W_t)'(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A - \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes W_t)x_k^A \\
&= -\left(y^{A*}-x^A\beta^A\right)'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A - \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes W_t)x_k^A
\end{aligned}$$

$$\underset{1\times 1}{\frac{\partial^2 l}{\partial \delta_k^A \partial \beta_l^A}} = \frac{\nu^{A\prime}}{\partial \beta_l^A}\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A) = -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})x_l^A\right]' \Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A) = -x_l^{A\prime}(I_T \otimes (R_t^A)^{-1})'(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A) =$$

$$= -x_l^{A\prime}((I_T \otimes (R_t^A)^{-1\prime}))[\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})']^{-1}(\iota_T \otimes \bar{x}_k^A) = -x_l^{A\prime}((I_T \otimes (R_t^A)^{-1\prime}))\Omega_{\xi^A}^{-1}(\iota_T \otimes \bar{x}_k^A) \quad \text{for all} \quad k = 1, ..., K; l = 1, ..., K$$

$$\underset{1\times 1}{\frac{\partial^2 l}{\partial \delta_k^A \partial \delta_l^A}} = \frac{\nu^{A\prime}}{\partial \delta_l^A}\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A) = -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_l^A)\right]' \Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A) = -(\iota_T' \otimes \bar{x}_l^{A\prime})(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A)$$

$$= -(\iota_T' \otimes \bar{x}_l^{A\prime})[\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})']^{-1}(\iota_T \otimes \bar{x}_k^A) = -(\iota_T' \otimes \bar{x}_l^{A\prime})\Omega_{\xi^A}^{-1}(\iota_T \otimes \bar{x}_k^A) \quad \text{for all} \quad k = 1, ..., K; l = 1, ..., K$$

$$\underset{1\times 1}{\frac{\partial^2 l}{\partial \delta_k^A \partial \rho^A}} = \frac{\nu^{A\prime}}{\partial \rho^A}\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A)$$

$$= -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right]' \Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A)$$

$$= -\left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A)$$

$$= -\left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(\iota_T \otimes \bar{x}_k^A)$$

$$
\begin{aligned}
\underset{1\times 1}{\frac{\partial^2 l}{\partial \rho^A \partial \beta_k^A}} &= \frac{\nu^{A\prime}}{\partial \beta_k^A}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) - \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)x_k^A \\
&= -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})x_k^A\right]'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) - \nu'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)x_k^A \\
&= -x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) - \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)x_k^A \\
&= -x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'[\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})']^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) - \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)x_k^A \\
&= -x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) - \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)x_k^A
\end{aligned}
$$

$$
\begin{aligned}
\underset{1\times 1}{\frac{\partial^2 l}{\partial \rho^A \partial \delta_k^A}} &= \frac{\nu^{A\prime}}{\partial \delta_k^A}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) \\
&= -\left[\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T \otimes \bar{x}_k^A)\right]'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) \\
&= -(\iota_T' \otimes \bar{x}_k^{A\prime})(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) \\
&= -(\iota_T' \otimes \bar{x}_k^{A\prime})[\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})']^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) \\
&= -(\iota_T' \otimes \bar{x}_k^{A\prime})\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)
\end{aligned}
$$

$$
\begin{aligned}
\underset{1\times 1}{\frac{\partial^2 l}{\partial \rho^A \partial \rho^A}} &= -Tr\left(I_N \otimes \frac{\partial R_t^A}{\partial \rho^A}W_t\right) + \frac{\nu^{A\prime}}{\partial \rho^A}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) \\
&= -Tr\left(I_T \otimes R_t^A W_t R_t^A W_t\right) - \left[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right]'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) \\
&= -Tr\left(I_T \otimes R_t^A W_t R_t^A W_t\right) - \left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'(\Omega_{\xi^A}^{-\frac{1}{2}})'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) \\
&= -Tr\left(I_T \otimes R_t^A W_t R_t^A W_t\right) - \left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)
\end{aligned}
$$

2.2.3 Information Matrix

Recall:

$$\begin{aligned}
y^{A*} &= x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A \\
u^A &= (I_T \otimes R_t^A)\xi^A \\
\xi^A &= (I_T \otimes (R_t^A)^{-1})u^A \\
\nu^A &= \Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})u^A
\end{aligned}$$

In this section we will make use of the following expected values (implicitly conditioning on covariates $x^A, \bar{x}^A$):

$$\begin{aligned}
E[\xi^A] &= 0 \\
E[u^A] &= E[(I_T \otimes R_t^A)\xi^A] = (I_T \otimes R_t^A)E[\xi^A] = 0 \\
E[\nu^A] &= E[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})u^A] = \Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})E[u^A] = 0 \\
E[y^{A*}] &= E[x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A] = x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + E[u^A] = x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A \\
E[\xi^A\xi^{A\prime}] &= \Omega_{\xi^A} \\
E[u^A u^{A\prime}] &= E[(I_T \otimes R_t^A)\xi^A\xi^{A\prime}(I_T \otimes R_t^A)'] = (I_T \otimes R_t^A)E[\xi^A\xi^{A\prime}](I_T \otimes R_t^A)' = (I_T \otimes R_t^A)\Omega_{\xi^A}(I_T \otimes R_t^A)' \\
E[\nu^A\nu^{A\prime}] &= E[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})u^A u^{A\prime}(I_T \otimes (R_t^A)^{-1})'(\Omega_{\xi^A}^{-\frac{1}{2}})'] = \Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})E[u^A u^{A\prime}](I_T \otimes (R_t^A)^{-1})'(\Omega_{\xi^A}^{-\frac{1}{2}})' \\
&= \Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes (R_t^A)^{-1})(I_T \otimes R_t^A)\Omega_{\xi^A}(I_T \otimes R_t^A)'(I_T \otimes (R_t^A)^{-1})'(\Omega_{\xi^A}^{-\frac{1}{2}})' = \Omega_{\xi^A}^{-\frac{1}{2}} I_{TN}\Omega_{\xi^A}^{\frac{1}{2}}(\Omega_{\xi^A}^{\frac{1}{2}})' I_{TN}(\Omega_{\xi^A}^{-\frac{1}{2}})' = I_{TN}
\end{aligned}$$

$$
\begin{aligned}
E[y^{A*}y^{A*\prime}] &= E\left\{[x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A][x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A]'\right\} \\
&= E[[x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A][x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]' + \\
&+ [x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]u^{A\prime} + u^A[x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]' + u^A u^{A\prime}] \\
&= E[[x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A][x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]'] + \\
&+ E[[x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]u^{A\prime}] + E[u^A[x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]'] + E[u^A u^{A\prime}] \\
&= [x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A][x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]' + \\
&+ [x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]E[u^{A\prime}] + E[u^A][x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]' + (I_T \otimes R_t^A)\Omega_{\xi^A}(I_T \otimes R_t^A)' \\
&= [x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A][x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A]' + (I_T \otimes R_t^A)\Omega_{\xi^A}(I_T \otimes R_t^A)'
\end{aligned}
$$

The Information Matrix $\mathcal{I}(\theta^A)$ is defined as:

$$
\begin{aligned}
\mathcal{I}(\theta^A) &= -E\left[\frac{\partial^2 l}{\partial\theta^A\partial\theta^{A\prime}}\right] \\
&= -E\begin{bmatrix}
\frac{\partial^2 l}{\partial\beta_1^A\partial\beta_1^A} & \cdots & \frac{\partial^2 l}{\partial\beta_1^A\partial\beta_K^A} & \frac{\partial^2 l}{\partial\beta_1^A\partial\delta_1^A} & \cdots & \frac{\partial^2 l}{\partial\beta_1^A\partial\delta_R^A} & \frac{\partial^2 l}{\partial\beta_1^A\partial\rho^A} \\
\vdots & & & & & & \vdots \\
\frac{\partial^2 l}{\partial\beta_K^A\partial\beta_1^A} & \cdots & \frac{\partial^2 l}{\partial\beta_K^A\partial\beta_K^A} & \frac{\partial^2 l}{\partial\beta_K^A\partial\delta_1^A} & \cdots & \frac{\partial^2 l}{\partial\beta_K^A\partial\delta_R^A} & \frac{\partial^2 l}{\partial\beta_K^A\partial\rho^A} \\
\frac{\partial^2 l}{\partial\delta_1^A\partial\beta_1^A} & \cdots & \frac{\partial^2 l}{\partial\delta_1^A\partial\beta_K^A} & \frac{\partial^2 l}{\partial\delta_1^A\partial\delta_1^A} & \cdots & \frac{\partial^2 l}{\partial\delta_1^A\partial\delta_R^A} & \frac{\partial^2 l}{\partial\delta_1^A\partial\rho^A} \\
\vdots & & & & & & \vdots \\
\frac{\partial^2 l}{\partial\delta_R^A\partial\beta_1^A} & \cdots & \frac{\partial^2 l}{\partial\delta_R^A\partial\beta_K^A} & \frac{\partial^2 l}{\partial\delta_R^A\partial\delta_1^A} & \cdots & \frac{\partial^2 l}{\partial\delta_R^A\partial\delta_R^A} & \frac{\partial^2 l}{\partial\delta_R^A\partial\rho^A} \\
\frac{\partial^2 l}{\partial\rho^A\partial\beta_1^A} & \cdots & \frac{\partial^2 l}{\partial\rho^A\partial\beta_K^A} & \frac{\partial^2 l}{\partial\rho^A\partial\delta_1^A} & \cdots & \frac{\partial^2 l}{\partial\rho^A\partial\delta_R^A} & \frac{\partial^2 l}{\partial\rho^A\partial\rho^A}
\end{bmatrix}
\end{aligned}
$$

The elements of the Information matrix are given by:

$$-E\left[\frac{\partial^2 l}{\partial\beta_k^A\partial\beta_l^A}\right] = -E\left[-x_l^{A\prime}((I_T\otimes(R_t^A)^{-1\prime}))\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\right] = x_l^{A\prime}((I_T\otimes(R_t^A)^{-1\prime}))\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A$$

$$-E\left[\frac{\partial^2 l}{\partial\beta_k^A\partial\delta_l^A}\right] = -E\left[-(\iota_T'\otimes\bar{x}_l^{A\prime})\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\right] = (\iota_T'\otimes\bar{x}_l^{A\prime})\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A$$

$$\begin{aligned}
-E\left[\frac{\partial^2 l}{\partial\beta_k^A\partial\rho^A}\right] &= -E\left[-\left(y^{A*}-x^A\beta^A\right)'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A-\nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes W_t)x_k^A\right]\\
&= \left(E\left[y^{A*}\right]-x^A\beta^A\right)'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A+E\left[\nu^{A\prime}\right]\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes W_t)x_k^A\\
&= \left[x^A\beta^A+(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A-x^A\beta^A\right]'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\\
&= \left[(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\right]'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\\
&= \delta^{A\prime}(\iota_T\otimes\bar{x}^A)'(I_T\otimes R_t^A)'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\\
&= \delta^{A\prime}(\iota_T\otimes\bar{x}^A)'(I_T\otimes A^{-1\prime}W_t')\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A
\end{aligned}$$

$$-E\left[\frac{\partial^2 l}{\partial\delta_k^A\partial\beta_l^A}\right] = -E\left[-x_l^{A\prime}((I_T\otimes(R_t^A)^{-1\prime}))\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\right] = x_l^{A\prime}((I_T\otimes(R_t^A)^{-1\prime}))\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)$$

$$-E\left[\frac{\partial^2 l}{\partial\delta_k^A\partial\delta_l^A}\right] = -E\left[-(\iota_T'\otimes\bar{x}_l^{A\prime})\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\right] = (\iota_T'\otimes\bar{x}_l^{A\prime})\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)$$

$$\begin{aligned}
-E\left[\frac{\partial^2 l}{\partial\delta_k^A\partial\rho^A}\right] &= -E\left[-\left(y^{A*}-x^A\beta^A\right)'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\right]\\
&= \left(E\left[y^{A*}\right]-x^A\beta^A\right)'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\\
&= \left[x^A\beta^A+(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A-x^A\beta^A\right]'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\\
&= \delta^{A\prime}(\iota_T\otimes\bar{x}^A)'(I_T\otimes R_t^A)'(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\\
&= \delta^{A\prime}(\iota_T\otimes\bar{x}^A)'(I_T\otimes A^{-1\prime}W_t')\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)
\end{aligned}$$

$$
\begin{aligned}
-E\left[\frac{\partial^2 l}{\partial \rho^A \partial \beta_k^A}\right] &= -E\left[-x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right) - \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)x_k^A\right] \\
&= x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(E\left[y^{A*}\right] - x^A\beta^A\right) + E\left[\nu^{A\prime}\right]\Omega_{\xi^A}^{-\frac{1}{2}}(I_T \otimes W_t)x_k^A \\
&= x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A - x^A\beta^A\right) \\
&= x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A \\
&= x_k^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A
\end{aligned}
$$

$$
\begin{aligned}
-E\left[\frac{\partial^2 l}{\partial \rho^A \partial \delta_k^A}\right] &= -E\left[-(\iota_T' \otimes \bar{x}_k^{A\prime})\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right] \\
&= (\iota_T' \otimes \bar{x}_k^{A\prime})\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(E\left[y^{A*}\right] - x^A\beta^A\right) \\
&= (\iota_T' \otimes \bar{x}_k^{A\prime})\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(x^A\beta^A + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A - x^A\beta^A\right) \\
&= (\iota_T' \otimes \bar{x}_k^{A\prime})\Omega_{\xi^A}^{-1}(I_T \otimes W_t)(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A \\
&= (\iota_T' \otimes \bar{x}_k^{A\prime})\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A
\end{aligned}
$$

$$
\begin{aligned}
-E\left[\frac{\partial^2 l}{\partial \rho^A \partial \rho^A}\right] &= -E\left[-Tr\left(I_T \otimes R_t^A W_t R_t^A W_t\right) - \left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right] \\
&= E\left[Tr\left(I_T \otimes R_t^A W_t R_t^A W_t\right)\right] + E\left[\left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right] \\
&= Tr\left(I_T \otimes R_t^A W_t R_t^A W_t\right) + E\left[\left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right]
\end{aligned}
$$

The second expectation in the above term requires some more detailed explanation. First, we introduce a trace operator as the term inside the expectation is a scalar and the trace of a scalar a is itself, i.e. $Tr(a) = a$.

$$\begin{aligned} & E\left[\left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right] \\ = \; & E\left[Tr\left[\left(y^{A*} - x^A\beta^A\right)'(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\right]\right] \end{aligned}$$

Then we make use of the cyclic property of the trace (i.e. the fact that it is invariant to cyclic permutations: $Tr(ABCD) = Tr(BCDA) = Tr(CDAB) = Tr(DABC)$)

$$\begin{aligned} = \; & E\left[Tr\left[(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left(y^{A*} - x^A\beta^A\right)\left(y^{A*} - x^A\beta^A\right)'\right]\right] \\ = \; & Tr\left[(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)E\left[\left(y^{A*} - x^A\beta^A\right)\left(y^{A*} - x^A\beta^A\right)'\right]\right] \end{aligned}$$

Next, we substitute $y^{A*} - x^A\beta^A = (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A$.

$$= \; Tr\left[(I_T \otimes W_t)'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)E\left[\left((I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right)\left((I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right)'\right]\right]$$

To derive $E\left[\left((I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right)\left((I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right)'\right]$, we expand the term inside the expectation:

$$\begin{aligned} & E\left[\left((I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right)\left((I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right)'\right] \\ = \; & E\left[(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\delta^{A\prime}(\iota_T \otimes \bar{x}^A)'(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)' + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A u^{A\prime} + u^A\delta^{A\prime}(\iota_T \otimes \bar{x}^A)'(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)' + u^A u^{A\prime}\right] \\ = \; & E\left[(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\delta^{A\prime}(\iota_T \otimes \bar{x}^A)'(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)'\right] + E\left[(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A u^{A\prime}\right] + \\ + \; & E\left[u^A\delta^{A\prime}(\iota_T \otimes \bar{x}^A)'(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)'\right] + E\left[u^A u^{A\prime}\right] \\ = \; & (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\delta^{A\prime}(\iota_T \otimes \bar{x}^A)'(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)' + (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A E\left[u^{A\prime}\right] + \\ + \; & E\left[u^A\right]\delta^{A\prime}(\iota_T \otimes \bar{x}^A)'(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)' + (I_T \otimes R_t^A)\Omega_{\xi^A}(I_T \otimes R_t^A)' \\ = \; & (I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A\delta^{A\prime}(\iota_T \otimes \bar{x}^A)'(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)' + (I_T \otimes R_t^A)\Omega_{\xi^A}(I_T \otimes R_t^A)' \end{aligned}$$

Putting things together, we get

$$
\begin{aligned}
-E\left[\frac{\partial^2 l}{\partial\rho^A\partial\rho^A}\right] &= Tr\left(I_T\otimes R_t^A W_t R_t^A W_t\right) + \\
&+ Tr\left[(I_T\otimes W_t)'\Omega_{\xi^A}^{-1}(I_T\otimes W_t)\left[(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\delta^{A\prime}(\iota_T\otimes\bar{x}^A)'(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)' + (I_T\otimes R_t^A)\Omega_{\xi^A}(I_T\otimes R_t^A)'\right]\right]
\end{aligned}
$$

3 Deriving the Asymptotic Distribution of θ^B

3.1 Murphy-Topel Correction

To derive the asymptotic distribution of our second-stage parameters we follow again closely Greene's (2011) derivation of the Murphy-Topel correction for the case of MLE and adapt it to the NLLS case.[16]

From Appendix 2.1 we know

$$\begin{aligned}\sqrt{TN}(\hat{\theta}^A - \theta^A) \quad &\xrightarrow{d} \quad \left[-\bar{H}(\theta^A)\right]^{-1} \sqrt{TN}\bar{g}^A(\theta^A) \\ &\xrightarrow{d} \quad N\left(0, TN\mathcal{I}(\theta^A)^{-1}\right),\end{aligned}$$

where $\mathcal{I}(\theta^A)$ denotes the Information Matrix.

In the following we will derive the asymptotic distribution of θ^B using a similar line of argumentation as in Appendix 2.1.

By the definition of the NLLS estimator we know that

$$g^B(\hat{\theta}^A, \hat{\theta}^B) = 0,$$

i.e. the gradient of the sum of squared residuals evaluated at $\hat{\theta}^A$ and $\hat{\theta}^B$ is equal to zero, since $\hat{\theta}^A$ and $\hat{\theta}^B$ minimize the sum of squared residuals.

Using a Taylor series expansion around the true parameter θ^A and θ^B (neglecting all higher order terms except the second, which we can do by the mean value theorem), this set of derivatives becomes:

$$\begin{aligned}g(\hat{\theta}^A, \hat{\theta}^B) = g^B(\theta^A, \theta^B) \quad &+ \quad H^B(\theta^A, \theta^B)(\hat{\theta}^B - \theta^B) + \\ &+ \quad H^A(\theta^A, \theta^B)(\hat{\theta}^A - \theta^A) = 0,\end{aligned}$$

with

$$\begin{aligned}\theta^A \quad &= \quad k\hat{\theta}^A + (1-k)\theta^A, \\ \theta^B \quad &= \quad k\hat{\theta}^B + (1-k)\theta^B \quad \text{for some} \quad 0 < k < 1.\end{aligned}$$

[16] Again, we refrain from proving consistency of the NLLS estimator. All proofs that follow assume implicitly that we can estimate all parameters consistently.

Solving for $(\hat{\theta}^B - \theta^B)$:

$$\begin{aligned}(\hat{\theta}^B - \theta^B) &= \left[-H^B(\theta^A, \theta^B)\right]^{-1} \left[g^B(\theta^A, \theta^B) + H^A(\theta^A, \theta^B)(\hat{\theta}^A - \theta^A)\right] \\ &= \left[-H^B(\theta^A, \theta^B)\right]^{-1} g^B(\theta^A, \theta^B) + \left[-H^B(\theta^A, \theta^B)\right]^{-1} H^A(\theta^A, \theta^B)(\hat{\theta}^A - \theta^A)\end{aligned}$$

Multiplying both sides of the equation with $\sqrt{TN}$ yields:

$$\begin{aligned}\sqrt{TN}(\hat{\theta}^B - \theta^B) &= \left[-H^B(\theta^A, \theta^B)\right]^{-1} \sqrt{TN} g^B(\theta^A, \theta^B) + \\ &+ \left[-H^B(\theta^A, \theta^B)\right]^{-1} H^A(\theta^A, \theta^B)\sqrt{TN}(\hat{\theta}^A - \theta^A)\end{aligned}$$

Since we have assumed consistency of $\hat{\theta}^A$ and $\hat{\theta}^B$.

$$\begin{aligned}\hat{\theta}^A &\xrightarrow{p} \theta^A, \\ \hat{\theta}^B &\xrightarrow{p} \theta^B,\end{aligned}$$

it follows

$$\begin{aligned}\theta^A &\xrightarrow{p} \theta^A, \\ \theta^B &\xrightarrow{p} \theta^B.\end{aligned}$$

By the Continuous Mapping Theorem:

$$\begin{aligned}\sqrt{TN}(\hat{\theta}^B - \theta^B) &\xrightarrow{p} \left[-H^B(\theta^A, \theta^B)\right]^{-1} \sqrt{TN} g^B(\theta^A, \theta^B) + \\ &+ \left[-H^B(\theta^A, \theta^B)\right]^{-1} H^A(\theta^A, \theta^B)\sqrt{TN}(\hat{\theta}^A - \theta^A)\end{aligned}$$

Expanding the right-hand-side by $1 = \frac{TN}{TN}$:

$$\begin{aligned}\sqrt{TN}(\hat{\theta}^B - \theta^B) &\xrightarrow{p} \left[-\frac{1}{TN} H^B(\theta^A, \theta^B)\right]^{-1} \sqrt{TN} \frac{1}{TN} g^B(\theta^A, \theta^B) + \\ &+ \left[-\frac{1}{TN} H^B(\theta^A, \theta^B)\right]^{-1} \frac{1}{TN} H^A(\theta^A, \theta^B)\sqrt{TN}(\hat{\theta}^A - \theta^A) \\ &= \left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} \sqrt{TN} \bar{g}^B(\theta^A, \theta^B) + \\ &+ \left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} \bar{H}^A(\theta^A, \theta^B)\sqrt{TN}(\hat{\theta}^A - \theta^A)\end{aligned}$$

Since $\sqrt{TN}(\hat{\theta}^A - \theta^A) \xrightarrow{d} \left[-\bar{H}(\theta^A)\right]^{-1} \sqrt{TN} \bar{g}^A(\theta^A)$:

$$\begin{aligned}\sqrt{TN}(\hat{\theta}^B - \theta^B) &\xrightarrow{d} \left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} \sqrt{TN} \bar{g}^B(\theta^A, \theta^B) + \\ &+ \left[-\bar{H}^B(\theta^A, \theta^B)\right]^{-1} \bar{H}^A(\theta^A, \theta^B) \left[-\bar{H}(\theta^A)\right]^{-1} \sqrt{TN} \bar{g}^A(\theta^A)\end{aligned}$$

Moreover:

$$\begin{aligned}
\sqrt{TN}\bar{g}^A(\theta^A) &\overset{d}{\to} N\left(0, \Omega_{g^A}\right) \\
\sqrt{TN}\bar{g}^B(\theta^A, \theta^B) &\overset{d}{\to} N\left(0, \Omega_{g^B}\right)
\end{aligned}$$

Note that $\Omega_{g^A} = TN\mathcal{I}(\theta^A)^{-1} = \Omega_{\theta^A}$ and that we will derive Ω_{g^B} in the next section. Finally:

$$\begin{aligned}
-\bar{H}(\theta^A) &\overset{p}{\to} -E\left[\bar{H}(\theta^A)\right] = \frac{1}{TN}\mathcal{I}(\theta^A) = \Omega_{\theta^A}^{-1} \\
-\bar{H}^A(\theta^A, \theta^B) &\overset{p}{\to} -E\left[\bar{H}^A(\theta^A, \theta^B)\right] \\
-\bar{H}^B(\theta^A, \theta^B) &\overset{p}{\to} -E\left[\bar{H}^B(\theta^A, \theta^B)\right]
\end{aligned}$$

Then again by Slutzky's Theorem:

$$\sqrt{TN}(\hat{\theta}^B - \theta^B) \overset{d}{\to} N\left(0, \Omega_{\theta^B}\right)$$

where $\Omega_{\theta^B} = Var\left[\sqrt{TN}(\hat{\theta}^B - \theta^B)\right]$.

This VC matrix is given by:

$$
\begin{aligned}
& Var\left[\sqrt{TN}(\hat{\theta}^B - \theta^B)\right] \\
= \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}\Omega_{g^B}E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1} + \\
+ \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}E\left[\bar{H}^A(\theta^A,\theta^B)\right]E\left[-\bar{H}(\theta^A)\right]^{-1}\Omega_{g^A}E\left[-\bar{H}(\theta^A)\right]^{-1}E\left[\bar{H}^A(\theta^A,\theta^B)\right]'E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1} + \\
+ \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}\Omega_{g^{BA}}E\left[-\bar{H}(\theta^A)\right]^{-1}E\left[\bar{H}^A(\theta^A,\theta^B)\right]'E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1} + \\
+ \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}E\left[\bar{H}^A(\theta^A,\theta^B)\right]E\left[-\bar{H}(\theta^A)\right]^{-1}\Omega_{g^{AB}}E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1} \\
= \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}\Omega_{g^B}E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1} + \\
+ \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}E\left[\bar{H}^A(\theta^A,\theta^B)\right]\Omega_{\theta^A}E\left[\bar{H}^A(\theta^A,\theta^B)\right]'E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1} + \\
+ \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}\Omega_{g^{BA}}\Omega_{\theta^A}E\left[\bar{H}^A(\theta^A,\theta^B)\right]'E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1} + \\
+ \;& E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}E\left[\bar{H}^A(\theta^A,\theta^B)\right]\Omega_{\theta^A}\Omega_{g^{AB}}E\left[-\bar{H}^B(\theta^A,\theta^B)\right]^{-1}
\end{aligned}
$$

3.2 Asymptotic Distribution of Score Vectors

The asymptotic joint distribution of the score vectors

$$\sqrt{TN}\bar{g}^A(\theta^A) = \frac{1}{\sqrt{TN}}\sum_t^T\sum_i^N \frac{\partial q^A\left(x_{ti}^A, \theta^A\right)}{\partial\theta_1} \quad \text{and}$$

$$\sqrt{TN}\bar{g}^B(\theta^A, \theta^B) = \frac{1}{\sqrt{TN}}\sum_t^T\sum_i^N \frac{\partial q^B\left(x_{ti}^{A0}, x_{ti}^B, \theta^A, \theta^B\right)}{\partial\theta^B}$$

is given by

$$\begin{pmatrix} \sqrt{TN}\bar{g}^A(\theta^A) \\ \sqrt{TN}\bar{g}^B(\theta^A, \theta^B) \end{pmatrix} \overset{d}{\to} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Omega_{g^A} & \Omega_{g^{AB}} \\ \Omega_{g^{BA}} & \Omega_{g^B} \end{pmatrix}\right),$$

by the Central Limit Theorem (CLM).

We know that Ω_{g^A} is the variance of the first stage score vector. Since the first stage is estimated by MLE, this is equivalent to $Var\left[\frac{\partial l}{\partial\theta^A}\right]$, which we know is the same as $-E\left[\frac{\partial^2 l}{\partial\theta^A\partial\theta^A}\right]^{-1} = \mathcal{I}(\theta^A)^{-1}$ given the Information Matrix Equality (Greene, 2011).

To compute the remaining terms of the VC matrix of the score vectors, we need to derive $\Omega_{g^{AB}} = \Omega'_{g^{BA}}$ and Ω_{g^B}. In order to do so, we compute the score vector of the second stage.

3.2.1 Derivation of First-Order Derivatives

Let

$$q^B = \nu^{B\prime}\nu^B \quad \text{with}$$

$$\nu^B = y^B - x^B\beta^B - (I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B - \tau\Lambda$$

$$\text{(Sample Selection)}$$

$$\nu^B = y^B - \alpha y^A - x^B\beta^B - (I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B - \tau\Lambda^g$$

$$\text{(Treatment Selection)},$$

Where we have defined

$$\Lambda = \begin{bmatrix} \psi_1\lambda_{11} \\ \vdots \\ \psi_N\lambda_{1N} \\ \vdots \\ \psi_1\lambda_{T1} \\ \vdots \\ \psi_N\lambda_{TN} \end{bmatrix} \quad \text{and} \quad \Lambda^g = \begin{bmatrix} \psi_1\lambda_{11}^g \\ \vdots \\ \psi_N\lambda_{1N}^g \\ \vdots \\ \psi_1\lambda_{T1}^g \\ \vdots \\ \psi_N\lambda_{TN}^g \end{bmatrix}$$

The first-order derivatives are then given by

$$\underset{1\times 1}{\frac{\partial q^B}{\partial \alpha}} = \frac{\partial \nu^{B\prime}\nu^B}{\partial \alpha} = \frac{\partial \nu^{B\prime}}{\partial \alpha}\nu^B + \nu^{B\prime}\frac{\partial \nu^B}{\partial \alpha} = 2\nu^{B\prime}\frac{\partial \nu^B}{\partial \alpha} = -2\nu^{B\prime} y^A \quad \text{(Treatment Selection)}$$

$$\underset{1\times 1}{\frac{\partial q^B}{\partial \beta_r^B}} = \frac{\partial \nu^{B\prime}\nu^B}{\partial \beta_r^B} = \frac{\partial \nu^{B\prime}}{\partial \beta_r^B}\nu^B + \nu^{B\prime}\frac{\partial \nu^B}{\partial \beta_r^B} = 2\nu^{B\prime}\frac{\partial \nu^B}{\partial \beta_r^B} = -2\nu^{B\prime} x_r^B$$

$$\underset{1\times 1}{\frac{\partial q^B}{\partial \delta_r^B}} = \frac{\partial \nu^{B\prime}\nu^B}{\partial \delta_r^B} = \frac{\partial \nu^{B\prime}}{\partial \delta_r^B}\nu^B + \nu^{B\prime}\frac{\partial \nu^B}{\partial \delta_r^B} = 2\nu^{B\prime}\frac{\partial \nu^B}{\partial \delta_r^B} = -2\nu^{B\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)$$

$$\underset{1\times 1}{\frac{\partial q^B}{\partial \tau}} = \frac{\partial \nu^{B\prime}\nu^B}{\partial \tau} = \frac{\partial \nu^{B\prime}}{\partial \tau}\nu^B + \nu^{B\prime}\frac{\partial \nu^B}{\partial \tau} = 2\nu^{B\prime}\frac{\partial \nu^B}{\partial \tau} = -2\nu^{B\prime}\Lambda \quad \text{(Sample Selection)}$$

$$\underset{1\times 1}{\frac{\partial q^B}{\partial \tau}} = \frac{\partial \nu^{B\prime}\nu^B}{\partial \tau} = \frac{\partial \nu^{B\prime}}{\partial \tau}\nu^B + \nu^{B\prime}\frac{\partial \nu^B}{\partial \tau} = 2\nu^{B\prime}\frac{\partial \nu^B}{\partial \tau} = -2\nu^{B\prime}\Lambda^g \quad \text{(Treatment Selection)}$$

$$\begin{aligned}
\underset{1\times 1}{\frac{\partial q^B}{\partial \rho^B}} &= \frac{\partial \nu^{B\prime}\nu^B}{\partial \rho^B} = \frac{\partial \nu^{B\prime}}{\partial \rho^B}\nu^B + \nu^{B\prime}\frac{\partial \nu^B}{\partial \rho^B} = 2\nu^{B\prime}\frac{\partial \nu^B}{\partial \rho^B} = -2\nu^{B\prime}\left[(I_T \otimes \frac{\partial R_t^B}{\partial \rho^B})(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right] \\
&= -2\nu^{B\prime}\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right] \quad \text{(Sample Selection)} \\
\underset{1\times 1}{\frac{\partial q^B}{\partial \rho^B}} &= \frac{\partial \nu^{B\prime}\nu^B}{\partial \rho^B} = \frac{\partial \nu^{B\prime}}{\partial \rho^B}\nu^B + \nu^{B\prime}\frac{\partial \nu^B}{\partial \rho^B} = 2\nu^{B\prime}\frac{\partial \nu^B}{\partial \rho^B} = -2\nu^{B\prime}\left[(I_T \otimes \frac{\partial R_t^B}{\partial \rho^B})(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right] \\
&= -2\nu^{B\prime}\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right] \quad \text{(Treatment Selection)}
\end{aligned}$$

The term $\frac{\partial \Lambda}{\partial \rho^B}$ can be derived as follows. First recall that $\psi_{ti} = \frac{\sum_{j=1}^N r_{tij}^B r_{tij}^A}{\sum_j^N (r_{tij}^A)^2}$ and hence $\frac{\partial \psi_{ti}}{\partial \rho^B} = \frac{\sum_{j=1}^N \frac{\partial r_{tij}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2}$. Further, note that $\frac{\partial r_{tij}^B}{\partial \rho^B}$ is the ijth element of the matrix $\frac{\partial R_t^B}{\partial \rho^B} = R_t^B W_t R_t^B$. Since λ_{ti} and λ_{ti}^g are not functions of ρ^B it follows that

$$\frac{\partial \Lambda}{\partial \rho^B} = \begin{bmatrix} \frac{\partial \psi_1}{\partial \rho^B} \lambda_{11} \\ \vdots \\ \frac{\partial \psi_N}{\partial \rho^B} \lambda_{1N} \\ \vdots \\ \frac{\partial \psi_1}{\partial \rho^B} \lambda_{T1} \\ \vdots \\ \frac{\partial \psi_N}{\partial \rho^B} \lambda_{TN} \end{bmatrix} = \begin{bmatrix} \frac{\sum_{j=1}^N \frac{\partial r_{t1j}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{11} \\ \vdots \\ \frac{\sum_{j=1}^N \frac{\partial r_{tNj}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{1N} \\ \vdots \\ \frac{\sum_{j=1}^N \frac{\partial r_{t1j}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{T1} \\ \vdots \\ \frac{\sum_{j=1}^N \frac{\partial r_{tNj}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{TN} \end{bmatrix} \quad \text{and} \quad \frac{\partial \Lambda^g}{\partial \rho^B} = \begin{bmatrix} \frac{\partial \psi_1}{\partial \rho^B} \lambda_{11}^g \\ \vdots \\ \frac{\partial \psi_N}{\partial \rho^B} \lambda_{1N}^g \\ \vdots \\ \frac{\partial \psi_1}{\partial \rho^B} \lambda_{T1}^g \\ \vdots \\ \frac{\partial \psi_N}{\partial \rho^B} \lambda_{TN}^g \end{bmatrix} = \begin{bmatrix} \frac{\sum_{j=1}^N \frac{\partial r_{t1j}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{11}^g \\ \vdots \\ \frac{\sum_{j=1}^N \frac{\partial r_{tNj}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{1N}^g \\ \vdots \\ \frac{\sum_{j=1}^N \frac{\partial r_{t1j}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{T1}^g \\ \vdots \\ \frac{\sum_{j=1}^N \frac{\partial r_{tNj}^B}{\partial \rho^B} r_{tij}^A}{\sum_j^N (r_{tij}^A)^2} \lambda_{TN}^g \end{bmatrix}$$

3.2.2 Variance and Covariance of Score Vectors

Recall the first-order derivatives of the score vectors of the first and second stage:

First-order derivatives of l

$$\begin{aligned}
\underset{1\times 1}{\frac{\partial l}{\partial \beta_k^A}} &= \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A = \Bigg[\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})\underbrace{\left[y^{A*}-x^A\beta^A-(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\right]}_{u^A}\Bigg]'\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A \\
&= u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-\frac{1}{2}}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})x_k^A = u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A \\
\underset{1\times 1}{\frac{\partial l}{\partial \delta_k^A}} &= \nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(\iota_T\otimes\bar{x}_k^A) = u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A) \\
\underset{1\times 1}{\frac{\partial l}{\partial \rho^A}} &= -Tr\left(I_T\otimes R_t^A W_t\right)+\nu^{A\prime}\Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes W_t)\left(y^{A*}-x^A\beta^A\right) = -Tr\left(I_T\otimes R_t^A W_t\right)+u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t)\Bigg(\underbrace{y^{A*}-x^A\beta^A}_{(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A+u^A}\Bigg) \\
&= -Tr\left(I_T\otimes R_t^A W_t\right)+u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t)\Bigg(\underbrace{y^{A*}-x^A\beta^A}_{(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A+u^A}\Bigg) \\
&= -Tr\left(I_T\otimes R_t^A W_t\right)+u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t)\left[(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A+u^A\right],
\end{aligned}$$

where we have used $\nu^A = \Omega_{\xi^A}^{-\frac{1}{2}}(I_T\otimes(R_t^A)^{-1})\underbrace{\left[y^{A*}-x^A\beta^A-(I_T\otimes R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\right]}_{u^A}$.

First-order derivatives of q

$$
\begin{aligned}
\underset{1\times 1}{\frac{\partial q^B}{\partial \alpha}} &= -2\nu^{B\prime} y^A \quad \text{(Treatment Selection)} \\
\underset{1\times 1}{\frac{\partial q^B}{\partial \beta_r^B}} &= -2\nu^{B\prime} x_r^B \\
\underset{1\times 1}{\frac{\partial q^B}{\partial \delta_r^B}} &= -2\nu^{B\prime} (I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) \\
\underset{1\times 1}{\frac{\partial q^B}{\partial \tau}} &= -2\nu^{B\prime} \Lambda \quad \text{(Sample Selection)} \\
\underset{1\times 1}{\frac{\partial q^B}{\partial \tau}} &= -2\nu^{B\prime} \Lambda^g \quad \text{(Treatment Selection)} \\
\underset{1\times 1}{\frac{\partial q^B}{\partial \rho^B}} &= -2\nu^{B\prime} \left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B} \right] \quad \text{(Sample Selection)} \\
\underset{1\times 1}{\frac{\partial q^B}{\partial \rho^B}} &= -2\nu^{B\prime} \left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda^g}{\partial \rho^B} \right] \quad \text{(Treatment Selection)}
\end{aligned}
$$

Sample Selection

$$
\begin{aligned}
\Omega_{g^B} &= Var\left[\frac{\partial q}{\partial \theta^B}\right] = E\left[\frac{\partial q}{\partial \theta^B}\left(\frac{\partial q}{\partial \theta^B}\right)'\right] - E\left[\frac{\partial q}{\partial \theta^B}\right]E\left[\frac{\partial q}{\partial \theta^B}\right]' \\
&= E\left[\begin{pmatrix} \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\begin{pmatrix} \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}'\right] - E\left[\begin{pmatrix} \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right]E\left[\begin{pmatrix} \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right]' \\
&= E\begin{bmatrix}
\frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & \\
\frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & \\
\frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \rho^B}\right)'
\end{bmatrix} - E\left[\begin{pmatrix} \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right]E\left[\begin{pmatrix} \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right]'
\end{aligned}
$$

The expectation of the elements of the vector of first derivatives are

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \beta_r^B}\right] &= E\left[-2\nu^{B\prime} x_r^B\right] = -2E\left[\nu^{B\prime}\right] x_r^B = 0 \\
E\left[\frac{\partial q}{\partial \delta_r^B}\right] &= E\left[-2\nu^{B\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\right] = -2E\left[\nu^{B\prime}\right](I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) = 0 \\
E\left[\frac{\partial q}{\partial \tau}\right] &= E\left[-2\nu^{B\prime}\Lambda\right] = -2E\left[\nu^{B\prime}\right]\Lambda = 0 \\
E\left[\frac{\partial q}{\partial \rho^B}\right] &= E\left[-2\nu^{B\prime}\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]\right] = -2E\left[\nu^{B\prime}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right] = 0.
\end{aligned}
$$

Thus Ω_{g^B} reduces to the expectation of the matrix of all cross-products of first derivatives and its elements are given by

$$
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] = E\left[-2\nu^{B\prime}x_r^B\left(-2\nu^{B\prime}x_s^B\right)'\right] = 4E\left[\nu^{B\prime}x_r^B x_s^{B\prime}\nu^B\right]
$$

Since the expression inside the expectation operator is a scalar, we make use again of the fact that the trace of a scalar is the trace itself and of the cyclic property of the trace:

$$
\begin{aligned}
&= 4E\left[Tr\left[\nu^{B\prime}x_r^B x_s^{B\prime}\nu^B\right]\right] \\
&= 4E\left[Tr\left[x_r^B x_s^{B\prime}\nu^B\nu^{B\prime}\right]\right] \\
&= 4Tr\left[x_r^B x_s^{B\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] &= E\left[-2\nu^{B\prime}x_r^B\left(-2\nu^{B\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\right)'\right] = 4E\left[\nu^{B\prime}x_r^B(\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))\nu^B\right] \\
&= 4E\left[Tr\left[\nu^{B\prime}x_r^B(\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))\nu^B\right]\right] = 4E\left[Tr\left[x_r^B(\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))E\left[\nu^B\nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right] \\
&= 4Tr\left[x_r^B(\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))E\left[E\left[\nu^B\nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= E\left[-2\nu^{B\prime}x_r^B\left(-2\nu^{B\prime}\Lambda\right)'\right] = 4E\left[\nu^{B\prime}x_r^B\Lambda'\nu^B\right] = 4E\left[Tr\left[\nu^{B\prime}x_r^B\Lambda'\nu^B\right]\right] \\
&= 4E\left[Tr\left[x_r^B\Lambda'\nu^B\nu^{B\prime}\right]\right] = 4Tr\left[x_r^B\Lambda' E\left[E\left[\nu^B\nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= E\left[-2\nu^{B\prime}x_r^B\left(-2\nu^{B\prime}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes \bar{x}^B)\delta^B+\tau\frac{\partial \Lambda}{\partial \rho^B}\right]\right)'\right] \\
&= 4E\left[\nu^{B\prime}x_r^B\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes \bar{x}^B)\delta^B+\tau\frac{\partial \Lambda}{\partial \rho^B}\right]'\nu^B\right] \\
&= 4E\left[Tr\left[\nu^{B\prime}x_r^B\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes \bar{x}^B)\delta^B+\tau\frac{\partial \Lambda}{\partial \rho^B}\right]'\nu^B\right]\right] \\
&= 4E\left[Tr\left[x_r^B\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes \bar{x}^B)\delta^B+\tau\frac{\partial \Lambda}{\partial \rho^B}\right]'\nu^B\nu^{B\prime}\right]\right] \\
&= 4Tr\left[x_r^B\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes \bar{x}^B)\delta^B+\tau\frac{\partial \Lambda}{\partial \rho^B}\right]'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right]
\end{aligned}
$$

Similarly, we can derive the remaining elements of matrix Ω_{g^B} for the sample selection case.

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] &= 4Tr\left[(I_T\otimes R_t^B)(\iota_T\otimes \bar{x}^B)x_s^{B\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] &= 4Tr\left[(I_T\otimes R_t^B)(\iota_T\otimes \bar{x}^B)(\iota_T\otimes \bar{x}^{B\prime})((I_T\otimes R_t^{B\prime}))E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= 4Tr\left[(I_T\otimes R_t^B)(\iota_T\otimes \bar{x}^B)\Lambda' E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= 4Tr\left[(I_T\otimes R_t^B)(\iota_T\otimes \bar{x}^B)\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes \bar{x}^B)\delta^B+\tau\frac{\partial \Lambda}{\partial \rho^B}\right]'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right]
\end{aligned}
$$

$$E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] = 4Tr\left[\Lambda x_s^{B\prime} E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] = 4Tr\left[\Lambda(\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \tau}\right)'\right] = 4Tr\left[\Lambda \Lambda' E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] = 4Tr\left[\Lambda\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]' E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] = 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right] x_s^{B\prime} E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] = 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right](\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] = 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]\Lambda' E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] = 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]' E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

$$
\begin{aligned}
\Omega_{g^{AB}} &= E\left[\frac{\partial l}{\partial \theta^A}\left(\frac{\partial q^B}{\partial \theta^B}\right)'\right] \\
&= E\left[\begin{pmatrix} \frac{\partial l}{\partial \beta_1^A} \\ \vdots \\ \frac{\partial l}{\partial \beta_R^A} \\ \frac{\partial l}{\partial \delta_1^A} \\ \vdots \\ \frac{\partial l}{\partial \delta_R^A} \\ \frac{\partial l}{\partial \rho^A} \end{pmatrix} \begin{pmatrix} \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}'\right] \\
&= E\begin{bmatrix}
\frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & \\
\frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & \\
\frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \rho^B}\right)'
\end{bmatrix}
\end{aligned}
$$

Thus $\Omega_{g^{AB}}$ is given by the expectation of the matrix of all cross-products of first derivatives of the selection and outcome equation and its elements are given by

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \beta_k^A}\left(\frac{\partial q}{\partial \beta_r^B}\right)'\right] &= E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\left(-2\nu^{B\prime}x_r^B\right)'\right] \\
&= -2E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A x_r^{B\prime}\nu^B\right] \\
&= -2E\left[Tr\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A x_r^{B\prime}\nu^B\right]\right] \\
&= -2E\left[Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A x_r^{B\prime}\nu^B u^{A\prime}\right]\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \beta_k^B}\left(\frac{\partial q}{\partial \delta_r^B}\right)'\right] &= E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\left(-2\nu^{B\prime}((I_T \otimes R_t^B))(\iota_T \otimes \bar{x}^B)\right)'\right] \\
&= -2E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A(\iota_T \otimes \bar{x}^B)'(I_T \otimes R_t^B)'\nu^B\right] \\
&= -2E\left[Tr\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A(\iota_T \otimes \bar{x}^B)'(I_T \otimes R_t^B)'\nu^B\right]\right] \\
&= -2E\left[Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A(\iota_T \otimes \bar{x}^B)'(I_T \otimes R_t^B)'\nu^B u^{A\prime}\right]\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A(\iota_T \otimes \bar{x}^B)'(I_T \otimes R_t^B)'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \beta_k^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\left(-2\nu^{B\prime}\Lambda\right)'\right] \\
&= -2E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\Lambda'\nu^B\right] \\
&= -2E\left[Tr\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\Lambda'\nu^B\right]\right] \\
&= -2E\left[Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\Lambda'\nu^B u^{A\prime}\right]\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\Lambda'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \beta_k^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= E\left[u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\left(-2\nu^{B\prime}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]\right)'\right] \\
&= -2E\left[u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'\nu^B\right] \\
&= -2E\left[Tr\left[u^{A\prime}(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'\nu^B\right]\right] \\
&= -2E\left[Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'\nu^B u^{A\prime}\right]\right] \\
&= -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes(R_t^A)^{-1})x_k^A\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]
\end{aligned}
$$

Similarly, we can derive the remaining elements of matrix $\Omega_{g^{AB}}$ for the sample selection case.

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \beta_r^B}\right)'\right] &= -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \delta_r^B}\right)'\right] &= -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)(\iota_T\otimes\bar{x}^B)'(I_T\otimes R_t^B)'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\Lambda' E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]
\end{aligned}
$$

The derivation of the following expectations require more details:

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \beta_r^B}\right)'\right] &= E\left[\left[-Tr\left(I_T \otimes R_t^A W_t\right) + u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left[(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right]\right]\left(-2\nu^{B\prime}x_r^B\right)'\right] \\
&= -2E\left[-Tr\left(I_T \otimes R_t^A W_t\right)x_r^{B\prime}\nu^B + u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)\left[(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A + u^A\right]x_r^{B\prime}\nu^B\right] \\
&= -2E[-Tr\left(I_T \otimes R_t^A W_t\right)x_r^{B\prime}\nu^B \\
&+ u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)(I_T \otimes R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A x_r^{B\prime}\nu^B + u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)u^A x_r^{B\prime}\nu^B] \\
&= -2E[-Tr\left(I_T \otimes R_t^A W_t\right)x_r^{B\prime}\nu^B \\
&+ Tr\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A x_r^{B\prime}\nu^B\right] + u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)u^A x_r^{B\prime}\nu^B] \\
&= -2E[-Tr\left(I_T \otimes R_t^A W_t\right)x_r^{B\prime}\nu^B \\
&+ Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A x_r^{B\prime}\nu^B u^{A\prime}\right] + u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)u^A x_r^{B\prime}\nu^B] \\
&= 2Tr\left(I_T \otimes R_t^A W_t\right)x_r^{B\prime}\underbrace{E\left[\nu^B\right]}_{=0} \\
&- 2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A x_r^{B\prime}E\left[\nu^B u^{A\prime}\right]\right] \\
&- 2E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)u^A x_r^{B\prime}\nu^B\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*} > 0, x^{A0}, x^B\right]\right]\right] \\
&- 2E\left[E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)u^A|\nu^B, x^{A0}, x^B\right]x_r^{B\prime}\nu^B\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*} > 0, x^{A0}, x^B\right]\right]\right] \\
&- 2\underbrace{E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t)u^A|\nu^B, x^{A0}, x^B\right]}_{\text{deterministic}}x_r^{B\prime}\underbrace{E\left[\nu^B\right]}_{=0} \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes W_t R_t^A)(\iota_T \otimes \bar{x}^A)\delta^A x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*} > 0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \delta_r^B}\right)'\right] &= -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_tR_t^A)(\iota_T\otimes\bar{x}^A)\delta^A(\iota_T\otimes\bar{x}^B)'(I_T\otimes R_t^B)'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_tR_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\Lambda'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= -2Tr[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_tR_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\left[(I_T\otimes R_t^BW_tR_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'\times \\
&\times\;\; E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]]
\end{aligned}
$$

Treatment Selection

$$
\begin{aligned}
\Omega_{g^B} &= Var\left[\frac{\partial q}{\partial \theta^B}\right] = E\left[\frac{\partial q}{\partial \theta^B}\left(\frac{\partial q}{\partial \theta^B}\right)'\right] - E\left[\frac{\partial q}{\partial \theta^B}\right] E\left[\frac{\partial q}{\partial \theta^B}\right]' \\
&= E\left[\begin{pmatrix} \frac{\partial q}{\partial \alpha} \\ \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\begin{pmatrix} \frac{\partial q}{\partial \alpha} \\ \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}'\right] - E\left[\begin{pmatrix} \frac{\partial q}{\partial \alpha} \\ \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right] E\left[\begin{pmatrix} \frac{\partial q}{\partial \alpha} \\ \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right]' \\
&= E\begin{bmatrix}
\frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \beta_1^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & & \\
\frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \beta_R^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \delta_1^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & & \\
\frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \delta_R^B}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \rho^B}\right)'
\end{bmatrix} -
\end{aligned}
$$

$$-E\left[\begin{pmatrix} \frac{\partial q}{\partial \alpha} \\ \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right] E\left[\begin{pmatrix} \frac{\partial q}{\partial \alpha} \\ \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}\right]'$$

The expectation of the elements of the vector of first derivatives are

$$E\left[\frac{\partial q}{\partial \alpha}\right] = E\left[-2\nu^{B\prime} y^A\right] = -2E\left[\nu^{B\prime} y^A\right] = -2E\left[\nu^{B\prime}\right] E\left[y^A\right] = 0$$

The second to last equality follows from the fact that after correcting for selection into treatment, the treatment indicator and the residual of the second stage are independent.

$$E\left[\frac{\partial q}{\partial \beta_r^B}\right] = E\left[-2\nu^{B\prime} x_r^B\right] = -2E\left[\nu^{B\prime}\right] x_r^B = 0$$

$$E\left[\frac{\partial q}{\partial \delta_r^B}\right] = E\left[-2\nu^{B\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\right] = -2E\left[\nu^{B\prime}\right](I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) = 0$$

$$E\left[\frac{\partial q}{\partial \tau}\right] = E\left[-2\nu^{B\prime}\Lambda^g\right] = -2E\left[\nu^{B\prime}\right]\Lambda^g = 0$$

$$E\left[\frac{\partial q}{\partial \rho^B}\right] = E\left[-2\nu^{B\prime}\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]\right] = -2E\left[\nu^{B\prime}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right] = 0.$$

Thus Ω_{g^B} reduces to the expectation of the matrix of all cross-products of first derivatives and its elements are given by

$$E\left[\frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \alpha}\right)'\right] = E\left[-2\nu^{B\prime} y^A\left(-2\nu^{B\prime} y^A\right)'\right] = 4E\left[\nu^{B\prime} y^A y^{A\prime}\nu^B\right] = 4E\left[Tr\left[\nu^{B\prime} y^A y^{A\prime}\nu^B\right]\right] = 4E\left[Tr\left[y^A y^{A\prime}\nu^B\nu^{B\prime}\right]\right]$$

Since ν^B and y^A are independent:

$$= 4Tr\left[E\left[y^A y^{A\prime}\right]E\left[\nu^B \nu^{B\prime}\right]\right] = 4Tr\left[E\left[y^A y^{A\prime}\right]E\left[E\left[\nu^B \nu^{B\prime} | y^{A*} > 0, x^{A0}, x^B\right]\right]\right]$$

From the properties of the indicator function we know that

$$\begin{aligned} E\left[y^A_{ti}\right] &= Pr(y^A_{ti} = 1) = \sigma_{u^A_{ti}}\Phi(z_{ti}) = \tilde{\Phi}(z_{ti}), \quad Var\left[y^A_{ti}\right] = Pr(y^A_{ti} = 1)(1 - Pr(y^A_{ti} = 1)) = \tilde{\Phi}(z_{ti})(1 - \tilde{\Phi}(z_{ti})), \\ \text{and} \quad Cov\left[y^A_{ti}, y^A_{sj}\right] &= P(y^A_{ti} = 1 \cap y^A_{sj} = 1) - P(y^A_{ti} = 1)P(y^A_{sj} = 1) = P(y^{A*}_{ti} > 0 \cap y^{A*}_{sj} > 0) - \tilde{\Phi}(z_{ti})\tilde{\Phi}(z_{sj}) \\ &= (1 - F(u^A_{ti}, u^A_{sj})) - \tilde{\Phi}(z_{ti})\tilde{\Phi}(z_{sj}), \end{aligned}$$

where $F(\cdot)$ denotes the cumulative joint bivariate normal distribution.
Since $Var\left[y^A_{ti}\right] = E\left[(y^A_{ti})^2\right] - E\left[y^A_{ti}\right]E\left[y^A_{ti}\right]$ it follows

$$E\left[(y^A_{ti})^2\right] = Var\left[y^A_{ti}\right] + E\left[y^A_{ti}\right]E\left[y^A_{ti}\right] = \tilde{\Phi}(z_{ti})(1 - \tilde{\Phi}(z_{ti})) + \tilde{\Phi}(z_{ti})\tilde{\Phi}(z_{ti}) = \tilde{\Phi}(z_{ti}) - \tilde{\Phi}(z_{ti})\tilde{\Phi}(z_{ti}) + \tilde{\Phi}(z_{ti})\tilde{\Phi}(z_{ti}) = \tilde{\Phi}(z_{ti}).$$

Finally, $E\left[y^A_{ti} y^A_{sj}\right]$ is given by

$$\begin{aligned} E\left[y^A_{ti} y^A_{sj}\right] &= Cov\left[y^A_{ti}, y^A_{sj}\right] + E\left[y^A_{ti}\right]E\left[y^A_{sj}\right] = \\ &= (1 - F(u^A_{ti}, u^A_{sj})) \end{aligned}$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \beta_r^B}\right)'\right] &= E\left[-2\nu^{B\prime}y^A\left(-2\nu^{B\prime}x_r^B\right)'\right] = 4E\left[\nu^{B\prime}y^A x_r^{B\prime}\nu^B\right] = 4E\left[Tr\left[\nu^{B\prime}y^A x_r^{B\prime}\nu^B\right]\right] = 4Tr\left[E\left[y^A\right]x_r^{B\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
&= 4Tr\left[\tilde{\Phi}(z)x_r^{B\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \delta_r^B}\right)'\right] &= E\left[-2\nu^{B\prime}y^A\left(-2\nu^{B\prime}(I_T\otimes R_t^B)(\iota_T\otimes\bar{x}^B)\right)'\right] = 4Tr\left[\tilde{\Phi}(z)(\iota_T\otimes\bar{x}^B)'(I_T\otimes R_t^B)'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= E\left[-2\nu^{B\prime}y^A\left(-2\nu^{B\prime}\Lambda^g\right)'\right] = 4Tr\left[\tilde{\Phi}(z)\Lambda^{g\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \alpha}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= E\left[-2\nu^{B\prime}y^A\left(-2\nu^{B\prime}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\right)'\right] \\
&= 4Tr\left[\tilde{\Phi}(z)\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \alpha}\right)'\right] &= 4Tr\left[x_r^B\tilde{\Phi}(z)'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] &= 4Tr\left[x_r^B x_s^{B\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] &= 4Tr\left[x_r^B(\iota_T\otimes\bar{x}^{B\prime})((I_T\otimes R_t^{B\prime}))E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= 4Tr\left[x_r^B\Lambda^{g\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \beta_r^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= 4Tr\left[x_r^B\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0,x^{A0},x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \alpha}\right)'\right] &= 4Tr\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\tilde{\Phi}(z)'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] &= 4Tr\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)x_s^{B\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] &= 4Tr\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)(\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= 4Tr\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\Lambda^{g\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \delta_r^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= 4Tr\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]' E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \alpha}\right)'\right] &= 4Tr\left[\Lambda^g\tilde{\Phi}(z)'E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] &= 4Tr\left[\Lambda^g x_s^{B\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] &= 4Tr\left[\Lambda^g(\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime}))E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= 4Tr\left[\Lambda^g\Lambda^{g\prime}E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \tau}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= 4Tr\left[\Lambda^g\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]' E\left[E\left[\nu^B\nu^{B\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial alpha}\right)'\right] &= 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]\tilde{\Phi}(z)' E\left[E\left[\nu^B \nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \beta_s^B}\right)'\right] &= 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right] x_s^{B\prime} E\left[E\left[\nu^B \nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \delta_s^B}\right)'\right] &= 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right](\iota_T \otimes \bar{x}^{B\prime})((I_T \otimes R_t^{B\prime})) E\left[E\left[\nu^B \nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]\Lambda^{g\prime} E\left[E\left[\nu^B \nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial q}{\partial \rho^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= 4Tr\left[\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]' E\left[E\left[\nu^B \nu^{B\prime}|y^{A*} > 0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
\Omega_{g^{AB}} &= E\left[\frac{\partial l}{\partial \theta^A}\left(\frac{\partial q^B}{\partial \theta^B}\right)'\right] \\
&= E\left[\begin{pmatrix} \frac{\partial l}{\partial \beta_1^A} \\ \vdots \\ \frac{\partial l}{\partial \beta_R^A} \\ \frac{\partial l}{\partial \delta_1^A} \\ \vdots \\ \frac{\partial l}{\partial \delta_R^A} \\ \frac{\partial l}{\partial \rho^A} \end{pmatrix}\begin{pmatrix} \frac{\partial q}{\partial \alpha} \\ \frac{\partial q}{\partial \beta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \beta_R^B} \\ \frac{\partial q}{\partial \delta_1^B} \\ \vdots \\ \frac{\partial q}{\partial \delta_R^B} \\ \frac{\partial q}{\partial \tau} \\ \frac{\partial q}{\partial \rho^B} \end{pmatrix}'\right] \\
&= E\begin{bmatrix}
\frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial}{\partial \beta_1^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & & \\
\frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \beta_K^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \delta_1^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\vdots & & & & & & & & \\
\frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \delta_K^A}\left(\frac{\partial q}{\partial \rho^B}\right)' \\
\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \alpha}\right)' & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \beta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \beta_R^B}\right)' & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \delta_1^B}\right)' & \cdots & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \delta_R^B}\right)' & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \tau}\right)' & \frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \rho^B}\right)'
\end{bmatrix}
\end{aligned}
$$

Thus $\Omega_{g^{AB}}$ is given by the expectation of the matrix of all cross-products of first derivatives of the selection and outcome equation and its elements are given by

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \beta_k^A}\left(\frac{\partial q}{\partial \alpha}\right)'\right] &= -2E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\left(-2\nu^{B\prime}y^A\right)'\right] \\
&= -2E\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A y^{A\prime}\nu^B\right] \\
&= -2E\left[Tr\left[u^{A\prime}(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A y^{A\prime}\nu^B\right]\right] \\
&= -2E\left[Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A y^{A\prime}\nu^B u^{A\prime}\right]\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A E\left[y^{A\prime}\nu^B u^{A\prime}\right]\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A E\left[y^{A\prime}\underbrace{E\left[\nu^B u^{A\prime}|y^{A*}>0, x^{A0}, x^B\right]}_{deterministic}\right]\right] \\
&= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\tilde{\Phi}(z)'E\left[\nu^B u^{A\prime}|y^{A*}>0, x^{A0}, x^B\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial l}{\partial \beta_k^A}\left(\frac{\partial q}{\partial \beta_r^B}\right)'\right] &= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \beta_k^B}\left(\frac{\partial q}{\partial \delta_r^B}\right)'\right] &= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A(\iota_T \otimes \bar{x}^B)'(I_T \otimes R_t^B)'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \beta_k^B}\left(\frac{\partial q}{\partial \tau}\right)'\right] &= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\Lambda^{g\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0, x^{A0}, x^B\right]\right]\right] \\
E\left[\frac{\partial l}{\partial \beta_k^B}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= -2Tr\left[(I_T \otimes (R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T \otimes (R_t^A)^{-1})x_k^A\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0, x^{A0}, x^B\right]\right]\right]
\end{aligned}
$$

Similarly, we can derive the remaining elements of matrix $\Omega_{g^{AB}}$ for the sample selection case.

$$E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \alpha}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\tilde{\Phi}(z)'E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]$$

$$E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \beta_r^B}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]$$

$$E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \delta_r^B}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)(\iota_T\otimes\bar{x}^B)'(I_T\otimes R_t^B)'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]$$

$$E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \tau}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\Lambda^{g\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]$$

$$E\left[\frac{\partial l}{\partial \delta_k^A}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(\iota_T\otimes\bar{x}_k^A)\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]' E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]$$

$$E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \alpha}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\tilde{\Phi}(z)\prime E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]$$

$$E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \beta_r^B}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A x_r^{B\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]$$

$$E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \delta_r^B}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A(\iota_T\otimes\bar{x}^B)'(I_T\otimes R_t^B)'E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]$$

$$E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \tau}\right)'\right] = -2Tr\left[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\Lambda^{g\prime}E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]\right]$$

$$\begin{aligned}E\left[\frac{\partial l}{\partial \rho^A}\left(\frac{\partial q}{\partial \rho^B}\right)'\right] &= -2Tr[(I_T\otimes(R_t^A)^{-1})'\Omega_{\xi^A}^{-1}(I_T\otimes W_t R_t^A)(\iota_T\otimes\bar{x}^A)\delta^A\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]'\times\\ &\times\ E\left[E\left[\nu^B u^{A\prime}|y_{ti}^{A*}>0,x^{A0},x^B\right]\right]]\end{aligned}$$

3.3 Spatial/Network Truncated Variance-covariance Matrix

3.3.1 Ω_{g^B}

The elements of matrix Ω_{g^B} are functions of the following conditional expectations (see Appendix 3.2.2 for the specific elements of this matrix):

$$
\begin{aligned}
&E\left[(\nu_{ti}^B)^2|y_{ti}^{A*}>0,x^{A0},x^B\right], \quad E\left[\nu_{ti}^B\nu_{tj}^B|y_{ti}^{A*}>0,x^{A0},x^B\right]\\
&E\left[\nu_{ti}^B\nu_{si}^B|y_{ti}^{A*}>0,x^{A0},x^B\right], \quad E\left[\nu_{ti}^B\nu_{sj}^B|y_{ti}^{A*}>0,x^{A0},x^B\right]
\end{aligned}
$$

To calculate these expressions, we use the following theorems (adapted to our model):

Theorem 1: Variance of the incidentally truncated bivariate normal distribution (Greene 2008, p. 883; using our notation and adapting to panel data case)

$$
\begin{aligned}
E\left[(\nu_{ti}^B)^2|y_{ti}^{A*}>0,x^{A0},x^B\right] &= Var\left[\nu_{ti}^B|y_{ti}^{A*}>0,x^{A0},x^B\right]\\
&= Var\left[u_{ti}^B\right]\left[\varrho_{|t=s,i=j}^{BB}-(\varrho_{|t=s,i=j}^{AB})^2\zeta_{ti}\right]
\end{aligned}
$$

Using our assumptions on the spatial/network error components:

$$
\begin{aligned}
&= \sigma_{\xi^B}^2\sum_{j=1}^N\left(r_{tij}^B\right)^2\left[1-\left(\frac{\sigma_{\xi^{AB}}\sum_{j=1}^N r_{tij}^A r_{tij}^B}{\sqrt{\sigma_{\xi^A}^2\sum_{j=1}^N\left(r_{tij}^A\right)^2}\sqrt{\sigma_{\xi^B}^2\sum_{j=1}^N\left(r_{tij}^B\right)^2}}\right)^2\zeta_{ti}\right]\\
&= \sigma_{\xi^B}^2\sum_{j=1}^N\left(r_{tij}^B\right)^2-\left(\frac{\sigma_{\xi^{AB}}\sum_{j=1}^N r_{tij}^A r_{tij}^B}{\sqrt{\sigma_{\xi^A}^2\sum_{j=1}^N\left(r_{tij}^A\right)^2}}\right)^2\zeta_{ti}\\
&= \sigma_{\xi^B}^2\sum_{j=1}^N\left(r_{tij}^B\right)^2-\tau^2\psi_{ti}^2\zeta_{ti}
\end{aligned}
$$

Theorem 2: Covariance of the incidentally truncated trivariate normal distribution (Kotz, Balakrishnan, & Johnson, 2000, pp. 317-318; using our notation and adapting to panel data case)

$$
\begin{aligned}
& E\left[\nu_{ti}^{B}\nu_{tj}^{B}|y_{ti}^{A*}>0,x^{A0},x^{B}\right] \\
= \; & Cov\left[\nu_{ti}^{B}\nu_{tj}^{B}|y_{ti}^{A*}>0,x^{A0},x^{B}\right] \\
= \; & \sqrt{Var\left[u_{ti}^{B}\right]}\sqrt{Var\left[u_{tj}^{B}\right]}\left[\varrho^{BB}_{|t=s,i\neq j}-\varrho^{AB}_{|t=s,i=j}\varrho^{AB}_{|t=s,i\neq j}\zeta_{ti}\right]
\end{aligned}
$$

Using our assumptions on the spatial/network error components:

$$
\begin{aligned}
= \; & \sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{tji}\right)^2} \\
\times \; & \left[\frac{\sigma^2_{\xi^B}\sum_{m=1}^{N}r^B_{tim}r^B_{tjm}}{\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{tji}\right)^2}}-\frac{\sigma_{\xi^{AB}}\sum_{j=1}^{N}r^A_{tij}r^B_{tij}}{\sqrt{\sigma^2_{\xi^A}\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}}\frac{\sigma_{\xi^{AB}}\sum_{m=1}^{N}r^A_{tim}r^B_{tjm}}{\sqrt{\sigma^2_{\xi^A}\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{tji}\right)^2}}\zeta_{ti}\right] \\
= \; & \sigma^2_{\xi^B}\sum_{m=1}^{N}r^B_{tim}r^B_{tjm}-\left(\frac{\sigma_{\xi^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\right)^2\frac{\sum_{j=1}^{N}r^A_{tij}r^B_{tij}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\frac{\sum_{m=1}^{N}r^A_{tim}r^B_{tjm}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\zeta_{ti} \\
= \; & \sigma^2_{\xi^B}\sum_{m=1}^{N}r^B_{tim}r^B_{tjm}-\tau^2\psi_{ti}\frac{\sum_{m=1}^{N}r^A_{tim}r^B_{tjm}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\zeta_{ti}
\end{aligned}
$$

with $\zeta_{ti}=\lambda_{ti}^2+\frac{x_{ti}^{A\prime}\beta^A+\sum_{j=1}^{N}r^A_{tij}\bar{x}_j^{A\prime}\delta^A}{\sqrt{\sigma^2_{\xi^A}\sum_{j}^{N}(r^A_{tij})^2}}\lambda_{ti}$ and $\varrho^{AB}=\frac{Cov\left[u^A_{ti},u^B_{sj}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}}$.

Similarly:

$$
\begin{aligned}
& E\left[\nu_{ti}^{B}\nu_{si}^{B}|y_{ti}^{A*}>0,x^{A0},x^{B}\right] \\
= \; & Cov\left[\nu_{ti}^{B}\nu_{si}^{B}|y_{ti}^{A*}>0,x^{A0},x^{B}\right] \\
= \; & \sqrt{Var\left[u_{ti}^{B}\right]}\sqrt{Var\left[u_{si}^{B}\right]}\left[\varrho^{BB}_{|t\neq s,i=j}-\varrho^{AB}_{|t=s,i=j}\varrho^{AB}_{|t\neq s,i=j}\zeta_{ti}\right]
\end{aligned}
$$

Again, using our assumptions on the spatial/network error components:

$$
\begin{aligned}
&= \sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{sij}\right)^2} \\
&\times \left[\frac{\sigma^2_{\mu^B}\sum_{j=1}^{N} r^B_{tij} r^B_{sij}}{\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{sij}\right)^2}} - \frac{\sigma_{\xi^{AB}}\sum_{j=1}^{N} r^A_{tij} r^B_{tij}}{\sqrt{\sigma^2_{\xi^A}\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}}\frac{\sigma_{\mu^{AB}}\sum_{j=1}^{N} r^A_{tij} r^B_{sij}}{\sqrt{\sigma^2_{\xi^A}\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{sij}\right)^2}}\zeta_{ti}\right] \\
&= \sigma^2_{\mu^B}\sum_{j=1}^{N} r^B_{tij} r^B_{sij} - \frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sigma_{\xi^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sum_{j=1}^{N} r^A_{tij} r^B_{tij}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\frac{\sum_{j=1}^{N} r^A_{tij} r^B_{sij}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\zeta_{ti} \\
&= \sigma^2_{\mu^B}\sum_{j=1}^{N} r^B_{tij} r^B_{sij} - \tau\psi_{ti}\frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sum_{j=1}^{N} r^A_{tij} r^B_{sij}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\zeta_{ti}
\end{aligned}
$$

$$
\begin{aligned}
& E\left[\nu^B_{ti}\nu^B_{sj}|y^{A*}_{ti} > 0, x^{A0}, x^B\right] \\
&= Cov\left[\nu^B_{ti}\nu^B_{sj}|y^{A*}_{ti} > 0, x^{A0}, x^B\right] \\
&= \sqrt{Var\left[u^B_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}\left[\varrho^{BB}_{|t\neq s, i\neq j} - \varrho^{AB}_{|t=s, i=j}\varrho^{AB}_{|t\neq s, i\neq j}\zeta_{ti}\right]
\end{aligned}
$$

Again, using our assumptions on the spatial/network error components:

$$
\begin{aligned}
&= \sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{sji}\right)^2} \\
&\times \left[\frac{\sigma^2_{\mu^B}\sum_{m=1}^{N} r^B_{tim} r^B_{sjm}}{\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{sji}\right)^2}} - \frac{\sigma_{\xi^{AB}}\sum_{j=1}^{N} r^A_{tij} r^B_{tij}}{\sqrt{\sigma^2_{\xi^A}\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}} \frac{\sigma_{\mu^{AB}}\sum_{m=1}^{N} r^A_{tim} r^B_{sjm}}{\sqrt{\sigma^2_{\xi^A}\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{sji}\right)^2}}\zeta_{ti}\right] \\
&= \sigma^2_{\mu^B}\sum_{m=1}^{N} r^B_{tim} r^B_{sjm} - \frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sigma_{\xi^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sum_{j=1}^{N} r^A_{tij} r^B_{tij}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\frac{\sum_{m=1}^{N} r^A_{tim} r^B_{sjm}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\zeta_{ti} \\
&= \sigma^2_{\mu^B}\sum_{m=1}^{N} r^B_{tim} r^B_{sjm} - \tau\psi_{ti}\frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma^2_{\xi^A}}}\frac{\sum_{m=1}^{N} r^A_{tim} r^B_{sjm}}{\sqrt{\sum_{j=1}^{N}\left(r^A_{tij}\right)^2}}\zeta_{ti}
\end{aligned}
$$

3.3.2 $\Omega_{g^{AB}}$

The elements of matrix $\Omega_{g^{AB}}$ are functions of the following conditional expectations:

$$
\begin{gathered}
E\left[u^A_{ti}\nu^B_{ti}|y^{A*}_{ti}>0,x^{A0},x^B\right], \quad E\left[u^A_{ti}\nu^B_{tj}|y^{A*}_{ti}>0,x^{A0},x^B\right] \\
E\left[u^A_{ti}\nu^B_{si}|y^{A*}_{ti}>0,x^{A0},x^B\right], \quad E\left[u^A_{ti}\nu^B_{sj}|y^{A*}_{ti}>0,x^{A0},x^B\right]
\end{gathered}
$$

These expressions can be derived base on the following theorem:

Theorem 3: Covariance of the incidentally truncated bivariate normal distribution (Kotz, Balakrishnan, & Johnson, 2000, pp. 311-312; using our notation and adapting to panel data case)

$$
\begin{aligned}
E\left[u^A_{ti}\nu^B_{ti}|y^{A*}_{ti}>0,x^{A0},x^B\right] &= Cov\left[u^A_{ti},\nu^B_{ti}|y^{A*}_{ti}>0,x^{A0},x^B\right] \\
&= Cov\left[u^A_{ti},u^B_{ti}\right]\left[1-\zeta_{ti}\right]
\end{aligned}
$$

Using our assumptions on the spatial/network error components:

$$= \sigma_{\xi^{AB}} \sum_{j=1}^{N} r_{tij}^{A} r_{tij}^{B} \left[1 - \zeta_{ti}\right]$$

And similarly:

$$\begin{aligned} E\left[u_{ti}^{A} \nu_{tj}^{B} | y_{ti}^{A*} > 0, x^{A0}, x^{B}\right] &= Cov\left[u_{ti}^{A}, \nu_{tj}^{B} | y_{ti}^{A*} > 0, x^{A0}, x^{B}\right] \\ &= Cov\left[u_{ti}^{A}, u_{tj}^{B}\right]\left[1 - \zeta_{ti}\right] \end{aligned}$$

Again, using our assumptions on the spatial/network error components:

$$= \sigma_{\xi^{AB}} \sum_{m=1}^{N} r_{tim}^{A} r_{tjm}^{B} \left[1 - \zeta_{ti}\right]$$

$$\begin{aligned} E\left[u_{ti}^{A} \nu_{si}^{B} | y_{ti}^{A*} > 0, x^{A0}, x^{B}\right] &= Cov\left[u_{ti}^{A}, \nu_{si}^{B} | y_{ti}^{A*} > 0, x^{A0}, x^{B}\right] \\ &= Cov\left[u_{ti}^{A}, u_{si}^{B}\right]\left[1 - \zeta_{ti}\right] \end{aligned}$$

Again, using our assumptions on the spatial/network error components:

$$= \sigma_{\mu^{AB}} \sum_{j=1}^{N} r_{tij}^{A} r_{sij}^{B} \left[1 - \zeta_{ti}\right]$$

$$\begin{aligned} E\left[u_{ti}^{A} \nu_{sj}^{B} | y_{ti}^{A*} > 0, x^{A0}, x^{B}\right] &= Cov\left[u_{ti}^{A}, \nu_{sj}^{B} | y_{ti}^{A*} > 0, x^{A0}, x^{B}\right] \\ &= Cov\left[u_{ti}^{A}, u_{sj}^{B}\right]\left[1 - \zeta_{ti}\right] \end{aligned}$$

Again, using our assumptions on the spatial/network error components:

$$= \sigma_{\mu^{AB}} \sum_{m=1}^{N} r_{tim}^{A} r_{sjm}^{B} \left[1 - \zeta_{ti}\right]$$

3.3.3 Derivation of ϱ^{BB} and ϱ^{AB}

$$
\begin{aligned}
\varrho^{AB}_{|t=s,i=j} &= \frac{Cov\left[u^A_{ti}, u^B_{ti}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{ti}\right]}} = \frac{E\left[u^A_{ti}u^B_{ti}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{ti}\right]}} = \frac{E\left[\sum_{j=1}^N r^A_{tij}(\mu^A_j + \varepsilon^A_{tj}) \sum_{i=1}^N r^B_{tij}(\mu^B_j + \varepsilon^B_{tj})\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{ti}\right]}} \\
&= \frac{E\left[\left(r^A_{ti1}(\mu^A_1 + \varepsilon^A_{t1}) + r^A_{ti2}(\mu^A_2 + \varepsilon^A_{t2}) + ... + r^A_{tiN}(\mu^A_N + \varepsilon^A_{tN})\right)\left(r^B_{ti1}(\mu^B_1 + \varepsilon^B_{t1}) + r^B_{ti2}(\mu^B_2 + \varepsilon^B_{t2}) + ... + r^B_{tiN}(\mu^B_N + \varepsilon^B_{tN})\right)\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{ti}\right]}} \\
&= \frac{\sigma_{\xi^{AB}} \sum_{j=1}^N r^A_{tij} r^B_{tij}}{\sqrt{\sigma^2_{\xi^A} \sum_{j=1}^N \left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B} \sum_{j=1}^N \left(r^B_{tij}\right)^2}} \\
\varrho^{AB}_{|t=s,i\neq j} &= \frac{Cov\left[u^A_{ti}, u^B_{tj}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{tj}\right]}} = \frac{E\left[u^A_{ti}u^B_{tj}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{tj}\right]}} = \frac{E\left[\sum_{j=1}^N r^A_{tij}(\mu^A_j + \varepsilon^A_{tj}) \sum_{i=1}^N r^B_{tji}(\mu^B_i + \varepsilon^B_{ti})\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{tj}\right]}} \\
&= \frac{E\left[\left(r^A_{ti1}(\mu^A_1 + \varepsilon^A_{t1}) + r^A_{ti2}(\mu^A_2 + \varepsilon^A_{t2}) + ... + r^A_{tiN}(\mu^A_N + \varepsilon^A_{tN})\right)\left(r^B_{tj1}(\mu^B_1 + \varepsilon^B_{t1}) + r^B_{tj2}(\mu^B_2 + \varepsilon^B_{t2}) + ... + r^B_{tjN}(\mu^B_N + \varepsilon^B_{tN})\right)\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{tj}\right]}} \\
&= \frac{\sigma_{\xi^{AB}} \sum_{m=1}^N r^A_{tim} r^B_{tjm}}{\sqrt{\sigma^2_{\xi^A} \sum_{j=1}^N \left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B} \sum_{i=1}^N \left(r^B_{tji}\right)^2}} \\
\varrho^{AB}_{|t\neq s,i=j} &= \frac{Cov\left[u^A_{ti}, u^B_{si}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{si}\right]}} = \frac{E\left[u^A_{ti}u^B_{si}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{si}\right]}} = \frac{E\left[\sum_{j=1}^N r^A_{tij}(\mu^A_j + \varepsilon^A_{tj}) \sum_{i=1}^N r^B_{sij}(\mu^B_j + \varepsilon^B_{sj})\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}} \\
&= \frac{E\left[\left(r^A_{ti1}(\mu^A_1 + \varepsilon^A_{t1}) + r^A_{ti2}(\mu^A_2 + \varepsilon^A_{t2}) + ... + r^A_{tiN}(\mu^A_N + \varepsilon^A_{tN})\right)\left(r^B_{si1}(\mu^B_1 + \varepsilon^B_{s1}) + r^B_{si2}(\mu^B_2 + \varepsilon^B_{s2}) + ... + r^B_{siN}(\mu^B_N + \varepsilon^B_{sN})\right)\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{si}\right]}} \\
&= \frac{\sigma_{\mu^{AB}} \sum_{j=1}^N r^A_{tij} r^B_{sij}}{\sqrt{\sigma^2_{\xi^A} \sum_{j=1}^N \left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B} \sum_{j=1}^N \left(r^B_{sij}\right)^2}} \\
\varrho^{AB}_{|t\neq s,i\neq j} &= \frac{Cov\left[u^A_{ti}, u^B_{sj}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}} = \frac{E\left[u^A_{ti}u^B_{sj}\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}} = \frac{E\left[\sum_{j=1}^N r^A_{tij}(\mu^A_j + \varepsilon^A_{tj}) \sum_{i=1}^N r^B_{sji}(\mu^B_i + \varepsilon^B_{si})\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}} \\
&= \frac{E\left[\left(r^A_{ti1}(\mu^A_1 + \varepsilon^A_{t1}) + r^A_{ti2}(\mu^A_2 + \varepsilon^A_{t2}) + ... + r^A_{tiN}(\mu^A_N + \varepsilon^A_{tN})\right)\left(r^B_{sj1}(\mu^B_1 + \varepsilon^B_{s1}) + r^B_{sj2}(\mu^B_2 + \varepsilon^B_{s2}) + ... + r^B_{sjN}(\mu^B_N + \varepsilon^B_{sN})\right)\right]}{\sqrt{Var\left[u^A_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}} \\
&= \frac{\sigma_{\mu^{AB}} \sum_{m=1}^N r^A_{tim} r^B_{sjm}}{\sqrt{\sigma^2_{\xi^A} \sum_{j=1}^N \left(r^A_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B} \sum_{i=1}^N \left(r^B_{sji}\right)^2}}
\end{aligned}
$$

And similarly $\varrho^{BB} = \frac{Cov\left[u^B_{ti}, u^B_{sj}\right]}{\sqrt{Var\left[u^B_{ti}\right]}\sqrt{Var\left[u^B_{sj}\right]}}$, and hence

$$
\begin{aligned}
\varrho^{BB}_{|t=s,i=j} &= \frac{\sigma^2_{\xi^B}\sum_{j=1}^{N}(r^B_{tij})^2}{\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}} = 1 \\
\varrho^{BB}_{|t=s,i\neq j} &= \frac{\sigma^2_{\xi^B}\sum_{m=1}^{N} r^B_{tim} r^B_{tjm}}{\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{tji}\right)^2}} \\
\varrho^{BB}_{|t\neq s,i=j} &= \frac{\sigma^2_{\mu^B}\sum_{j=1}^{N} r^B_{tij} r^B_{sij}}{\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{sij}\right)^2}} \\
\varrho^{BB}_{|t\neq s,i\neq j} &= \frac{\sigma^2_{\mu^B}\sum_{m=1}^{N} r^B_{tim} r^B_{sjm}}{\sqrt{\sigma^2_{\xi^B}\sum_{j=1}^{N}\left(r^B_{tij}\right)^2}\sqrt{\sigma^2_{\xi^B}\sum_{i=1}^{N}\left(r^B_{sji}\right)^2}}
\end{aligned}
$$

3.3.4 Estimation of $\hat{\sigma}^2_{\xi^B}$, $\hat{\sigma}_{\xi^{AB}}$, $\hat{\sigma}^2_{\mu^B}$, and $\hat{\sigma}_{\mu^{AB}}$

Note that for each individual i at time t, the true truncated variance of the residual ν^B_{ti} is given by (13):

$$\sigma^2_{\nu^B_{ti}} = \sigma^2_{\xi^B} \sum_{j=1}^{N} (r^B_{tij})^2 - \tau^2 \psi^2_{ti} \zeta_{ti}$$

The sample average of the truncated variance converges in probability to

$$\frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \hat{\sigma}^2_{\nu^B_{ti}} \xrightarrow{p} \sigma^2_{\xi^B} \overline{\sum_{j=1}^{N} (r^B_j)^2} - \tau^2 \overline{\psi^2 \zeta},$$

since

$$\frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \left[\sum_{j=1}^{N} (\hat{r}^B_{tij})^2 \right] \xrightarrow{p} \overline{\sum_{j=1}^{N} (r^B_j)^2} \quad \text{and}$$

$$\frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\hat{\psi}^2_{ti} \hat{\zeta}_{ti} \right) \xrightarrow{p} \overline{\psi^2 \zeta}.$$

We estimate $\frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \hat{\sigma}^2_{\nu^B_{ti}}$ by $\frac{\sum_{t=1}^{T} \sum_{i=1}^{N} (\hat{\nu}^B_{ti})^2}{TN}$ and solve the following equation for $\hat{\sigma}_{\xi^B}$:

$$\frac{\sum_{t=1}^{T} \sum_{j=1}^{N} (\hat{\nu}^B_{ti})^2}{TN} = \hat{\sigma}^2_{\xi^B} \widehat{\overline{\sum_{j=1}^{N} (r^B_j)^2}} - \hat{\tau}^2 \widehat{\overline{\psi^2 \zeta}}$$

Therefore, the estimates of the variance of the error components of the second stage $\hat{\sigma}^2_{\xi^B}$ and covariance of the error components of the first and the second stage $\hat{\sigma}_{\xi^{AB}}$ are given by

$$\hat{\sigma}^2_{\xi^B} = \frac{1}{\widehat{\overline{\sum_{j=1}^{N} (r^B_j)^2}}} \left(\frac{\sum_{t=1}^{T} \sum_{j=1}^{N} (\hat{\nu}^B_{ti})^2}{TN} + \hat{\tau}^2 \widehat{\overline{\psi^2 \zeta}} \right) \quad \text{and}$$

$$\hat{\sigma}_{\xi^{BA}} = \frac{\hat{\tau}}{\sqrt{\hat{\sigma}^2_{\xi^B}}}$$

Next, in order to estimate $\sigma^2_{\mu^B}$, and $\sigma_{\mu^{AB}}$ we follow a similar procedure as before but based on (14) and (15). For each individual i at time t and s, the true truncated covariance of residuals u^A_{ti} and ν^B_{si} is given by (15):

$$\sigma_{u^A_{ti} \nu^B_{si}} = \sigma_{\mu^{AB}} \sum_{j=1}^{N} r^A_{tij} r^B_{sij} \left[1 - \zeta_{ti} \right] = \sigma_{\mu^{AB}} \sum_{j=1}^{N} r^A_{tij} r^B_{sij} - \sigma_{\mu^{AB}} \sum_{j=1}^{N} r^A_{tij} r^B_{sij} \zeta_{ti}$$

The sample average of the truncated covariance converges in probability to

$$\frac{1}{TN}\sum_{s<t}\sum_{i=1}^{N}\hat{\sigma}_{u_{ti}^{A}\nu_{si}^{B}} \xrightarrow{p} \sigma_{\mu^{AB}}\overline{\sum_{j=1}^{N} r_j^A r_j^B} - \sigma_{\mu^{AB}}\overline{\sum_{j=1}^{N} r_j^A r_j^B \zeta},$$

since

$$\frac{1}{TN}\sum_{s<t}\sum_{i=1}^{N}\left[\sum_{j=1}^{N}\hat{r}_{tij}^{A}\hat{r}_{sij}^{B}\right] \xrightarrow{p} \overline{\sum_{j=1}^{N} r_j^A r_j^B} \quad \text{and}$$

$$\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}\hat{r}_{tij}^{A}\hat{r}_{sij}^{B}\hat{\zeta}_{ti}\right) \xrightarrow{p} \overline{\sum_{j=1}^{N} r_j^A r_j^B \zeta}.$$

We estimate $\frac{1}{TN}\sum_{s<t}\sum_{i=1}^{N}\hat{\sigma}_{u_{ti}^{A}\nu_{si}^{B}}$ by $\frac{\sum_{s<t}\sum_{i=1}^{N}(\hat{u}_{ti}^{A}\hat{\nu}_{si}^{B})^2}{TN}$ and solve the following equation for $\hat{\sigma}_{\mu^{AB}}$:

$$\frac{\sum_{s<t}\sum_{i=1}^{N}\hat{u}_{ti}^{A}\hat{\nu}_{si}^{B}}{TN} = \hat{\sigma}_{\mu^{AB}}\widehat{\overline{\sum_{j=1}^{N} r_j^A r_j^B}} - \hat{\sigma}_{\mu^{AB}}\widehat{\overline{\sum_{j=1}^{N} r_j^A r_j^B \zeta}}$$

Therefore, the estimate of the covariance of the individual-specific time-invariant component of the first and the second stage $\hat{\sigma}_{\mu^{AB}}$ is given by

$$\hat{\sigma}_{\mu^{AB}} = \frac{\sum_{s<t}\sum_{i=1}^{N}\hat{u}_{ti}^{A}\hat{\nu}_{si}^{B}}{TN}\left[\widehat{\overline{\sum_{j=1}^{N} r_j^A r_j^B}} - \widehat{\overline{\sum_{j=1}^{N} r_j^A r_j^B \zeta}}\right]^{-1}$$

Finally, to obtain $\hat{\sigma}_{\mu}^{2}$ we use (14), which gives us the true truncated covariance of residuals ν_{ti}^{B} and ν_{si}^{B}.

$$\sigma_{\nu_{ti}^{B}\nu_{si}^{B}} = \sigma_{\mu^B}^2\sum_{j=1}^{N} r_{tij}^{B} r_{sij}^{B} - \tau\psi_{ti}\frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma_{\xi^A}^2}}\frac{\sum_{j=1}^{N} r_{tij}^{A} r_{sij}^{B}}{\sqrt{\sum_{j=1}^{N}\left(r_{tij}^{A}\right)^2}}\zeta_{ti}$$

The sample average of the truncated covariance converges in probability to

$$\frac{1}{TN}\sum_{s<t}\sum_{i=1}^{N}\hat{\sigma}_{\nu_{ti}^{B}\nu_{si}^{B}} \xrightarrow{p} \sigma_{\mu^B}^2\overline{\sum_{j=1}^{N} r_j^B r_j^B} - \tau\frac{\sigma_{\mu^{AB}}}{\sqrt{\sigma_{\xi^A}^2}}\overline{\psi\frac{\sum_{j=1}^{N} r_j^A r_j^B}{\sqrt{\sum_{j=1}^{N}\left(r_j^A\right)^2}}\zeta},$$

since

$$\frac{1}{TN}\sum_{s<t}\sum_{i=1}^{N}\left[\sum_{j=1}^{N}\hat{r}_{tij}^{B}\hat{r}_{sij}^{B}\right] \xrightarrow{p} \overline{\sum_{j=1}^{N} r_j^B r_j^B} \quad \text{and}$$

$$\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(\hat{\psi}_{ti}\frac{\sum_{j=1}^{N}\hat{r}_{tij}^{A}\hat{r}_{sij}^{B}}{\sqrt{\sum_{j=1}^{N}\left(\hat{r}_{tij}^{A}\right)^2}}\hat{\zeta}_{ti}\right) \xrightarrow{p} \overline{\psi\frac{\sum_{j=1}^{N} r_j^A r_j^B}{\sqrt{\sum_{j=1}^{N}\left(r_j^A\right)^2}}\zeta}.$$

We estimate $\frac{1}{TN}\sum_{s<t}\sum_{i=1}^{N}\hat{\sigma}_{\nu_{ti}^{B}\nu_{si}^{B}}$ by $\frac{\sum_{s<t}\sum_{i=1}^{N}\hat{\nu}_{ti}^{B}\hat{\nu}_{si}^{B}}{TN}$ and solve the following equation for $\hat{\sigma}^2_{\mu^B}$:

$$\frac{\sum_{s<t}\sum_{i=1}^{N}\hat{\nu}_{ti}^{B}\hat{\nu}_{si}^{B}}{TN} = \hat{\sigma}^2_{\mu^B}\overline{\widehat{\sum_{j=1}^{N} r_j^B r_j^B}} - \hat{\tau}\frac{\hat{\sigma}_{\mu^{AB}}}{\sqrt{\hat{\sigma}^2_{\xi^A}}}\psi\overline{\widehat{\frac{\sum_{j=1}^{N} r_j^A r_j^B}{\sqrt{\sum_{j=1}^{N}\left(r_j^A\right)^2}}}}\zeta$$

Therefore, the estimate of the covariance of the individual-specific time-invariant component of the first and the second stage $\hat{\sigma}_{\mu^{AB}}$ is given by

$$\hat{\sigma}^2_{\mu^B} = \frac{1}{\overline{\widehat{\sum_{j=1}^{N} r_j^B r_j^B}}}\left[\frac{\sum_{s<t}\sum_{i=1}^{N}\hat{\nu}_{ti}^{B}\hat{\nu}_{si}^{B}}{TN} + \hat{\tau}\frac{\hat{\sigma}_{\mu^{AB}}}{\sqrt{\hat{\sigma}^2_{\xi^A}}}\psi\overline{\widehat{\frac{\sum_{j=1}^{N} r_j^A r_j^B}{\sqrt{\sum_{j=1}^{N}\left(r_j^A\right)^2}}}}\zeta\right]$$

3.4 Hessians of Second-Stage Parameters

The elements of the Hessian of q with respect to θ^B and θ^B are given by:

(recall: $\nu^B = y^B - \alpha y^A - x^B \beta^B - (I_T \otimes R_t^B)(\iota_t \otimes \bar{x}^B)\delta^B - \tau \Lambda^g$)

$$
\begin{aligned}
\frac{\partial^2 q}{\partial \alpha \partial \alpha} &= -2 \frac{\partial \nu^{B\prime}}{\partial \alpha} y^A = 2 y^{A\prime} y^A = 2Tr\left[y^A y^{A\prime}\right] \\
\frac{\partial^2 q}{\partial \alpha \partial \beta_r^B} &= -2 \frac{\partial \nu^{B\prime}}{\partial \beta_r^B} y^A = 2 x_r^{B\prime} y^A \\
\frac{\partial^2 q}{\partial \alpha \partial \delta_r^B} &= -2 \frac{\partial \nu^{B\prime}}{\partial \delta_r^B} y^A = 2\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}_r^B)\right]' y^A \\
\frac{\partial^2 q}{\partial \alpha \partial \tau} &= -2 \frac{\partial \nu^{B\prime}}{\partial \tau} y^A = 2 \Lambda^{g\prime} y^A \\
\frac{\partial^2 q}{\partial \alpha \partial \tau} &= -2 \frac{\partial \nu^{B\prime}}{\partial \rho^B} y^A = 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda^g}{\partial \rho^B}\right]' y^A
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 q}{\partial \beta_r^B \partial \alpha} &= -2 \frac{\partial \nu^{B\prime}}{\partial \alpha} x_r^B = 2 y^{A\prime} x_r^B \\
\frac{\partial^2 q}{\partial \beta_r^B \partial \beta_s^B} &= -2 \frac{\partial \nu^{B\prime}}{\partial \beta_s^B} x_r^B = 2 x_s^{B\prime} x_r^B \\
\frac{\partial^2 q}{\partial \beta_r^B \partial \delta_s^B} &= -2 \frac{\partial \nu^{B\prime}}{\partial \delta_s^B} x_r^B = 2\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}_s^B)\right]' x_r^B \\
\frac{\partial^2 q}{\partial \beta_r^B \partial \tau} &= -2 \frac{\partial \nu^{B\prime}}{\partial \tau} x_r^B = 2 \Lambda' x_r^B \quad \text{(Sample Selection)} \\
\frac{\partial^2 q}{\partial \beta_r^B \partial \tau} &= -2 \frac{\partial \nu^{B\prime}}{\partial \tau} x_r^B = 2 \Lambda^{g\prime} x_r^B \quad \text{(Treatment Selection)} \\
\frac{\partial^2 q}{\partial \beta_r^B \partial \rho^B} &= -2 \frac{\partial \nu^{B\prime}}{\partial \rho^B} x_r^B = 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right]' x_r^B \quad \text{(Sample Selection)} \\
\frac{\partial^2 q}{\partial \beta_r^B \partial \rho^B} &= -2 \frac{\partial \nu^{B\prime}}{\partial \rho^B} x_r^B = 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda^g}{\partial \rho^B}\right]' x_r^B \quad \text{(Treatment Selection)}
\end{aligned}
$$

$$\frac{\partial^2 q}{\partial \delta_r^B \partial \alpha} = -2\frac{\partial \nu^{B\prime}}{\partial \alpha}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) = 2y^{A\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)$$

$$\frac{\partial^2 q}{\partial \delta_r^B \partial \beta_s^B} = -2\frac{\partial \nu^{B\prime}}{\partial \beta_s^B}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) = 2x_s^{B\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)$$

$$\frac{\partial^2 q}{\partial \delta_r^B \partial \delta_s^B} = -2\frac{\partial \nu^{B\prime}}{\partial \delta_s^B}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) = 2\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}_s^B)\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)$$

$$\frac{\partial^2 q}{\partial \delta_r^B \partial \tau} = -2\frac{\partial \nu^{B\prime}}{\partial \tau}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) = 2\Lambda'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) \quad \text{(Sample Selection)}$$

$$\frac{\partial^2 q}{\partial \delta_r^B \partial \tau} = -2\frac{\partial \nu^{B\prime}}{\partial \tau}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) = 2\Lambda^{g\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) \quad \text{(Treatment Selection)}$$

$$\begin{aligned}\frac{\partial^2 q}{\partial \delta_r^B \partial \rho^B} &= -2\left\{\frac{\partial \nu^{B\prime}}{\partial \rho^B}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) + \nu^{B\prime}\frac{\partial\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\right]}{\partial \rho^B}\right\} \\ &= -2\left\{-\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) + \nu^{B\prime}(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\right\} \quad \text{(Sample Selection)}\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 q}{\partial \delta_r^B \partial \rho^B} &= -2\left\{\frac{\partial \nu^{B\prime}}{\partial \rho^B}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) + \nu^{B\prime}\frac{\partial\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\right]}{\partial \rho^B}\right\} \\ &= -2\left\{-\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) + \nu^{B\prime}(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\right\} \quad \text{(Treatment Selection)}\end{aligned}$$

Sample Selection

$$\frac{\partial^2 q}{\partial \tau \partial \alpha} = -2\frac{\partial \nu^{B\prime}}{\partial \alpha}\Lambda = 2y^{A\prime}\Lambda$$

$$\frac{\partial^2 q}{\partial \tau \partial \beta_r^B} = -2\frac{\partial \nu^{B\prime}}{\partial \beta_r^B}\Lambda = 2x_r^{B\prime}\Lambda$$

$$\frac{\partial^2 q}{\partial \tau \partial \delta_r^B} = -2\frac{\partial \nu^{B\prime}}{\partial \delta_r^B}\Lambda = 2\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}_r^B)\right]'\Lambda$$

$$\frac{\partial^2 q}{\partial \tau \partial \tau} = -2\frac{\partial \nu^{B\prime}}{\partial \tau}\Lambda = 2\Lambda'\Lambda$$

$$\frac{\partial^2 q}{\partial \tau \partial \rho^B} = -2\left\{\frac{\partial \nu^{B\prime}}{\partial \rho^B}\Lambda + \nu^{B\prime}\frac{\partial \Lambda}{\partial \rho^B}\right\} = -2\left\{-\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]'\Lambda + \nu^{B\prime}\frac{\partial \Lambda}{\partial \rho^B}\right\}$$

Treatment Selection

$$
\begin{aligned}
\frac{\partial^2 q}{\partial\tau\partial\alpha} &= -2\frac{\partial\nu^{B\prime}}{\partial\alpha}\Lambda^g = 2y^{A\prime}\Lambda^g \\
\frac{\partial^2 q}{\partial\tau\partial\beta_r^B} &= -2\frac{\partial\nu^{B\prime}}{\partial\beta_r^B}\Lambda^g = 2x_r^{B\prime}\Lambda^g \\
\frac{\partial^2 q}{\partial\tau\partial\delta_r^B} &= -2\frac{\partial\nu^{B\prime}}{\partial\delta_r^B}\Lambda^g = 2\left[(I_T\otimes R_t^B)(\iota_T\otimes\bar{x}_r^B)\right]'\Lambda^g \\
\frac{\partial^2 q}{\partial\tau\partial\tau} &= -2\frac{\partial\nu^{B\prime}}{\partial\tau}\Lambda^g = 2\Lambda^{g\prime}\Lambda^g \\
\frac{\partial^2 q}{\partial\tau\partial\rho^B} &= -2\left\{\frac{\partial\nu^{B\prime}}{\partial\rho^B}\Lambda^g + \nu^{B\prime}\frac{\partial\Lambda^g}{\partial\rho^B}\right\} = -2\left\{-\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]'\Lambda^g + \nu^{B\prime}\frac{\partial\Lambda^g}{\partial\rho^B}\right\}
\end{aligned}
$$

Sample Selection

$$
\begin{aligned}
\frac{\partial^2 q}{\partial\rho^B\partial\alpha} &= -2\frac{\nu^{B\prime}}{\partial\alpha}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] = 2y^{A\prime}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] \\
\frac{\partial^2 q}{\partial\rho^B\partial\beta_r^B} &= -2\frac{\nu^{B\prime}}{\partial\beta_r^B}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] = 2x_r^B\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] \\
\frac{\partial^2 q}{\partial\rho^B\partial\delta_r^B} &= -2\left\{\frac{\nu^{B\prime}}{\partial\delta_r^B}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] + \nu^{B\prime}\frac{\partial\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right]}{\partial\delta_r^B}\right\} \\
&= -2\left\{-(I_T\otimes R_t^B)(\iota_T\otimes\bar{x}^B)\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] + \nu^{B\prime}(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}_r^B)\right\} \\
\frac{\partial^2 q}{\partial\rho^B\partial\tau} &= -2\left\{\frac{\nu^{B\prime}}{\partial\tau}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] + \nu^{B\prime}\frac{\partial\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right]}{\partial\rho^B}\right\} \\
&= -2\left\{-\Lambda'\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] + \nu^{B\prime}\frac{\partial\Lambda}{\partial\rho^B}\right\}
\end{aligned}
$$

$$
\frac{\partial^2 q}{\partial\rho^B\partial\rho^B} = -2\left\{\frac{\partial\nu^{B\prime}}{\partial\rho^B}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] + \nu^{B\prime}\frac{\partial\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right]}{\partial\rho^B}\right\}
$$

We derive the partial derivative in the second summand inside the curly bracket first:

$$\frac{\partial\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right]}{\partial\rho^B} = \left\{(I_T \otimes \left[\frac{\partial R_t^B}{\partial\rho^B} W_t R_t^B + R_t^B W \frac{\partial R_t^B}{\partial\rho^B}\right])(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial^2\Lambda}{\partial\rho^B\rho^B}\right\}$$

Therefore:

$$\begin{aligned}
\frac{\partial^2 q}{\partial\rho^B\partial\rho^B} &= -2\{-\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right] \\
&+ \nu^{B\prime}\left\{(I_T \otimes \left[\frac{\partial R_t^B}{\partial\rho^B} W_t R_t^B + R_t^B W \frac{\partial R_t^B}{\partial\rho^B}\right])(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial^2\Lambda}{\partial\rho^B\rho^B}\right\}\}
\end{aligned}$$

Treatment Selection

$$\begin{aligned}
\frac{\partial^2 q}{\partial\rho^B\partial\alpha} &= 2y^{A\prime}\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right] \\
\frac{\partial^2 q}{\partial\rho^B\partial\beta_r^B} &= 2x_r^B\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right] \\
\frac{\partial^2 q}{\partial\rho^B\partial\delta_r^B} &= -2\left\{-(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right] + \nu^{B\prime}(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}_r^B)\right\} \\
\frac{\partial^2 q}{\partial\rho^B\partial\tau} &= -2\left\{-\Lambda^{g\prime}\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right] + \nu^{B\prime}\frac{\partial\Lambda^g}{\partial\rho^B}\right\} \\
\frac{\partial^2 q}{\partial\rho^B\partial\rho^B} &= -2\{-\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right] \\
&+ \nu^{B\prime}\left\{(I_T \otimes \left[\frac{\partial R_t^B}{\partial\rho^B} W_t R_t^B + R_t^B W \frac{\partial R_t^B}{\partial\rho^B}\right])(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial^2\Lambda^g}{\partial\rho^B\rho^B}\right\}\}
\end{aligned}$$

Finally, the expectation of these second-derivatives:

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial\alpha\partial\alpha}\right] &= E\left[2Tr\left[y^A y^{A\prime}\right]\right] = 2Tr\left[E\left[y^A y^{A\prime}\right]\right] \\
E\left[\frac{\partial^2 q}{\partial\alpha\partial\beta_r^B}\right] &= E\left[2x_r^{B\prime} y^A\right] = 2x_r^{B\prime} E\left[y^A\right] = 2x_r^{B\prime}\tilde{\Phi}(z) \\
E\left[\frac{\partial^2 q}{\partial\alpha\partial\delta_r^B}\right] &= 2\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}_r^B)\right]'\tilde{\Phi}(z) \\
E\left[\frac{\partial^2 q}{\partial\alpha\partial\tau}\right] &= 2\Lambda^{g\prime}\tilde{\Phi}(z) \\
E\left[\frac{\partial^2 q}{\partial\alpha\partial\tau}\right] &= 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]'\tilde{\Phi}(z)
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial\beta_r^B\partial\alpha}\right] &= 2\tilde{\Phi}(z)' x_r^B \\
E\left[\frac{\partial^2 q}{\partial\beta_r^B\partial\beta_s^B}\right] &= 2x_s^{B\prime} x_r^B \\
E\left[\frac{\partial^2 q}{\partial\beta_r^B\partial\delta_s^B}\right] &= 2\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}_s^B)\right]' x_r^B \\
E\left[\frac{\partial^2 q}{\partial\beta_r^B\partial\tau}\right] &= 2\Lambda' x_r^B \quad \text{(Sample Selection)} \\
E\left[\frac{\partial^2 q}{\partial\beta_r^B\partial\tau}\right] &= 2\Lambda^{g\prime} x_r^B \quad \text{(Treatment Selection)} \\
E\left[\frac{\partial^2 q}{\partial\beta_r^B\partial\rho^B}\right] &= 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda}{\partial\rho^B}\right]' x_r^B \quad \text{(Sample Selection)} \\
E\left[\frac{\partial^2 q}{\partial\beta_r^B\partial\rho^B}\right] &= 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]' x_r^B \quad \text{(Treatment Selection)}
\end{aligned}
$$

$$E\left[\frac{\partial^2 q}{\partial \delta_r^B \partial \alpha}\right] = 2\tilde{\Phi}(z)'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)$$

$$E\left[\frac{\partial^2 q}{\partial \delta_r^B \partial \beta_s^B}\right] = 2x_s^{B\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)$$

$$E\left[\frac{\partial^2 q}{\partial \delta_r^B \partial \delta_s^B}\right] = 2\left[(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}_s^B)\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)$$

$$E\left[\frac{\partial^2 q}{\partial \delta_r^B \partial \tau}\right] = 2\Lambda'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) \quad \text{(Sample Selection)}$$

$$E\left[\frac{\partial^2 q}{\partial \delta_r^B \partial \tau}\right] = 2\Lambda^{g\prime}(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) \quad \text{(Treatment Selection)}$$

$$\begin{aligned} E\left[\frac{\partial^2 q}{\partial \delta_r^B \partial \rho^B}\right] &= -2\left\{-\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) + E\left[\nu^{B\prime}\right](I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\right\} \\ &= 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda}{\partial \rho^B}\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) \quad \text{(Sample Selection)} \end{aligned}$$

$$\begin{aligned} E\left[\frac{\partial^2 q}{\partial \delta_r^B \partial \rho^B}\right] &= -2\left\{-\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) + E\left[\nu^{B\prime}\right](I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\right\} \\ &= 2\left[(I_T \otimes R_t^B W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau\frac{\partial \Lambda^g}{\partial \rho^B}\right]'(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B) \quad \text{(Treatment Selection)} \end{aligned}$$

Sample Selection

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial\tau\partial\alpha}\right] &= 2\tilde{\Phi}(z)'\Lambda \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\beta_r^B}\right] &= 2x_r^{B\prime}\Lambda \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\delta_r^B}\right] &= 2\left[(I_T\otimes R_t^B)(\iota_T\otimes\bar{x}_r^B)\right]'\Lambda \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\tau}\right] &= 2\Lambda'\Lambda \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\rho^B}\right] &= -2\left\{-\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'\Lambda+E\left[\nu^{B\prime}\right]\frac{\partial\Lambda}{\partial\rho^B}\right\} \\
&= 2\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda}{\partial\rho^B}\right]'\Lambda
\end{aligned}
$$

Treatment Selection

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial\tau\partial\alpha}\right] &= 2\tilde{\Phi}(z)'\Lambda^g \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\beta_r^B}\right] &= 2x_r^{B\prime}\Lambda^g \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\delta_r^B}\right] &= 2\left[(I_T\otimes R_t^B)(\iota_T\otimes\bar{x}_r^B)\right]'\Lambda^g \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\tau}\right] &= 2\Lambda^{g\prime}\Lambda^g \\
E\left[\frac{\partial^2 q}{\partial\tau\partial\rho^B}\right] &= -2\left\{-\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]'\Lambda^g+E\left[\nu^{B\prime}\right]\frac{\partial\Lambda^g}{\partial\rho^B}\right\} \\
&= 2\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]'\Lambda^g
\end{aligned}
$$

Sample Selection

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial \rho^B \partial \alpha}\right] &= 2\tilde{\Phi}(z)'\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right] \\
E\left[\frac{\partial^2 q}{\partial \rho^B \partial \beta_r^B}\right] &= 2x_r^B\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right] \\
E\left[\frac{\partial^2 q}{\partial \rho^B \partial \delta_r^B}\right] &= -2\left\{-(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right] + E\left[\nu^{B\prime}\right](I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}_r^B)\right\} \\
&= 2(I_T \otimes R_t^B)(\iota_T \otimes \bar{x}^B)\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right] \\
E\left[\frac{\partial^2 q}{\partial \rho^B \partial \tau}\right] &= -2\left\{-\Lambda'\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right] + E\left[\nu^{B\prime}\right]\frac{\partial \Lambda}{\partial \rho^B}\right\} \\
&= 2\Lambda'\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial \rho^B \partial \rho^B}\right] &= -2\{-\left[((I_T \otimes R_t^B) W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right] \\
&+ E\left[\nu^{B\prime}\right]\left\{(I_T \otimes \left[\frac{\partial R_t^B}{\partial \rho^B} W_t R_t^B + R_t^B W \frac{\partial R_t^B}{\partial \rho^B}\right])(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial^2 \Lambda}{\partial \rho^B \rho^B}\right\}\} \\
&= 2\left[((I_T \otimes R_t^B) W R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right]\left[(I_T \otimes R_t^B W_t R_t^B)(\iota_T \otimes \bar{x}^B)\delta^B + \tau \frac{\partial \Lambda}{\partial \rho^B}\right]
\end{aligned}
$$

Treatment Selection

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial\rho^B\partial\alpha}\right] &= 2\tilde{\Phi}(z)'\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\\
E\left[\frac{\partial^2 q}{\partial\rho^B\partial\beta_r^B}\right] &= 2x_r^B\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\\
E\left[\frac{\partial^2 q}{\partial\rho^B\partial\delta_r^B}\right] &= -2\left\{-(I_T\otimes R_t^B)(\iota_T\otimes\bar{x}^B)\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]+E\left[\nu^{B\prime}\right](I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}_r^B)\right\}\\
&= 2(I_T\otimes R_t^B)(\iota_T\otimes\bar{x}^B)\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\\
E\left[\frac{\partial^2 q}{\partial\rho^B\partial\tau}\right] &= -2\left\{-\Lambda^{g\prime}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]+E\left[\nu^{B\prime}\right]\frac{\partial\Lambda^g}{\partial\rho^B}\right\}\\
&= 2\Lambda^{g\prime}\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]
\end{aligned}
$$

$$
\begin{aligned}
E\left[\frac{\partial^2 q}{\partial\rho^B\partial\rho^B}\right] &= -2\{-\left[((I_T\otimes R_t^B)WR_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\\
&+ E\left[\nu^{B\prime}\right]\left\{(I_T\otimes\left[\frac{\partial R_t^B}{\partial\rho^B}W_t R_t^B+R_t^B W\frac{\partial R_t^B}{\partial\rho^B}\right])(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial^2\Lambda^g}{\partial\rho^B\rho^B}\right\}\}\\
&= 2\left[((I_T\otimes R_t^B)WR_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]\left[(I_T\otimes R_t^B W_t R_t^B)(\iota_T\otimes\bar{x}^B)\delta^B+\tau\frac{\partial\Lambda^g}{\partial\rho^B}\right]
\end{aligned}
$$