EIEF WORKING PAPER SERIES

EINAUDI INSTITUTE FOR ECONOMICS AND FINANCE

EIEF Working Paper 15/09

October 2015

Parametric and Semiparametric IV Estimation of Network Models with Selectivity

by

Tiziano Arduini

(Sapienza University of Rome)

Eleonora Patacchini

(Cornell University, Sapienza University of Rome, EIEF, IZA and CEPR)

Edoardo Rainone

(Bank of Italy)

Parametric and Semiparametric IV Estimation of Network Models with Selectivity

Tiziano Arduini * Eleonora Patacchini † Edoardo Rainone ‡

Abstract

We propose parametric and semiparametric IV estimators for spatial autoregressive models with network data where the network structure is endogenous. We embed a dyadic network formation process in the control function approach as in Heckman and Robb (1985). In the semiparametric case, we use power series to approximate the correction terms. We establish the consistency and asymptotic normality for both parametric and semiparametric cases. We also investigate their finite sample properties via Monte Carlo simulation.

Key Words: Spatial Autoregressive Model, Endogenous Network Formation, Semiparametric estimation, Sample selection models, Two-step estimation, Series estimation.

JEL Codes: C14, C31, C51

*Sapienza University of Rome. E-mail: tiziano.arduini@uniroma1.it.

†Cornell University, Sapienza University of Rome, EIEF, IZA and CEPR. E-mail: ep454@cornell.edu.

‡Bank of Italy, DG for Markets and Payment Systems. E-mail: edoardo.rainone@bancaditalia.it. The views expressed here do not necessarily reflect those of the Bank of Italy.

1 Introduction

The estimation procedures for the spatial autoregressive (SAR) models with network data and an exogenous network structure are now well established (see Lee, 2007; Lee et al., 2010; Liu, 2013b; Liu and Lee, 2010). In almost all social and economic contexts, however, network is chosen exogenously and behaviors were derived given this network structure. Network formation and behavior over networks can happen simultaneously. Fail to account individual characteristics driving both can result serious endogeneity issues.

Despite the problem of endogeneity, estimation methods for network models with selectivity are not widely available. Hsieh and Lee (2014) and Goldsmith-Pinkham and Imbens (2013) propose a simultaneous estimation of dyadic network formation and outcome choice using a Bayesian modeling approach. They use MCMC methods to draw from the posterior distribution due to the complexity of their likelihood function. However, a behavioral foundation under such a framework is challenging. As networks can form under various rules, one needs to find conditions that match the properties of equilibrium outcomes for many different games. Bayesian methods are also unfeasible when the sample size is large. Blume et al. (2014, forthcoming) suggests using control function approach to address network formation.[1] They consider the development of the control function approach to address network formation as an important direction of future research.[2]

This paper develops this approach. An independent study by Qu and Lee (2015) proposes a control function method to estimate a SAR model with an endogenous spatial weight matrix. They construct the spatial weights using distances between units in terms of economic variables (such as GDP or trade volumes). Endogeneity arises when these weights are correlated with the outcome. They thus employ an OLS estimator for the selection equation (first stage) and correct the SAR outcome model with the estimated residuals (second stage). We consider network data, such that the entries of the spatial weights matrix are binary. We modeled a dyadic network formation mechanism for the selection equation as in Graham (2014).[3] Graham (2014) presents a comprehensive tractation of dyadic empirical models of network formation. In particular, he parametrizes link formation in a logit form and derives the statistical properties of a maximum likelihood estimator.[4] Due to the technical complication in estimating a selection correction term at the individual level (i.e. with a dyadic network formation, as in Graham, 2014), no estimation method has been proposed for this case so far.[5]

We propose two estimation methods. The first method is a simple two-stage instrumental variable estimator with a parametric selection procedure (2SP_IV).[6] With explicit distributional assumptions on the disturbances, we show that the selectivity bias is a multivariate inverse Mills ratio. The asymptotic theory is derived using asymptotic inference under near-epoch dependence (NED) from Jenish and Prucha (2012). The results are similar to the ones in Qu and Lee (2015). The second estimator we proposed uses a power series to approximate selectivity bias term, in the spirit of Newey (2009). This is a two-stage semiparametric instrumental variable estimator (2SSP_IV). We also derive statistical properties for our estimator that is very easy to implement in applied work.

The main aim of this paper is to show consistency and asymptotic normality of the 2SSP_IV and the 2SP_IV. Lee (1982) proposes two-stage instrumental variable estimators for selection models with flexible correction

[1] Blume et al. (2014, forthcoming); Brock and Durlauf (2001a,b) discuss how selection correction mechanisms can be used to overcome the reflection problem.

[2] The endogeneity of spatial weights has also been pointed out as an important future direction of spatial econometrics by Pinkse and Slade (2010). They argue that many of the issues arising from this problem are still waiting for good solutions and that the endogeneity problem "can admittedly be challenging."

[3] This framework can be motivated from the behavior of utility maximization (McFadden et al., 1973).

[4] A simplified version of the model has been used by Fafchamps and Gubert (2007), Apicella et al. (2012), Attanasio et al. (2012).

[5] Horrace et al. (2014, forthcoming) deal with the estimation of network production models where the working team (network) is chosen by a manager. In their framework, teams are not overlapping and they can employ a polychotomous Heckman-type selection correction.

[6] See Heckman and Robb (1985) and Vella (1998) for an overview of models and methods to address selection problems.

terms. The theory presented in this paper allows for the functional form of the selectivity bias term to be entirely unknown, with the number of approximating functions growing with the sample size to achieve $\sqrt{n}-$consistency and asymptotic normality. Our main results show the NED properties of the random variables and functions involved in the 2SSP_IV estimator.

The rest of the paper is organized as follows. The next section introduces the econometric specification of a network model with selectivity. We discuss different ways to correct for the selectivity bias. Section 3 proposes parametric and semiparametric estimation methods of our model. Consistency and asymptotic normality of estimators from these methods are derived in Section 4. We evaluate finite sample performance of our estimators in Section 5 using Monte Carlo simulations. Section 6 is conclusion. Appendix A contains some basic properties of NED of random fields and power series. Appendix B includes the proofs of NED properties of the key statistics of the estimators. Appendix C gives the proof of our main results.

2 A network model with selectivity

We consider n agents indexed by $i = 1, \dots, n$. Each agent is endowed with a predetermined location $l(i)$. Following Jenish and Prucha (2012) and Qu and Lee (2015), we consider spatial processes located on a (possibly) unevenly spaced lattice $D \subset R^{d_0}$, $d_0 \geq 1$. From now on, we maintain the following assumption concerning D.

Assumption 1. *The Lattice $D \subset R^{d_0}$, $d_0 \geq 1$, is infinitely countable. The location $l : \{1, \dots, n\} \to D_n \subset D$ is a mapping of agent i to its location $l(i) \in D_n$. All elements in D are located at distances of at least $\rho_0 > 0$ from each other, i.e., for all $l(i), l(j) \in D : \rho_{ij} \geq \rho_0$; w.l.o.g. we assume $\rho_0 = 1$.*

Physical distance plays a crucial role in our asymptotic methods. It ensures the growth of the sample size as the sample regions D_n expand, see e.g. Conley (1999) and Jenish and Prucha (2009, 2012). This means the asymptotic methods we employed are increasing domain asymptotics.

Agents in the sample regions may be connected. Let $g_{ij,n} = 1$ if agents i and j are connected and zero otherwise. Connections may be equivalently referred to as a links, friendships, edges or arcs depending on the context. Self-ties are rulled out so that $g_{ii,n} = 0$ for all i. The (i, j) pair is called a dyad. There are $N_d = n(n-1)$ dyads in the sample. For a sample size of n agents, we observe the vector of outcomes $Y_n = (y_{1,n}, \dots, y_{n,n})'$, the covariate matrix X_n with its element $\{x_{i,n}; l(i) \in D_n, n \in N\}$ being bounded in absolute value for all i and n. $G_n = \{g_{ij,n}\}$ is a binary matrix with zero diagonal.

Let $\{(v_{l(i),n}, u_{l(i),n}); l(i) \in D_n, n \in N\}$ be a triangular double array of real random fields defined on a probability space $(\Omega; F; P)$, where the index set $D_n \subset D$ is a finite set and D satisfies Assumption 1. For simplicity, we use the subscript i to indicate $l(i)$, i.e. $v_{l(i),n}$ becomes $v_{i,n}$.

Following Graham (2014), agents i and j form a link if the total surplus from doing so is positive

$$g_{ij,n} = I(C_{ij,n}\alpha + \beta_d \rho_{ij} + a_{i,n}^{in} + a_{j,n}^{out} - v_{ij,n} \geq 0), \tag{1}$$

where $I(\cdot)$ denotes the indicator function, $\{a_{i,n}^{dir}; l(i) \in D_n, n \in N\}$, with $dir \in \{in, out\}$ ("in" stands for indegree, "out" for outdegree), are two agent-specific unobserved characteristics that capture the propensity of having a indegree and a outdegree.[7] The term inside the indicator function corresponds to the net link surplus. Similar to Leung (2014), link surplus varies with observed dyad attributes $\{C_{ij,n}; l(i) \in D_n, l(j) \in D_n, n \in N\}$, distances ρ_{ij}, unobserved agent attributes $(a_{i,n}^{in}, a_{j,n}^{out})$, and an idiosyncratic component $(v_{ij,n})$. Model (1) satisfies a no externalities condition. The net surplus associated with a link does not vary with the presence or

[7]Observe that model (1) is the Graham (2014) network formation model augmented for indegree and outdegree agent level unobserved heterogeneity. It basically extend the model for a directed graph representation.

absence of other links in the network. For example, the vector $(C_{ij,n})$ does not include the number of friends i and j have in common, $\sum_{k=1}^{n} g_{ik,n} g_{jk,n}$.[8]

We assume that the outcome equation is a SAR model specified as

$$Y_n = \phi G_n Y_n + X_n \beta + U_n, \tag{2}$$

where $U_n = (u_{1,n}, \ldots, u_{n,n})'$ is a vector of disturbances.[9] In model (2), ϕ represents the level of *the endogenous effect*, where an agent's choice/outcome may depend on those of her peers on the same activity.[10] [11] β is a $k \times 1$ vector of coefficients.

2.1 The parametric approach

Let $C_n = (C_{12}, \ldots, C_{n,n-1})'$ be the $N_d \times k$ matrix of dyad characteristics and $A_n = (a_1, \ldots, a_n)'$ the $n \times 2$ vector of unobserved agent characteristics, where $a_i = [a_i^{in}; a_i^{out}]$. Following Graham (2014), we focus on the case where $v_{ij,n}$ is independently logistically distributed across dyads. Thus, the conditional probability of forming a link (i, j) equals

$$Pr(g_{ij,n} = 1 | C_n, A_n) = \frac{exp(\alpha C_{ij,n} + \beta_d \rho_{ij} + a_{i,n}^{in} + a_{j,n}^{out})}{1 + exp(\alpha C_{ij,n} + \beta_d \rho_{ij} + a_{i,n}^{in} + a_{j,n}^{out})}. \tag{3}$$

Graham (2014) assumed that links form independently conditional on C_n and A_n. This assumption is appropriate in settings where no strategic decisions are involved. This is true in some types of friendship networks, See Graham (2014) for further details.

2.1.1 The selectivity bias

In model (2) endogeneity of G_n may arise from the correlation between $u_{i,n}$ and $v_{ij,n}$, controlling for $a_{i,n}$. This means the shock on meeting opportunities $v_{ij,n}$, can be correlated with agents' unobservables in the outcome equation. The unobserved characteristic $a_{i,n}$ captures unobserved attributes of agent i that make her a good/bad link partner. In a risk-sharing network, for example, trustworthy agents may have high values for $a_{i,n}$ (Fafchamps and Gubert, 2007). Even we control the level of trustworthy in the village, meeting opportunities $(v_{ij,n})$ could still correlate with the agent's unobserved ability $(u_{i,n})$ which affect the probability of finding josb $(y_{i,n})$. Formally, we make the following assumption.

Assumption 2. *The error terms u_i and $v'_{i,n} = \{v_{lj}\}_{l=i, l\neq j}$ have a joint distribution $(u_{i,n}, v'_{i,n}) \sim i.i.d.(0, \Sigma_{uv})$, where $\Sigma_{uv} = \begin{pmatrix} \sigma_u^2 & \sigma'_{uv} \\ \sigma_{uv} & \Sigma_v \end{pmatrix}$ is positive definite, σ_u^2 is a scalar variance, σ_{uv} is a $n-1$ vector of covariances with constant elements across i and j, and $\Sigma_v = \sigma_v I_{n-1}$ is a $n-1 \times n-1$ matrix. The $\sup_{i,n} ||v'_{i,n}||^{4+\delta_\epsilon}$ and $\sup_{i,n} |u'_{i,n}|^{4+\delta_\epsilon}$ exist for some $\delta_\epsilon > 0$. Furthermore, $E(u_{i,n}|v'_{i,n}) = v'_{i,n}\gamma$ and $var(u_{i,n}|v_{i,n}) = \sigma_\epsilon^2$.*

[8]See Jackson and Wolinsky (1996), Jackson and Watts (2002) and Bala and Goyal (2000) for studies of network formation in the presence of interdependent preferences. See Christakis et al. (2010), Mele (2010), Goldsmith-Pinkham and Imbens (2013), Sheng (2014) and De Paula et al. (2014) for some recent attempts to study econometric models of network formation with interdependent preferences.

[9]We consider here an aggregate model specification (i.e. G which multiplies y in model (2) is not row-normalized. Our estimation method applies also to an average model (i.e. when G which multiplies y in model (2) is row-normalized) or to other functions of the outcome of connected agents (like max, min, variance, etc.). Without loss of generality, *contextual effects* are not included in equation (2). Contextual effects refer to characteristics of individual i's peers, i.e the average level of peers' exogenous covariates $(\sum g_{ij}/g_{i\cdot}x_j)$. Their inclusion would only make the notation heavier.

[10] When the G_n matrix is predetermined the endogeneous effects is identified under conditions provided by Bramoullé et al. (2009). See also Calvó-Armengol et al. (2009) and Lee (2007).

[11]See Ballester et al. (2006) and Calvó-Armengol et al. (2009) for a motivation of the use of the SAR model to study peer effects.

From the two conditional moments in Assumption 2 we have a $n-1$ column vector $\gamma = (\Sigma_v^{-1}\sigma_{uv})'$ and the scalar $\sigma_\epsilon^2 = \sigma_u^2 - \sigma_{uv}'\Sigma_v^{-1}\sigma_{uv}$. If σ_{uv} is zero, the adjacency matrix G_n might be treated as exogenously given and we can apply conventional methodology for the estimation of SAR models. However, stochastic dependence between $u_{i,n}$ and $v_{ij,n}$ (selection on unobservables) and stochastic dependence between $u_{i,n}$ and $c_{ij,n}$ (selection on observables) may generate a nonzero σ_{uv}, thus G_n becomes an endogenous adjacency matrix. Let $V_n = (v_{1,n}', \cdots v_{n,n}')$, given Assumption 2 we can decompose the error term in the outcome equation (2) in the following way

$$Y_n = \phi G_n Y_n + \beta X_n + \underbrace{\gamma V_n + \varepsilon_n}_{U_n}, \tag{4}$$

where $\varepsilon_n = (\varepsilon_{1,n}, \ldots, \varepsilon_{n,n})$, with $\varepsilon_{i,n}$ independent to $v_{i,n}$. Formally, the selectivity bias is equal to γV_n.

Let us define $G_{i,n}$ as the i^{th} row of the matrix G_n. Under Assumption 2 we have

$$\begin{aligned} \psi_{1,n}' = E(v_{i,n}|A_{i,n}, C_{i,n}, G_{i,n}) &= \quad (5)\\ E(v_{i,n}'|I(v_{i2,n} \geq -\alpha C_{i2,n} - a_{i,n} - a_{2,n}), \ldots, I(v_{in,n} \geq -\alpha C_{in,n} - a_{i,n} - a_{n,n}), A_n, C_n) &= \\ (E(v_{i1,n}|I(v_{i2,n} \geq -\alpha C_{i1,n} - a_{i,n} - a_{1,n})), \ldots, E(v_{in,n}|I(v_{in,n} \geq -\alpha C_{in,n} - a_{i,n} - a_{n,n}))' &= \quad (6) \end{aligned}$$

where in equation (6) we use the fact that v_{ij} are considered i.i.d. in the network formation model after controlling for node-specific heterogeneity. Denote $\Psi_n = (\psi_{1,n}', \ldots, \psi_{n,n}')'$ and $\xi_{i,n} = \gamma(v_{i,n} - \psi_{i,n}) + \varepsilon_{i,n}$. The selectivity corrected model (4) becomes

$$Y_n = \phi G_n Y_n + \beta X_n + \gamma \Psi_n + \xi_n, \tag{7}$$

where $E(\xi_{i,n}|G_{i,n}, A_n, C_n) = 0$ and $E(\xi_{i,n}^2|G_{i,n}, A_n, C_n) = \sigma_\epsilon^2$; and $\xi_{i,n}$'s are i.i.d. across i. Our asymptotic analysis will mainly rely on equation (7), where $\psi_{i,n}'$ are functions to control for the endogeneity of G_n.

In the parametric case, in order to have a closed form expression for $\psi_{i,n}'$, we need to characterize the joint distribution of $(u_{i,n}, v_{i,n}')$. So far we have assumed that the marginal distribution of $v_{ij,n}$ is logistic. Let us assume that the marginal distribution of u_i is normal. Lee (1982) suggests to transform the error $v_{ij,n}$ using a strictly increasing transformation, J. The conditional moments in Assumption 2 become $E(u_{i,n}|v_{i,n}^*) = v_{i,n}^{*'}\gamma$, where $v_{ij,n}^* = J(v_{ij,n})$. The Lee's approach provides a way to generate a large class of models with selectivity. By specifying different transformations, we allow different distributions for u and thus any specific probability choice model falls into our method of correcting the selectivity bias term. This approach is particularly useful when the marginal distribution of the outcome error $u_{i,n}$ is assumed to be normal and the marginal of $v_{ij,n}$ is known but not normal. Thus, the joint distribution of $u_{i,n}$ and the transformed error $v_{i,n}^*$ can be set as a bivariate normal and their dependence is captured by the correlation coefficient. The term in (5) becomes

$$\begin{aligned} \psi_{i,n}^* = E(v_{i,n}^{*'}|G_{i,n}, A_n, C_n) = \quad (8)\\ (g_{i1,n}E(v_{i1,n}^*|I(v_{i1,n}^* \geq -J(\alpha C_{i1,n} + a_{i,n} + a_{1,n}))) + (1 - g_{i1,n})E(v_{i1,n}^*|I(v*_{i1,n} < -J(\alpha C_{i1,n} + a_{i,n} + a_{1,n}))), \ldots, \\ (g_{in,n}E(v_{in,n}^*|I(v_{in,n}^* \geq -J(\alpha C_{in,n} + a_{i,n} + a_{n,n})))' + (1 - g_{in,n})E(v_{in,n}^*|I(v_{in,n}^* < -J(\alpha C_{in,n} + a_{i,n} + a_{n,n})))'. \end{aligned}$$

Let $f_J(\cdot)$ be the implied density function of $v_{ij,n}^*$ which is assumed to exist under the transformation J. Let us denote the incomplete first moment of the r.v. $v_{ij,n}^*$ evaluated at $J(\alpha C_{ij,n} + a_{i,n} + a_{j,n})$ as

$$\mu^{+}_{ij} = \mu^{+}(J(\alpha C_{ij,n} + a_{i,n} + a_{j,n})) = \int_{J(-\infty)}^{J(\alpha C_{ij,n}+a_{i,n}+a_{j,n})} v^{*}_{ij,n} f_J(v^{*}_{ij,n})\, dv^{*}_{ij,n}, \tag{9}$$

$$\mu^{-}_{ij} = \mu^{-}(J(\alpha C_{ij,n} + a_{i,n} + a_{j,n})) = \int_{J(\alpha C_{ij,n}+a_{i,n}+a_{j,n})}^{J(\infty)} v^{*}_{ij,n} f_J(v^{*}_{ij,n})\, dv^{*}_{ij,n}.$$

Let $F_{ij} = F(\alpha C_{ij,n} + a_{i,n} + a_{j,n}) = Pr(v_{ij,n} < \alpha C_{ij,n} + a_{i,n} + a_{j,n})$ be the probability that the event $g_{ij} = 0$ occurs. The selectivity bias term $\psi^{*}_{i,n}$ equals to $(g_{i1,n}\mu^{-}_{i1}/(1-F_{i1}) + (1-g_{i1,n})\mu^{+}_{i1}/F_{i1} \,\ldots,\, g_{in,n}\mu^{-}_{in}/(1-F_{in}) + (1-g_{in,n})\mu^{+}_{in}/F_{in})$. Conditional on the endogenous adjacency matrix, equation (7) becomes

$$Y_n = \phi G_n Y_n + \beta X_n + \gamma \Psi^{*}_n + \xi_n, \tag{10}$$

where $\gamma = [(1/\sigma_{v*})^2 I_{n-1}\sigma_{uv*}]'$ and $\xi_{i,n} = \gamma(v^{*}_{i,n} - \psi^{*}_{i,n}) + \varepsilon_{i,n}$.

2.2 The semiparametric approach

Model (2)-(1) can be reformulated using a multiple-index/partially-linear model specification. Let $L_{i,n} = (C_{i,n}, \rho_i)$, with $C_{i,n} = (c_{i1,n}, \ldots, c_{in,n})$ and $\rho_i = (\rho_{i2}, \ldots, \rho_{in})$, and $\lambda = (\alpha, \beta_d, A_n)$. Let us define $m_i = m(\lambda, L_{i,n})$ as a known function that determines the selection probability. Following Newey (2009), we assume a standard index sufficiency conditions (see, e.g. Powell, 1994)

Assumption 3. *Let*

$$\psi(m_i) = E(u_{i,n}|G_{i,n}, X_n) = E(u_{i,n}|m(\lambda, L_{i,n}), G_{i,n}), \tag{11}$$

$$P(g_{ij,n} = 1|C_{i,n}) = \pi(m(\lambda, L_{i,n})), \tag{12}$$

$$\psi(m_i) = w(\psi(m_{i1,n}), \ldots, \psi(m_{in,n})), \tag{13}$$

where $\psi(\cdot)$ and $\pi(\cdot)$ are unknown functions and $w(\cdot)$ is known.

Thus, equation (11) means that conditional on selection the mean of the outcome disturbances depend only on m_i.[12] Using this assumption, we can rewrite equation (2) as

$$y_{i,n} = \phi G_{i,n} Y_n + \beta x_{i,n} + \psi(m_i) + \varepsilon_{i,n}. \tag{14}$$

Under Assumption 2, if we characterize the distribution of the errors $(u_{i,n}, v'_{i,n})$, the function $\psi(m_i)$ becomes the multivariate inverse Mills ratio in *step 2* of Section 3.1. The term is a generalization of the correction term considered by Heckman and Robb (1985). In this paper, we allow $\psi(m_i)$ to have an unknown functional form. Equation (14) is an additive semiparametric regression like that considered by Jenish (2013) and Su (2012), except that the variable $m = m(\lambda, L_n)$ depends on unknown parameters.

3 Estimation methods

3.1 The two-stage parametric IV estimation

In practice, under Assumption 2 the parametric approach can be implemented as follows.

[12]This restriction is implied by the assumption of independence between disturbances and regressors.

Step 1: Estimate the dyadic network formation model (1) using Graham (2014) joint maximum likelihood estimator, let $\hat{\alpha}$, $\hat{\beta}_d$ and $\hat{a}^{in}_{i,n}, \hat{a}^{out}_{i,n}$ be the estimated parameters.

Step 2: Using the estimated parameters from *Step 1* to compute

$$\hat{\psi}_{i,n} = ((g_{i1}\hat{\mu}^{-}_{i1}/(1-\hat{F}_{i1}) + (1-g_{i1})\hat{\mu}^{+}_{i1}/\hat{F}_{i1}, \ldots, g_{in}\hat{\mu}^{-}_{in}/(1-\hat{F}_{in}) + (1-g_{in})\hat{\mu}^{+}_{in}/\hat{F}_{in})),$$

by assuming $J = \Phi^{-1}(\cdot)$, where Φ is the normal CDF and $f_j(\cdot)$ is a normal density function. We also assume that $(u_i, v_i'*)$ follows a joint normal distribution.

Step 3: With the estimated $\hat{\Psi}_n$, we consider the feasible counterpart of equation (10)

$$Y_n = \phi G_n Y_n + \beta X_n + \gamma \hat{\Psi}_n + \hat{\xi}_n, \tag{15}$$

where $\hat{\xi}_i$ is a zero mean normally distributed residual. Estimation of γ in equation (15) is made feasible assuming a constant covariance between $u_{i,n}$ and $v_{i,n}$ across i and j.[13] Thus, we can sum up all the selectivity correction terms. A weaker version of Assumption 2 is often used in the literature concerned with the estimation of polychotomous sample selection models (see Dahl, 2002; Schmertmann, 1994). We can relax this assumption, by allowing a less restrictive covariance structure as long as the number of parameters in γ is small and do not grow with n. We rewrite model (15) in a more compact form

$$Y_n = (G_n Y_n, X_n, \hat{\Psi}_n)\kappa + \hat{\xi}_n, \tag{16}$$

where $\hat{\xi}_n = \xi_n + \gamma(\Psi_n - \hat{\Psi}_n)$ and $\kappa = (\phi, \beta', \gamma')'$. For the estimation of equation (16), with the inclusion of the control functions in Ψ_n, G_n can be treated as exogenous. However, $G_n Y_n$ remains endogenous and we can use an IV estimation method in the spirit of Qu and Lee (2015). Let Q_n be an $n \times n$ matrix of IVs, then a 2SIV estimator of κ (2SP_IV) is

$$\begin{aligned} \hat{\kappa} &= [(G_n Y_n, X_n, \hat{\Psi}_n)' Q_n (Q_n' Q_n)^{-1} \\ &\times Q_n' (G_n Y_n, X_n, \hat{\Psi}_n)]^{-1} (G_n Y_n, X_n, \hat{\Psi}_n)' \\ &\times Q_n (Q_n' Q_n)^{-1} Q_n' Y_n. \end{aligned} \tag{17}$$

As the composite error $\xi_{in} = \gamma(v^*_{i,n} - \psi_{in}) + \varepsilon$ is heterosckedastic as its variance matrix is

$$\begin{aligned} \sigma^2_{\xi} = var(\xi_{in}|G_{i,n}, A_n, C_n) &= \gamma[E(v^{*\prime 2}_{i,n}|I(v_{i1,n} \geq -\alpha C_{i1,n} - a_{i,n} - a_{1,n}), \ldots, I(v_{in,n} \geq -\alpha C_{in,n} - a_{i,n} - a_{n,n}), A_n, C_n) \\ &- E(v^{*\prime}_{i,n}|I(v_{i1,n} \geq -\alpha C_{i1,n} - a_{i,n} - a_{1,n}), \ldots, I(v_{in,n} \geq -\alpha C_{in,n} - a_{i,n} - a_{n,n}), A_n, C_n)^2]\gamma' \\ &+ \sigma^2_u + \sigma'_{uv}\Sigma_v^{-1}\sigma_{uv}, \end{aligned}$$

we may also consider a generalized 2SIV (G2SIV). In a SAR model with endogenous weights matrix, Qu and Lee (2015) propose a G2SIV estimator. In our case its properties are the same as in Qu and Lee (2015).

[13] In practice, we also assumed that the covariance is constant between $u_{i,n}$ and $v_{ij,n}$, no matter whether the link is realized or not (when $g_{ij} = 1$ or $g_{ij} = 0$).

3.2 The two-stage semiparametric IV estimation

The asymptotic properties of the two-step estimator for semiparametric sample selection models have been derived by Newey (2009). This estimator works as the theoretical basis for ours. A similar kind of models has been studied also by Su (2012) and Jenish (2013), but in econometric frameworks without selectivity.[14] Following Newey (2009), we can estimate this model in the following two-step procedure.

Step 1: Let $(\hat{\alpha}, \hat{\beta}_d, \hat{A})$ and $\hat{m}_i$ be the semiparametric estimates from the link formation model (1).[15]

Step 2: The second step consists of a linear regression of Y_n on G_nY_n, X_n and functions of $\hat{m}$ that can approximate $\psi(m)$. Let us define $\tau(m, \eta)$ as a strictly monotonic transformation of each entry of the index m, depending on parameter η. Let $P^K(\tau) = (P_{1K}(\tau), \ldots, P_{KK}(\tau))'$ be a vector of functions with the properties that for large value of K a linear combination of $P^K(\tau)$ can approximate an unknown function of $\tau(\cdot)$. Let $\hat{\tau}_i = \tau(\hat{m}_i, \hat{\eta})$ and $\hat{p}_i = P^K(\hat{\tau}_i)$. Let us assume furthermore, that the approximating functions are power series given by[16]

$$P_{kK}(\tau) = \tau^{k-1}.$$

This power series can lead to several different types of selection correction, depending on the transformation τ one implements. As in Newey (2009), we give three examples of monotonic transformation. A power series approximates the index $\hat{m}$ (linear), the inverse Mills ratio $\frac{\phi(\cdot)}{\Phi(\cdot)}$(as in the parametric case), or the normal CDF $\Phi(\cdot)$. In the last two examples, it may be appropriate to undo a location and scale transformation imposed in most semiparametric estimators of $m(L_i, \lambda)$. Consider the third example of the normal CDF. To this end, let $\hat{\eta}$ be the coefficient from a probit estimation of G_i on $\hat{m}_i$. Then, the transformed observation for the third example is $\tau_{ij} = \Phi(\hat{\eta}\, \hat{m}_{ij})$.[17]

Thus, we can write model (14) as

$$y_{i,n} = \phi G_i Y_n + \beta x_{i,n} + \sum_{k=1}^{q} \gamma_k \tau_i^{k-1} + \varepsilon_{i,n}. \tag{18}$$

As in the parametric case, however, G_nY_n remains endogenous. Let us define $Z_n = [G_nY_n, X_n, \Psi(m)]$ is a $n \times (q+1+k)$ matrix. Let H_K be an $n \times n$ matrix of IVs. The feasible 2SIV estimator of $\mu = (\phi, \beta)'$ (2SSP_IV) for model (14) is

$$\hat{\mu} = (Z_1'\hat{R}_K(I - \hat{Q}_K)\hat{R}_K Z_1)^{-1} Z_1'\hat{R}_K'(I - \hat{Q}_K)\hat{R}_K Y_n, \tag{19}$$

where $\hat{R}_K = \hat{H}_n(\hat{H}_n'\hat{H}_n)^{-1}\hat{H}_n'$ and $Z_1 = [G_nY_n, X_n]$.

[14] An additional difference with their models consists in the approximating function, they use kernel-based method. As highlighted before, we suggest to use series estimator given their virtue of being easy to implement in our context.

[15] There are many distribution-free estimators that are available for the selection equation, including those of Manski (1975), Cosslett (1983), Powell et al. (1989), Ichimura (1993), Klein and Spady (1993), Khan (2013) and Lei (2013) in a SAR model.

[16] Other approximating functions can be used. The asymptotic properties of 2SSP_IV also hold with regression spline approximation by slightly changing the growth rate of the number of approximating functions. See, e.g. Newey (2009).

[17] Observe that we do not impose normality.

4 Asymptotic Properties

4.1 Key Statistics

The 2SP_IV

For the 2SP_IV estimator $\hat{\kappa}$

$$\begin{aligned} \hat{\kappa} - \kappa_0 &= [(G_n Y_n, X_n, \hat{\Psi}_n)' Q_n (Q_n' Q_n)^{-1} \\ &\times Q_n'(G_n Y_n, X_n, \hat{\Psi}_n)]^{-1} (G_n Y_n, X_n, \hat{\Psi}_n)' \\ &\times Q_n (Q_n' Q_n)^{-1} Q_n' \hat{\xi}_n. \end{aligned} \tag{20}$$

where the subscript 0 on parameters denotes their true values. From the reduced form of the model (15) we have

$$GnY_n = G_n(I - \phi_0 G_n)^{-1}(\beta_0 X_n + \gamma_0 \Psi_n + \xi_n) = S_n(\beta_0 X_n + \gamma_0 \Psi_n + \xi_n),$$

where $S_n = G_n(I - \phi_0 G_n)^{-1}$ and $S_n(\phi) = G_n(I - \phi G_n)^{-1}$. For the consistency and asymptotic distribution of $\hat{\kappa}$ terms we need to analyze are $Q_n' Q_n$, $Q_n'(G_n Y_n, X_n, \Psi_n)$, $Q_n'(\xi_n + \gamma(\Psi_n - \hat{\Psi}_n))$. For the choice of the IV matrix Q_n, its column vector can be linear combinations of X_n $G_n X_n$, $G^2 X_n$, $\ldots$, and columns in $\hat{\Psi}_n$. If for example we set the instrument matrix as $Q_n = (S_n X_n, X_n, \hat{\Psi}_n, S_n \hat{\Psi}_n)$, then terms which need to be analyzed for consistency via some law of large numbers (LLN) are

$$\frac{1}{n} X_n' S_n' S_n X_n, \frac{1}{n} X_n' S_n' X_n, \frac{1}{n} X_n' S_n' \hat{\Psi}_n,$$
$$\frac{1}{n} X_n' S_n' S_n \hat{\Psi}_n, \frac{1}{n} X_n' \hat{\Psi}_n,$$
$$\frac{1}{n} \hat{\Psi}_n' \hat{\Psi}_n, \frac{1}{n} \hat{\Psi}_n' S_n' S_n \hat{\Psi}_n, \frac{1}{n} \hat{\Psi}_n' S_n' S_n \Psi_n,$$
$$\frac{1}{n} X_n' S_n' \Psi_n, \frac{1}{n} X_n' S_n' S_n \xi_n, \frac{1}{n} X_n' S_n \xi_n, \frac{1}{n} \hat{\Psi}_n' S_n \xi_n, \frac{1}{n} \hat{\Psi}_n' S_n' S_n \xi_n,$$
$$\frac{1}{n} X_n' S_n' \hat{\xi}_n, \frac{1}{n} X_n' \hat{\xi}_n, \frac{1}{n} \hat{\Psi}_n' \hat{\xi}_n, \; and \; \frac{1}{n} \hat{\Psi}_n' S_n' \hat{\xi}_n.$$

For the asymptotic distribution of the estimator, we need to consider the stochastic convergence in distribution via central limit theorem (CLT) for some of those terms after proper rescaling.

The 2SSP_IV

For the 2SSP_IV estimator $\hat{\mu}$

$$\begin{aligned} \hat{\mu} - \mu_0 &= [(G_n Y_n, X_n)' \hat{H}_n (\hat{H}_n' \hat{H}_n)^{-1} \\ &\times \hat{H}_n' (I - \hat{Q}_K) \hat{H}_n (\hat{H}_n' \hat{H}_n)^{-1} \\ &\times \hat{H}_n' (G_n Y_n, X_n)]^{-1} (G_n Y_n, X_n)' \hat{H}_n (\hat{H}_n' \hat{H}_n)^{-1} \\ &\times \hat{H}_n' (I - \hat{Q}_K) \hat{H}_n (\hat{H}_n' \hat{H}_n)^{-1} \hat{H}_n' (\varepsilon + \Psi_0(m)). \end{aligned} \tag{21}$$

where the subscript 0 on parameters denotes their true values. From the reduced form of the model (15) we have

$$GnY_n = G_n(I - \phi_0 G_n)^{-1}(\beta_0 X_n + \Psi_0(m) + \varepsilon) = S_n(\beta_0 X_n + \Psi_0(m) + \varepsilon),$$

where $S_n = G_n(I - \phi_0 G_n)^{-1}$ and $S_n(\phi) = G_n(I - \phi G_n)^{-1}$. For the consistency and asymptotic distribution of $\hat{\mu}$ terms we need to analyze are $H_n' H_n$, $H_n'(G_n Y_n, X_n)$, $H_n'\hat{P}$, $\hat{P}'\hat{P}$, $H_n'(\varepsilon + \Psi_0(m))$. For the choice of the IV matrix H_n, its column vector can be linear combinations of $X_n\, G_n X_n$, $G^2 X_n$, ..., and columns in $\hat{P}$. For example, if we choose $\hat{H}_n = (G_n X_n, X_n, \hat{P})$, then the terms which be analyzed are

$$\frac{1}{n}X_n' G_n' S_n X_n,\ \frac{1}{n}X_n' G_n' S_n \Psi_0(m),\ \frac{1}{n}X_n' G_n' S_n \varepsilon,\ \frac{1}{n}X_n' G_n' X_n,\ \frac{1}{n}\hat{P}' S_n X_n,$$
$$\frac{1}{n}\hat{P}' S_n \Psi_0(m),\ \frac{1}{n}\hat{P}' S_n \varepsilon,\ \frac{1}{n}\hat{P}' X_n,\ \frac{1}{n}X_n' S_n X_n,\ \frac{1}{n}X_n' S_n \Psi_0(m),\ \frac{1}{n}X_n' S_n \varepsilon,\ \frac{1}{n}X_n' X_n,$$
$$\frac{1}{n}X_n'\hat{P},\ \frac{1}{n}\hat{P}'\hat{P},\ \frac{1}{n}\hat{P}X_n,\ \frac{1}{n}X_n' G_n' G_n X_n,\ \frac{1}{n}X_n' S_n \Psi_0(m),\ \frac{1}{n}X_n' S_n \varepsilon,\ \frac{1}{n}X_n' X_n,$$
$$\frac{1}{n}X_n' G_n' \varepsilon,\ \frac{1}{n}X_n' G_n' \Psi_0(m),\ \frac{1}{n}X_n' \varepsilon,\ \frac{1}{n}X_n' \Psi_0(m),\ \frac{1}{n}\hat{P}' \Psi_0(m),\ and\ \frac{1}{n}\hat{P}' \varepsilon,$$

for consistency via LLN and some properly rescaled terms for the asymptotic distribution of the estimator via CLT.

4.2 Assumptions

To analyze terms in the above key statistics, we need more topological structures and additional assumptions. Most of the assumptions are used to prove the asymptotic properties of both estimators. However, the 2SSP_IV estimator, given its generality, needs additional assumptions (9-12).

Assumption 4. *1. 2SP_IV $sup_n ||G_n||_\infty = c_g < \infty$.*

2. 2SSP_IV The elements $g_{ij,n} = O(1/h_n)$, uniformly in all i, j. The ratio $\frac{h_n}{n} \to 0$, as n goes to infinity. h_n is a divergent sequence.

Assumption 5. *The parameter $\theta = (\phi, \beta_d', \beta', A', \gamma) \in \Theta$ which is a compact set in the Euclidean space R^{k_θ}. The true parameter θ_0 is contained in the interior of Θ. Furthermore, $sup_{\phi \in \Phi}|\phi| c_g < 1$, where Φ is the parameter space for ϕ.*

Assumption 6. *The support of X_n and A is compact subset of R^K and R respectively. C_n is a bounded function of X_n.*

Assumption 7. *Let $i = 1, \ldots n$ index a random sample of agents from a population. ($g_{ij,n}$, $x_{i,n}$, $y_{i,n}$) are observable.*

Assumption 8. *If $\rho_{ij} > \rho_c$, then $g_{ij} = 0$, where $\rho_c = l_c - \epsilon$, and $l_c = \sup_{i,j \in D_n} \rho_{ij}$ with $\epsilon \to 0$.*

The following assumptions are needed to analyze the terms of the key statistics for the 2SSP_IV estimator.

Assumption 9. *There exists $z(L, G)$ such that for $z_i = z(L_i, G_i)$, $\sqrt{n}(\hat{\lambda} - \lambda) = \sum_{i=1}^{N_d} z_i/\sqrt{N_d} + o_p(1)$, $E(z_i) = 0$ and $E(z_i z_i')$ exist and is non-singular.*

Assumption 10. *$\Psi_0(m)$ is continuously differentiable in m, of orders $c \geq 1$.*

Assumption 11. *There is a* η_0 *with* $\sqrt{n}(\hat{\eta} - \eta_0) = O_p(1)$, *the distribution of* $\tau(m(L, \lambda_0), \eta_0)$ *has an absolutely continuous component with PDF bounded away from zero on its support, which is compact. Also, the first and the second partial derivatives of* $m(L, \lambda)$ *and* $\tau(m, \eta)$ *are bounded in a neighbourhood of* λ_0 *and* η_0, *respectively. w.l.o.g. the function* w *is equal to* τ *and additive.*

Assumption 12. $K = K_n$ *such that* $\sqrt{n}K^{-c+1} \xrightarrow{p} 0$ *and* $p^k(\tau)$ *is a power series with* $c \geq 5$, $K^7/n \to 0$ *and* $K^{1+2c}/h_n = O(1)$.

Assumptions 4 are standard assumptions in the spatial econometrics literature to limit the spatial correlation of the errors. For the 2SP_IV estimator the condition on the adjacency matrix is the same given in Qu and Lee (2015). We require the row of the adjacency matrices to be uniformly bounded, which in turn requires the number of direct links of each node in a network to be bounded. Intuitively, this assumption says, as the sample size increases, the number of direct links of a node cannot go to infinity.

However, in the case of the 2SSP_IV we need to impose a stronger assumption on the topology of the adjacency matrix in order to have key statistics that are NED on the input process. In some empirical applications, it is a practice to have G_n be row-normalized such that $G_i = (g_{i1}, \ldots, g_{in}/\sum_{j=1}^{n} g_{ij})$. In this case since g_{ij} is positive and uniformly bounded, if the $\sum_{j=1}^{n} g_{ij} = O(h_n)$ uniformly in i, then the resulting G_n will have the property assigned to the Assumption 4 (see, e.g., Lee, 2004).

As in Qu and Lee (2015), the distance plays a crucial role in Assumption 8. In fact agents might be linked to other agents in wide area, but once the geographic distance between two agents exceeds a threshold, the two units are not spatially interacted. The second part of the assumption is needed to fit the sparseness requirement of the SAR outcome equation with the network formation estimation as in Graham (2014).

Assumptions 9-12 are needed for the more general case of the 2SSP_IV estimator. Most of the assumptions directly follow from Newey (2009). Assumption 9 requires that $\hat{\lambda}$ be asymptotically equivalent to a sample average which is a function of L and G. It is satisfied by many semiparametric estimators of binary choice models that are $\sqrt{n}-$consistent (see, e.g. Klein and Spady, 1993). Assumption 10 imposes smoothness conditions on function $m(\cdot)$ to control the bias of the estimator. Assumption 11 imposes that the density of τ_i is bounded away from zero. This assumption might be restrictive as observed by Newey (2009). Assumption 12 imposes growth rate restrictions for the number of approximating terms.

Our asymptotic analysis of the proposed estimator is based on inference under NED. The following notion of NED for random field is closely related to Jenish and Prucha (2012).

Definition 1. *For any random vector* Z, $||Z||_p = (E|Z|^p)^{1/p}$ *denotes its* L_p-*norm where* $|\cdot|$ *denotes Eucledean norm. Denote* $\mathfrak{F}_{i,n}(s)$ *as a* $\sigma-$ *field generated by the random vectors* $\zeta_{j,n}$*'s located within the ball* $B_i(s)$, *which is a ball centered at the location* $l(i)$ *with a radius* s *in a* $d-$*dimensional Euclidean space* D.

Definition 2 (NED)**.** *Let* $T = \{T_{i,n};\, l(i) \in D_n,\, n \in N\}$ *and* $\zeta = \{\zeta_{i,n};\, l(i) \in D_n,\, n \in N\}$ *be random fields with* $||T_{i,n}||_p < \infty$, *with* $p \geq 1$, *where* $D_n \subset D$ *and its cardinality* $|D_n| = n$, *and let* $d = \{d_{i,n};\, l(i) \in D_n,\, n \in N\}$ *be an array of finite positive constants. Then the random field* T *is said to be* L_p-*near-epoch dependent on the random field* ζ *if* $||T_{i,n} - E(T_{i,n}|\mathfrak{F}_{i,n}(s))||_p \leq d_{i,n}\varphi(s)$ *for some sequence* $\varphi(s) \geq 0$ *such that* $\lim_{s\to\infty} \varphi(s) = 0$. $\varphi(s)$ *and* $d_{i,n}$ *are called respectively, the NED coefficient and the NED scaling factors.* T *is said to be* L_p-*NED on* ζ *of size* $-\alpha$ *if* $\varphi(s) = O(s^{-\mu})$ *for some* $\mu < \alpha < 0$. *Furthermore when* $\varphi(s) = O(\rho)$, *where* $0 < \rho < 1$ *then* T *is called geometrically* L_p*-NED on* ζ. *If* $\sup_n sup_{l(i)\in D_n} d_{i,n} < \infty$, *then* T *is said to be uniformly* L_p*-NED on* ζ.

4.2.1 Asymptotic inference of the key statistics

Let $\zeta^*_{i,n}$ be a vector valued function of the error term $\zeta_{i,n} = (v^{*'}_{i,n}, \xi_{i,n})'$ and the observed X_n., i.e., $\zeta^*_{i,n} = f_i(v^{*'}_{i,n}, \xi_{i,n}, X_n, \theta_0)$. As X_n is deterministic, $\zeta^*_{i,n}$ is purely determined by the location $l(i)$, independent of the error terms associated with any other places. Let $M_n = A_n' B_n$, where A_n and B_n are either G_n^{m1} or S_n^{m2}, with m_1 and m_2 being finite non-negative integers. The NED property of the statistics $a'\zeta_n^{*'} M_n \zeta_n^* b$, for some constant vectors a and b, and $\zeta_n^* = (\zeta_{1,n}^*, \ldots, \zeta_{n,n}^*)'$ as the basis for the NED is established in Appendix B. Then based on the asymptotic inference under NED, we have the following LLN.

Proposition 1. *Under Assumptions 1,2, 4-8, suppose $sup_{i,n}||\zeta_{i,n}^*||_4 < \infty$ for the 2SP_IV or Assumptions 4-8 hold for the 2SSP_IV then $\frac{1}{n}E|a'\zeta^{*'} M_n \zeta^* b| = O(1)$ and $\frac{1}{n}[a'\zeta^{*'} M_n \zeta^* b - E(a'\zeta^{*'} M_n \zeta^* b)] = o_p(1)$, where a and b are conformable vectors of constants.*

Furthermore, with the compactness of the parameter space of θ, we have the following ULLN.

Proposition 2. *Under Assumptions 1,2, 4-8, suppose $sup_{i,n}||\zeta_{i,n}^*||_4 < \infty$ for the 2SP_IV or Assumptions 4-8 hold for the 2SSP_IV , then $\frac{1}{n}a'\zeta^{*'}(\theta) M_n(\phi)\zeta^*(\theta) b$ is stochastic equicontinuous and*

$$\sup_{\theta\in\Theta} \frac{1}{n} |a'\zeta^{*'}(\theta) M_n(\phi)\zeta^*(\theta) b - E(a'\zeta^{*'}(\theta) M_n(\phi)\zeta^*(\theta) b)| = o_p(1),$$

where $\zeta_{i,n}^*(\theta)' = f_i(v_{i,n}^{*'}, \xi_{i,n}, X_n, \theta)$ and $M_n(\phi) = A_n' B_n$, where A_n and B_n are either G_n^{m1} or $S_n^{m2}(\phi)$, with m_1 and m_2 being finite non-negative integers.

Let us denote

$$R_n = \sum_{j=1}^{m} [a_j' \zeta^{*'} M_{jn} \zeta^* b_j - E(a_j' \zeta^{*'} M_{jn} \zeta^* b_j)] = \sum_{j=1}^{m} r_{i,n},$$

where each M_{jn} matrix can be expressed as $M_{jn} = A_{jn}' B_{jn}$, where A_{jn} and B_{jn} are either G_n^{m1} or S_n^{m2}. Denote $\sigma_{R_n}^2$ as the variance of R_n and $r_{i,n} = \sum_{j=1}^{m} \sum_{k=1}^{n} [a_j' \zeta_{i,n}^{*'} M_{jn}(i,k) \zeta_{k,n}^* b_j - E(a_j' \zeta_{i,n}^{*'} M_{jn}(i,k) \zeta_{k,n}^* b_j)]$. From Qu and Lee (2015) we habe the following CLT for R_n.

Proposition 3. *Under Assumptions 1,2, 4-8, suppose $sup_{i,n}||\zeta_{i,n}^*||_{4+\delta_\epsilon} < \infty$, for some $\delta_\epsilon > 0$ for the 2SP_IV or Assumptions 4-8 hold for the 2SSP_IV ; and $\inf_n \frac{1}{n}\sigma_{R_n}^2 > 0$, then $\frac{R_n}{\sigma_{R_n}^2} \xrightarrow{d} N(0,1)$.*

As in Qu and Lee (2015) these propositions provide essential tools for asymptotic analysis of the consistency and asymptotic normality of the 2SP_IV and the 2SSP_IV estimators.

4.2.2 Consistency and asymptotic normality

To show consistency and asymptotic normality of the 2SIV, we need certain rank conditions on relevant limiting matrices in addition to the convergence of each separated term.

The 2SP_IV

Assumption 13. *Column of Q_n are from $M_n q_n$ and $M_n \Psi_n$, where q_n is strictly exogeneous vector and $M_n = A_n' B_n$ in which A_n and $B_n$are either G_n^{m1} or S_n^{m2} being finite non-negative integers. X_n is a deterministic uniform bounded variable.*[18]

Assumption 14. *$lim_{n\to\infty} \frac{1}{n} E(Q_n' Q_n)$ exists and is nonsingular; $lim_{n\to\infty} \frac{1}{n} E[Q_n'(S_n(\beta_0 X_n + \gamma_0 \Psi_n), X_n, \Psi_n)]$ has full column rank.*

Let us define $\lambda = (\alpha, \beta, A)$.

[18] This assumption could be replaced by stochastic regressors with certain finite moment condition. It is used in order to simplify the derivations.

Theorem 1. *Under Assumption 1,2, 4-8, 13 and 14 the 2SIV estimator $\hat{\kappa}$ is consistent. Furthermore, $\sqrt{n}(\hat{\kappa}-\kappa_0) \xrightarrow{d} N(0, \Sigma_{IV})$ where,*

$$\Sigma_{IV} = plim_{n\to\infty}\frac{1}{n}(J_n' P_{qn} J_n)^{-1} J_n' P_{qn} \Sigma P_{qn} J_n (J_n' P_{qn} J_n)^{-1},$$

with $J_n = (S_n(\beta_0 X_n + \gamma_0 \Psi_n), X_n, \Psi_n)$, $P_{qn} = Q_n(Q_n' Q_n)^{-1} Q_n'$ and with $\Sigma = I_{n-1}\sigma_\xi^2 + (\gamma \frac{\partial}{\partial\lambda}\Psi_n) var(\lambda)(\frac{\partial}{\partial\lambda}\Psi_n'\gamma')$.

However, the estimator is not feasible because ϕ_0 in S_n is not known. In practice, we may use X_n $G_n X_n$, $G^2 X_n$, ..., to have an initial consistent estimate $\hat{\kappa}$ by 2SP_IV, and then using $S_n(\hat{\phi})X_n$ and $S_n(\hat{\phi})\hat{\Psi}_n$ to obtain the feasible best 2SP_IV estimator using Proposition 2. The two estimators has the same limiting distributions. For details see Qu and Lee (2015).

The 2SSP_IV

Let us denote $Q_n^{\perp} = I - Q_n$.

Assumption 15. *$E(\varepsilon^2|m, G_n)$ is bounded. X_n is a deterministic uniform bounded variable.*

Assumption 16. *Column of H_n are from $M_n q_n$ and column of Ψ_n, where q_n is strictly exogeneous vector and $M_n = A_n' B_n$ in which A_n and B_n are either G_n^{m1} or S_n^{m2} being finite non-negative integers.*

Assumption 17. *$lim_{n\to\infty}\frac{1}{n}E(H_n' H_n)$ exists and is nonsingular; $lim_{n\to\infty}\frac{1}{n}E[H_n'(S_n(\beta_0 X_n + \Psi_0(m)), X_n)]$ has full column rank, $lim_{n\to\infty}\frac{1}{n}E(H_n' Q_n^{\perp} H_n)$ has full column rank and $lim_{n\to\infty}\frac{1}{n}E(Q_n' Q_n)$ exists and is non-singular.*

Theorem 2. *Under Assumptions 1-12, and 15 - 17, the 2SIV estimator $\hat{\kappa}$ is consistent. Furthermore, $\sqrt{n}(\hat{\mu}-\mu_0) \xrightarrow{d} N(0, \Sigma_{SPIV})$ where,*

$$\Sigma_{SPIV} = plim_{n\to\infty}\frac{1}{n}(J_n' F_{Kn} Q_n^{\perp} F_{Kn} J_n)^{-1} J_n' F_{Kn} Q_n^{\perp} F_{Kn} \Sigma_2 J_n F_{Kn} Q_n^{\perp} F_{Kn} (J_n' F_{Kn} Q_n^{\perp} F_{Kn} J_n)^{-1},$$

with $\Sigma_2 = I_{n-1}\sigma_\varepsilon^2 + (\frac{\partial}{\partial\lambda}\Psi_0) var(\lambda)(\frac{\partial}{\partial\lambda}\Psi_0)'$.

4.3 Comparison of the estimation approaches

In this section, we discuss the pros and cons of each estimation approach. The parametric approach imposes strong parametric assumptions on the multivariate distributions of the selection and outcome equation disturbances, and it allows the selectivity bias to enter linearly in the outcome equation. The main advantage of this approach is its efficiency when the parametric assumptions are correct. However, to avoid the curse of dimensionality, one need to restrict $cov(u_i, v_{ij})$ to be constant across i and j (Assumption 2). Even though this restriction can be relaxed, it implies the computation of many multidimensional integrals, e.g. the multivariate inverse Mills ratios in equation (15).

The semiparametric approach allows the functional form of the correction term to be completely unknown. In order to include this unknown term one can use an approximating function. If the number of approximating functions grows with the sample size, the estimator is $\sqrt{n}$-consistent and asymptotic normal. In our network model, where the number of terms in the approximating function is embedded in a growing number of instruments (the spatial lags), the estimator precision may be poor if the sample size is low (see Section 2.2).

5 Simulation experiments

In order to evaluate the finite sample performance of the proposed estimators, we conduct a Monte Carlo simulation. The data generating process (DGP) is

$$Y_n = (I - \phi G_n)^{-1}(\beta X_n + U_n), \tag{22}$$

where X_n is an exogenous random variable normally distributed with mean zero and variance equal to $\sigma_x = 1$ and $\beta = 1$. u is generated from a normal distribution with mean equal to zero and variance equal to 8. Links are formed according with the following rule

$$g_{ij} = I(g^*_{ij} \geq 0) = I(\alpha + \gamma(x_i + x_j) + \iota|x_i - x_j| + v_{ij} \geq 0), \tag{23}$$

where $\alpha = -1.5$, $\gamma = -0.1$, and $\iota = -0.3$.[19] This DGP implies that we generate a direct graph, i.e. arcs are not necessarily reciprocal. v_{ij} is a logistically distributed random variable with mean equal to zero and $var(v_{ij}) = \sigma_v = 1$. To create endogeneity in the SAR outcome equation, we generate a bivariate normal random variable $(u_{i,n}, v{*}'_{i,n}) \sim i.i.d.(0, \Sigma_{uv})$, where $\Sigma_{uv} = \begin{pmatrix} 8 & \sigma_{uv} \\ \sigma_{uv} & I_{n-1} \end{pmatrix}$, σ_{uv} is a $n-1$ vector of covariances with constant elements across i and j.

In the simulation, we compare the 2SLS estimator which is commonly used when network exogeneity is assumed, with our parametric 2SP_IV and semiparametric 2SSP_IV estimators given in (17) and (19). We refer to these three methods as 2SLS, 2SP_IV, 2SSP_IV in the tables. Our 2SP_IV and 2SSP_IV treat G_n as endogenous, while the conventional 2SLS estimator estimates only the outcome equation and treats G_n as exogenous.[20] Our aim is to show the magnitude of the bias for the conventional method when links, in fact, form endogenously. To generate different degree of endogeneity, we set the $\sigma_{uv} = 0$, 0.15 and 0.30. To test the flexibility of the nonparametric correction terms, we use three different distributions of the bivariate random variable $(u_{i,n}, v{*}'_{i,n})$, namely student's T with 1, 3, 5, and 7 degree of freedom, Beta and Weibull distributions. All the distributions have the same mean and covariance matrix as the bivariate normal case.[21] We also allow the endogenous effects to be different setting $\phi = 0.07$, 0.08, 0.09, and 0.1 to investigate how the endogenous effects parameter affects estimates.

The setup of our simulations is as follows. The population numerosity is 50, 100 and 500 nodes. Each node is allowed to have endogenous connections. We estimate 1,000 times model (2) with 2SLS, model (4) using the 2SP_IV and 2SSP_IV estimators. The control variable X_n, the error terms and the network are randomly generated for each replication.[22]

Table 1-3 reports the empirical mean (Avg Point Estimation), the empirical standard deviation (Standard Deviation) and the mean squared errors (MSE) of each estimator. Tables 1-3 are given in Appendix D.

Table 1 reports on the performance of each estimator when we let ϕ and the sample size vary. The degree of endogeneity is medium and fixed to $\sigma_{uv} = 0.30$. For each value of ϕ, we show the performance of each estimator in small, medium and large samples (n=50, 100 and 500, respectively).

Table 2 reports on the performance of each estimator when ϕ and the degree of endogeneity changes. For each value of ϕ we allow the degree of endogeneity to vary between *exogeneity*, e.g. $\sigma_{uv} = 0$, *weak endogeneity*,

[19] Node specific fixed effects are removed for the sake of simplicity because we focus on the estimation of the SAR outcome equation in the case of endogenous adgacency matrix.

[20] Model (23) is estimated using logistic regression when the outcome equation estimator is 2SP_IV. Model (23) is estimated using the smoothed maximum score in the semiparametric case, see Horowitz (1992) for details.

[21] In order to keep the density of the network stable we changed the value of the constant of the model (23) for the different bivariate distribution of the errors. See footnote of Table 3.

[22] Qualitative results are not sensitive to the parameters values held constant in Table 1. For the sake of brevity we do not show the output of all simulations

e.g. $\sigma_{uv} = 0.15$ and *medium endogeneity* e.g. $\sigma_{uv} = 0.30$.

Table 3 reports the perfomance of each estimator when we change the bivariate distribution of the errors. Here, the values of ϕ and n are set to 0.1 and 100, respectively. The degree of endogeneity is medium ($\sigma_{uv} = 0.3$).

The simulation results are summarized as follows.

1. For the biases of the parameter estimators, 2SP_IV and 2SSP_IV have very small biases in all cases. The conventional 2SLS estimator is systematically upward biased. The magnitude of bias slightly decreases as the value of ϕ increases. The biases of the endogenous effects $\hat{\phi}$ are in general much higher than those for $\hat{\beta}$.

2. The biases of conventional 2SLS estimator vary with the degree of endogeneity. The higher the degree of endogeneity, i.e., the larger the covariance σ_{uv}, the larger the bias of estimators.

3. Comparing the empirical standard deviations of the estimators, the empirical standard deviation of the conventional 2SLS is systematically greater than 2SP_IV and 2SSP_IV. It increases as the value of σ_{uv} increases as can be seen from Table 2. The empirical standard deviations of 2SP_IV and 2SSP_IV are similar.

4. From Table 1 the bias and the standard deviation of those estimators decreases as the sample size increases.

5. From Table 3 the conventional 2SLS suffers severe upward bias, especially when the bivariate distribution of the errors is Student's T with 1 degree of freedom (more far from the normal distribution) or Weibull or Beta. In the Weibull case the bias of the 2SLS exceeds 100%. 2SP_IV performs slightly better than conventional 2SLS, however, it still suffers a lot from the parametric assumption. 2SSP_IV estimator outperforms all others. This result confirms the criticism on the reliance of distributional assumptions in the selectivity model (see, for example Vella, 1998).

6 Concluding remarks

Many applications of interaction models with network data are based on the assumption that links among economic agents are exogenous. Recent papers, like Hsieh and Lee (2014) and Goldsmith-Pinkham and Imbens (2013), propose a Bayesian approach to estimate network models with endogenous interactions. This paper considers a frequentist approach. Graham (2014) proposed a maximum likelihod estimator for the link formation and derived the statistical properties. Building on his work, we consider two sets of equations: one for the selection model (as in Graham, 2014) and the other for the SAR outcome. Endogeneity arises when there are unobserved factors driving both link formation and outcome. We propose two estimation methods: $2SP\_IV$ and $2SSP\_IV$. The consistency and asymptotic normality of these estimators are proved using the theory of asymptotic inference under near-epoch dependence. We provide Monte Carlo simulations to investigate finite sample properties of our proposed estimators and compare their performances with those under the exogenous weight matrix assumption.

The simulation results indicate that the commonly used estimators are upward biased when the network formation process is endogenous. On the other hand, our estimators have good finite sample properties. As the sample size increases our estimates quickly converge to the true parameters. As we expect, the $2SSP\_IV$ outperforms $2SP\_IV$ when the distribution of the errors departs from normality. This result has also been confirmed in the literature (see, for example Vella, 1998). The two advantages of our method comparing to a

Bayesian approach are efficiency and generalization. Bayesian models are highly time demanding in terms of computational time. Our method is easy to implement and fast to calculate. Second, Bayesian models highly depends on their prior assumptions -while the semiparametric method does not impose any functional form to the bivariate distribution of unobservables.

References

Amemiya, T. (1975). Qualitative response models, *Annals of Economic and Social Measurement, Volume 4, number 3*, NBER, pp. 363–372.

Apicella, C. L., Marlowe, F. W., Fowler, J. H. and Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers, *Nature* **481**(7382): 497–501.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G. and Meghir, C. (2012). Risk pooling, risk preferences, and social networ, *American Economic Journal: Applied Economics* **4**(2): 134–167.

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation, *Econometrica* **68**(5): 1181–1229.

Ballester, C., Calvó-Armengol, A. and Zenou, Y. (2006). Who's who in networks. wanted: the key player, *Econometrica* **74**(5): 1403–1417.

Blume, L. E., Brock, W. A., Durlauf, S. N. and Jayaraman, R. (2014, forthcoming). Linear social interactions models, *Journal of Political Economy* .

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects through social networks, *Journal of Econometrics* **150**: 41–55.

Brock, W. A. and Durlauf, S. N. (2001a). Discrete choice with social interactions, *The Review of Economic Studies* **68**(2): 235–260.

Brock, W. A. and Durlauf, S. N. (2001b). Interactions-based models, *Handbook of econometrics* **5**: 3297–3380.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2009). Peer effects and social networks in education, *The Review of Economic Studies* **76**(4): 1239–1267.

Christakis, N. A., Fowler, J. H., Imbens, G. W. and Kalyanaraman, K. (2010). An empirical model for strategic network formation, *Technical report*, National Bureau of Economic Research.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence, *Journal of econometrics* **92**(1): 1–45.

Cosslett, S. R. (1983). Distribution-free maximum likelihood estimator of the binary choice model, *Econometrica: Journal of the Econometric Society* pp. 765–782.

Dahl, G. B. (2002). Mobility and the return to education: Testing a roy model with multiple markets, *Econometrica* **70**(6): 2367–2420.

De Paula, A., Richards-Shubik, S. and Tamer, E. T. (2014). Identi cation of preferences in network formation games, *Available at SSRN 2577410* .

Fafchamps, M. and Gubert, F. (2007). Risk sharing and network formation, *The American economic review* pp. 75–79.

Goldsmith-Pinkham, P. and Imbens, G. (2013). Social networks and the identification of peer effects, *Journal of Business and Economic Statistics* **31**: 253–264.

Graham, B. S. (2014). An empirical model of network formation: detecting homophily when agents are heterogenous, *Technical report*, National Bureau of Economic Research.

Heckman, J. J. and Robb, R. (1985). Alternative methods for evaluating the impact of interventions: An overview, *Journal of econometrics* **30**(1): 239–267.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model, *Econometrica: journal of the Econometric Society* pp. 505–531.

Horrace, W. C., Liu, X. and Patacchini, E. (2014, forthcoming). Endogenous network production functions with selectivity, *Journal of Econometrics* .

Hsieh, C.-S. and Lee, L. F. (2014). A social interactions model with endogenous friendship formation and selectivity, *Journal of Applied Econometrics* .

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models, *Journal of Econometrics* **58**(1): 71–120.

Jackson, M. O. and Watts, A. (2002). The evolution of social and economic networks, *Journal of Economic Theory* **106**(2): 265–295.

Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks, *Journal of economic theory* **71**(1): 44–74.

Jenish, N. (2013). Spatial semiparametric model with endogenous regressors, *Econometric Theory* pp. 1–26.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields, *Journal of econometrics* **150**(1): 86–98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence, *Journal of econometrics* **170**(1): 178–190.

Khan, S. (2013). Distribution free estimation of heteroskedastic binary response models using probit/logit criterion functions, *Journal of Econometrics* **172**(1): 168–182.

Klein, R. W. and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models, *Econometrica: Journal of the Econometric Society* pp. 387–421.

Lee, L.-F. (1982). Some approaches to the correction of selectivity bias, *The Review of Economic Studies* **49**(3): 355–372.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica* **72**(6): 1899–1925.

Lee, L. F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics* **140**: 333–374.

Lee, L. F., Liu, X. and Lin, X. (2010). Specification and estimation of social interaction models with network structures, *The Econometrics Journal* **13**: 145–176.

Lei, J. (2013). Smoothed spatial maximum score estimation of spatial autoregressive binary choice panel models.

Leung, M. (2014). A random-field approach to inference in large models of network formation, *Available at SSRN* .

Liu, X. (2013b). Identification and efficient estimation of simultaneous equations network models, *Unpublished manuscript.* .

Liu, X. and Lee, L. F. (2010). Gmm estimation of social interaction models with centrality, *Journal of Econometrics* **159**: 99–115.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice, *Journal of econometrics* **3**(3): 205–228.

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

Mele, A. (2010). A structural model of segregation in social networks, *Available at SSRN 1694387* .

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics* **79**(1): 147–168.

Newey, W. K. (2009). Two-step series estimation of sample selection models, *The Econometrics Journal* **12**(s1): S217–S229.

Pinkse, J. and Slade, M. E. (2010). The future of spatial econometrics*, *Journal of Regional Science* **50**(1): 103–117.

Powell, J. L. (1994). Estimation of semiparametric models, *Handbook of econometrics* **4**: 2443–2521.

Powell, J. L., Stock, J. H. and Stoker, T. M. (1989). Semiparametric estimation of index coefficients, *Econometrica: Journal of the Econometric Society* pp. 1403–1430.

Qu, X. and Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix, *Journal of Econometrics* **184**(2): 209–232.

Schmertmann, C. P. (1994). Selectivity bias correction methods in polychotomous sample selection models, *Journal of Econometrics* **60**(1): 101–132.

Sheng, S. (2014). A structural econometric analysis of network formation games.

Su, L. (2012). Semiparametric gmm estimation of spatial autoregressive models, *Journal of Econometrics* **167**(2): 543–560.

Vella, F. (1998). Estimating models with sample selection bias: a survey, *Journal of Human Resources* pp. 127–169.

Appendix A: Some basic properties of NED of random fields and power series

NED of random fields

In the following proofs, we adopt asymptotic inference under NED on $\zeta_{i,n} = (v^{*'}_{i,n}, \xi_{i,n})'$ as basis for the NED process. The following claims are some basic results due to the topological structure in Assumption 1. These basic properties are derived in Jenish and Prucha (2012) and Qu and Lee (2015).

Claim A.1. *For any distance ρ, there are at most $c_1\rho_0^d$ points in $B_i(\rho)$, where c_1 is a positive constant.*

Claim A.1 derives from Jenish and Prucha (2012).

Claim A.2. *For any random field $T = \{T_{i,n};\, l(i) \in D_n,\, n \in N\}$, with $||T_{i,n}||_p < \infty$, we have*

$$||T_{i,n}(m) - E(T_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p \leq 2||T_{i,n}(m)||_p.$$

The results follows from the Minkowsky and conditional Jensen's inequalities.

Claim A.3. *If $||t_{1i,n}(m) - E(t_{1i,n}(m)|\mathfrak{F}_{i,n}(s))||_4 \leq C_1\varphi_1(s)$ and $||t_{2i,n}(m) - E(t_{2i,n}(m)|\mathfrak{F}_{i,n}(s))||_4 \leq C_2\varphi_2(s)$, with $\max(||t_{1i,n}(m)||_4, ||t_{2i,n}(m)||_4) \leq C$, then $||t_{1i,n}(m)t_{2i,n}(m) - E(t_{1i,n}(m)t_{2i,n}(m)|\mathfrak{F}_{i,n}(s))||_2 \leq C(C_1 + C_2)\varphi(s)$, with $\varphi(s) = \max(\varphi_1(s), \varphi_1(s))$.*

The results follows from Qu and Lee (2015) Claim B.3.

The LLN and the CLT under NED inference follows directly from Jenish and Prucha (2012).

Claim A.4. *Under Assumption 1, if the random field $T = \{T_{i,n};\, l(i) \in D_n,\, n \in N\}$ is $L_1 - NED$, the base $\zeta_{i,n}$'s are i.i.d., and $||T_{i,n}||_p < \infty$ uniformly, then $\frac{1}{n}\sum_{i=1}^n (T_{i,n} - ET_{i,n}) \to_{L_1} 0$.*

Claim A.5. *Let $T = \{T_{i,n};\, l(i) \in D_n,\, n \in N\}$ be a random field that is $L_2 - NED$ on the base $\zeta_{i,n}$'s that is i.i.d.. If assumption 1 and the following conditions are met:*

1. *$||T_{i,n}||_{2+\delta} < \infty$ uniformly, for some $\delta > 0$,*
2. *$\inf_n 1/n\sigma^2 > 0$, where $\sigma^2 = var(\sum_{i=1}^n T_{i,n})$,*
3. *NED coefficients satisfy $\sum_{r=1}^\infty r^{d_0-1}\varphi(r) < \infty$*
4. *NED scaling factors satisfy $sup_{n,i\in D} d_{i,n} < \infty$,*

then $1/\sigma \sum_{i=1}^n (T_{i,n} - ET_{i,n}) \xrightarrow{d} N(0,1)$.

Power Series

The following claims are some basic results derived in Newey (2009).

Claim A.6. *Under Assumptions 9 -11, we have that there is that a non-singular linear transformation of $\tilde{p}^K(\tau)$ of $p^K(\tau)$ such that for $\Pi_c(K) = CK^{1+2c}$*

$$E(\tilde{p}^K(\tau_i)\tilde{p}^K(\tau_i)') = I$$

$$\sup_{|\tau|\leq 1}\left|\left|\frac{\partial^c \tilde{p}^K(\tau_i)}{\partial \tau^c}\right|\right| \leq \Pi_c(K)$$

$$\Pi_1(K)K^{1/2}/\sqrt{n} \to 0,\ \Pi_1(K)K^{-c+1}/\sqrt{n} \to 0.$$

Following Newey (2009), we have that by $\frac{\partial w(m_i(L_i,\lambda))}{\partial \lambda}$ bounded and $\sqrt{n}-$consistency of $\hat{\lambda}$, and by $\frac{\partial \tau(m_i,\eta)}{\partial m}$ bounded, $\max_i |\hat{\tau} - \tau| = O_p(1/\sqrt{n})$. From location and scale transformation for power series, which will not change the regression, it can be assumed that $|\hat{\tau}| \leq 1$. As in Newey (1997), from Assumption 12 follows that there is a non-singular linear transformation of $\tilde{p}^K(\tau)$ of $p^K(\tau)$ such that for $\pi_c(K) = CK^{1+2c}$

$$E(\tilde{p}^K(\tau_i)\tilde{p}^K(\tau_i)') = I$$

$$\sup_{|\tau|\leq 1}\left|\left|\frac{\partial^c \tilde{p}^K(\tau_i)}{\partial \tau^c}\right|\right| \leq \pi_c(K)$$

$$\pi_1(K)K^{1/2}/\sqrt{n} \to 0,\ \ \pi_1(K)K^{-c+1}/\sqrt{n} \to 0.$$

Since a non-singular transformation does not change $\hat{\mu}$, without loss of generality we can set $\tilde{p} = p$.

Claim A.7. *Under Assumptions 9 -12, it follows as in Newey (1997, 2009) that there is a γ_K such that for $\psi_K(\tau) = p^k(\tau)'\gamma_K$,*

$$\sup_{|\tau|\leq 1} |\psi_0(\tau) - \psi_K(\tau)| \leq CK^{-c+1}, \qquad \sup_{|\tau|\leq 1}\left|\frac{\partial \psi_0(\tau)}{\partial \tau} - \frac{\partial \psi_K(\tau)}{\partial \tau}\right| \leq CK^{-c+1}.$$

Appendix B: Proofs of NED properties for relevant statistics

In what follows $e_{k,n} = (0, \ldots, 0, 1, \ldots, 0)'$ is the unit column vector with one in its kth entry and zeros in its other entries and $e_n = (1, \ldots, 1)' = \sum_{k=1}^{n} e_{k,n}$. Observe that $I_n = \sum_{i=1}^{n} e_{i,n}e_{i,n}'$.

2SP_IV

Claim B.1. *Under Assumptions 1, 4 and 8, for any positive integer m, $sup_m||G_n^m||_1 \leq c_g c_1 \rho_c^{d_0}$.*

Proof of Claim B.1. See Qu and Lee (2015) Claim C.2.1 □

Claim B.2. *Under Assumptions 1, 4, 5 and 8, for any positive integer m, $sup_{\phi\in\Phi}||S_n(\phi)||_1 \leq \infty$ and $sup_{\phi\in\Phi}||S_n(\phi)||_\infty \leq \infty$.*

Proof of Claim B.2. By applying Claim B.1 see proof Qu and Lee (2015) Claim C.2.2 □

Claim B.3. *Let $t_{i,n}(m) = e_{i,n}G_n^m\zeta_{i,n}^*a$, where $\zeta_{i,n}^* = f_i(\zeta_{i,n}, X_n)$ with $\zeta_{i,n} = (v_{i,n}^{*'}, \xi_{i,n})'$ is a vector-valued function and a is any conformable vector of constants. Under Assumptions 1, 4 and 8, we have that $sup_{i,n}||t_{i,n}(m)||_p < C_{ap}m^{d_0}c_g^m$ and $sup_{i,n}||t_{i,n}(m) - E(t_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p < C_{apm}\varphi(s)$ with C_{apm} and C_{ap} being positive constant; $\varphi(s) = 1$ if $s \leq m\rho c$ and $\varphi(s) = 0$ if $s > m\rho c$.*

Proof of Claim B.3. See proof Claim C.2.5 in Qu and Lee (2015). □

Claim B.4. *Let $h_{i,n}(m) = e_{i,n}'S_n^m(\phi)\zeta_n^*a$, where $\zeta_{i,n}^* = f_i(\zeta_{i,n}, X_n)$ with $\zeta_{i,n} = (v_{i,n}^{*'}, \xi_{i,n})'$ is a vector-valued function and a is any conformable vector of constants. Under Assumptions 1, 4 and 8 suppose $sup_{i,n}||\zeta_{i,n}^*||_p < \infty$, then $sup_{i,n}||h_{i,n}(m)||_p < \infty$ and $sup_{i,n}||h_{i,n}(m) - E(h_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p < C_{apm}\varphi(s)$ with C_{apm} being positive constant; $\varphi(s) = 1$ if $s \leq m\rho c$ and $\varphi(s) = s^{d_0+m-1}|\phi c_g|^{s/\rho_c}$ if $s > m\rho c$.*

Proof of Claim B.4. See proof Claim C.2.6 in Qu and Lee (2015). □

2SSP_IV

Claim B.5. *Under Assumptions 1, 4 and 8, for any positive integer m, $sup_n||G_n^m||_1 = O(\rho_c^{d_0}/h_n)$.*

Proof of Claim B.5. Let us define e_n Consider the kth column sum of G_n^m

$$e'_{i,n}G_n^m e_{k,n} = \sum_i e'_n G_n^{m-1} e_{i,n} e'_{i,n} G_n e_{k,n} \leq ||G_n^{m-1}||_\infty \sum_i e'_{i,n} G_n e_{k,n}. \tag{24}$$

Under Assumption 4 and 8, $\sum_i e'_{i,n}G_n e_{k,n} = \sum_{i\in B_k(\rho_c)} g_{ik,n} = O(h_n/h_n)O(\rho_c^{d_0}) = O(\rho_c^{d_0})$. Hence, $e'_{i,n}G_n^m e_{k,n} = O(h_n^{m-1}/h_n^m)O(\rho_c^{d_0}) = O(\rho_c^{d_0}/h_n)$. This follows for any k and n, we have $sup_n||G_n^m||_1 = O(\rho_c^{d_0}/h_n)$.

□

Claim B.6. *If the i,jth element of G_n^m is not zero, then $\rho_{ij} \leq m\rho_c$*

Proof of Claim B.6. See proof of Claim C.2.3 in Qu and Lee (2015). □

Claim B.7. *For any positive integer p and $0 < q < 1$, if $s \geq p/(-lnq) + 1$, then there exist a finite constant c such that $\sum_{l=[S]} l^p q^l < cs^p q^s$, where s is the largest integer less than or equal to s.*

Proof of Claim B.7. See proof of Claim C.2.4 in Qu and Lee (2015). □

Claim B.8. *Let $t_{i,n}(m) = e_{i,n}G_n^m \zeta_{i,n}^* a$, where $\zeta_{i,n}^* = f_i(\zeta_{i,n}, P)$ with $\zeta_{i,n} = (v_{i,n}^{*'}, \xi_{i,n})'$ is a vector-valued function and a is any conformable vector of constants. Under Assumptions 1, 4, 5, 8 - 12 we have that $sup_{i,n}||t_{i,n}(m)||_p < C_{ap}$ and $sup_{i,n}||t_{i,n}(m) - E(t_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p < C_{apm}\varphi(s)$ with C_{apm} being positive constant; $\varphi(s) = 1$ if $s \leq m\rho c$ and $\varphi(s) = 0$ if $s > m\rho c$.*

Proof of Claim B.8. From Claim B.6, $e'_{i,n}G_n^m e_{k,n} = 0$ if $k \notin B_i(m\rho_c)$. Therefore,

$$|t_{i,n}(m)| = |\sum_k e'_{i,n}G_n^m e_{k,n} e'_{k,n}\zeta_{i,n}^* a| = |\sum_{k\in B_i(m\rho_c)} e'_{i,n}G_n^m e_{k,n} e'_{k,n}\zeta_{i,n}^* a| \leq \max_{k,n}|e'_{i,n}G_n^m e_{k,n}| \sum_{k\in B_i(m\rho_c)} |\zeta_{i,n}^* a|. \tag{25}$$

Thus, we have

$$||t_{i,n}(m)||_p \leq c_1 \frac{(m\rho_c)^{d_0}}{h_n} \sum_{k\in B_i(m\rho_c)} ||\zeta_{i,n}^* a||_p \leq c_1 \frac{(m\rho_c)^{d_0}}{h_n} \pi_c(K), \tag{26}$$

where $sup_{i,n}||\zeta_{i,n}^* a||_p \leq \pi_0(K)$ and $C_{ap} = c_1 \frac{(m\rho_c)^{d_0}}{h_n}\pi_c(K)$. See claim A.6 for a definition of $\pi(\cdot)$. In this case under Assumption 11 $\pi_c(K) = CK^{(1+2c)}$ for power series, see Claim A.6.

Next we show the NED property. We have that all the chains of $e'_{i,n}G_n^m$ related to $t_{i,n}(m)$ are within the ball $B_i(m\rho_c)$. Hence, when $s > m\rho_c$, $(t_{i,n}(m) - E(t_{i,n}(m))|\mathfrak{F}_{i,n}(s)) = 0$. With $s \leq m\rho_c$,

$$||t_{i,n}(m) - E(t_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p \leq 2||t_{i,n}(m)||_p \leq 2C_{ap}.$$

The NED property follows if we choose $\varphi(s) = 1$ for $s \leq m\rho_c$, and $\varphi(s) = 0$ otherwise. □

Claim B.9. *Let $h_{i,n}(m) = e'_{i,n}S_n^m(\phi)\zeta_n^* a$, where $\zeta_{i,n}^* = f_i(\zeta_{i,n}, P)$ with $\zeta_{i,n} = (v_{i,n}^{*'}, \xi_{i,n})'$ is a vector-valued function and a is any conformable vector of constants. Under Assumptions 1, 4, 5, 8 - 12 we have that $sup_{i,n}||h_{i,n}(m)||_p < \infty$ and $sup_{i,n}||h_{i,n}(m) - E(h_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p < C_{apm}\varphi(s)$ with C_{apm} being positive constant; $\varphi(s) = 1$ if $s \leq m\rho c$ and $\varphi(s) = s^{d_0+m-1}|\phi c_g|^{s/\rho_c}$ if $s > m\rho c$.*

Proof of Claim B.9. From the proof of Claim C.1.7 in Qu and Lee (2015), we have $g_{i,n}(m) = \sum_{l=0}^{\infty} C_l^{l+m-1}\phi^l t_{i,n}(l+m)$, where C_l^{l+m-1} is a binomial coefficient. If $\phi = 0$, then $g_{i,n}(m) = t_{i,n}(m)$ and the claim follows from Claim B.8. For $\phi \neq 0$ by Claim B.8, for any i and n,

$$||g_{i,n}(m)||_p \leq c_1 \frac{(m\rho_c)^{d_0}}{h_n}\pi_c(K)\sum_{l=0}^{\infty}|\phi c_g|^l(l+m)^{d_0+m-1},$$

which is finite under Assumption 11 and denoted as C_m. Thus, for $s > 0$,

$$||g_{i,n}(m) - E(g_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p \leq 2||g_{i,n}(m)||_p \leq 2C_m.$$

Now consider the case, when $s > m\rho_c$. Given such as s, from Claim B.8, $t_{i,n}(m+l) - E(t_{i,n}(m+l)|\mathfrak{F}_{i,n}(s) = 0$ for any nonnegative integer l such that $s > (m+l)\rho_c$. Such a set of l will be determined by $l < (s/\rho_c - m)$. Thus, when $s > m\rho_c$

$$\begin{aligned} ||g_{i,n}(m) - E(g_{i,n}(m)|\mathfrak{F}_{i,n}(s))||_p &= \\ &= \left|\left|\sum_{l=[s/\rho_c-m]}^{\infty}(l+m)^{m-1}|\phi|^l||t_{i,n}(m+l)\right|\right|_p \\ &\leq 2C_{ap}c_g^m\sum_{l=[s/\rho_c-m]}^{\infty}(l+m)^{m-1+d_0}|\phi c_g|^l. \end{aligned} \tag{27}$$

The last inequality follows from Claim B.7. As $s/\rho_c > m$, we have

$$\begin{aligned} \sum_{l=[s/\rho_c-m]}^{\infty}(l+m)^{m-1+d_0}|\phi c_g|^l/|\phi|^m &= \sum_{l=[s/\rho_c]}^{\infty}(l)^{m-1+d_0}|\phi c_g|^l/|\phi|^m/|\phi|^m \\ &= O(s^{m-1+d_0}|\phi c_g|^{s/\rho_c}) \end{aligned}$$

if $s > m\rho_c$.

□

Appendix C: Proofs of main results

Proof of Proposition 1. Under Assumptions 1,2, 4-8, suppose $sup_{i,n}||\zeta_{i,n}^*||_4 < \infty$ for the 2SP_IV or Assumptions 4-8 hold for the 2SSP_IV , Claims A.3, B.8 and B.9 the result follows from Claim A.4 and Proposition 1 in Qu and Lee (2015). □

Proof of Proposition 2. From Proposition 1 we have that

$$\frac{1}{n}|a'\zeta^{*'}(\theta)M_n(\phi)\zeta^*(\theta)b - E(a'\zeta^{*'}(\theta)M_n(\phi)\zeta^*(\theta)b)| = o_p(1).$$

For each $\theta \in \Theta$. To show ULLN as the parameter space of θ is compact, we only need to show stochastic equicontinuity of $\frac{1}{n}|a'\zeta^{*'}(\theta)M_n(\phi)\zeta^*(\theta)b|$. Let us define $\lambda = (\alpha, \beta', A') \subset \Theta$, we have three cases: *(a)* $\zeta^{*'}(\theta) = \Psi(\lambda)$; *(b)* $\zeta^{*'}(\theta) \neq \Psi(\lambda)$; $\zeta^*(\lambda) = P(\tau)$.
(a) By the mean value theorem,

$$\begin{aligned} & |a'\Psi(\lambda_1)'M_n(\phi_1)\Psi(\lambda_1)b - a'\Psi(\lambda_2)'M_n(\phi_2)\Psi(\lambda_2)b| \qquad (28)\\ = \; & |(\lambda_1-\lambda_2)a'(\frac{\partial}{\partial\lambda}\Psi'(\bar{\lambda})M_n(\bar{\phi})\Psi(\bar{\lambda}) + \Psi(\bar{\lambda})M_n(\bar{\phi})\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))b \\ + \; & (\phi_1-\phi_2)a'\Psi'\frac{\partial}{\partial\phi}M_n(\bar{\theta})\Psi b| \\ \leq \; & |\lambda_1-\lambda_2||a'(\frac{\partial}{\partial\lambda}\Psi'(\bar{\lambda})M_n(\bar{\phi})\Psi(\bar{\lambda}))b| + |a'(\Psi(\bar{\lambda})M_n(\bar{\phi})\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))b| \\ + \; & |\phi_1-\phi_2||a'\Psi'(\bar{\lambda})\frac{\partial}{\partial\phi}M_n(\bar{\phi})\Psi(\bar{\lambda})b|. \end{aligned}$$

Let us focus on the first term $|\lambda_1 - \lambda_2||a'(\frac{\partial}{\partial\lambda}\Psi'(\bar{\theta})M_n(\bar{\phi})\Psi(\bar{\theta}))b|$. We have

$$\begin{aligned} |\lambda_1-\lambda_2||a'(\frac{\partial}{\partial\lambda}\Psi'(\bar{\lambda})'M_n(\bar{\phi})\Psi(\bar{\lambda}))b| \leq \; & |\lambda_1-\lambda_2||(a'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))a)^{1/2}((b'\Psi(\bar{\lambda}))'M_n(\bar{\phi})'M_n(\bar{\phi})\Psi(\bar{\lambda})b)^{1/2} \qquad (29)\\ \leq \; & |\lambda_1-\lambda_2||(a'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))a)^{1/2}((b'\Psi(\bar{\lambda}))'\Psi(\bar{\lambda})b)^{1/2} \\ \times \; & [\mu_{max}(M_n(\bar{\phi})'M_n(\bar{\phi}))]^{1/2} \\ \leq \; & |\lambda_1-\lambda_2||(a'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))a)^{1/2}((b'\Psi(\bar{\lambda}))'\Psi(\bar{\lambda})b)^{1/2} \\ \times \; & [\sup_{\phi\in\Phi}||M_n(\bar{\phi})'M_n(\bar{\phi})||_\infty]^{1/2}. \end{aligned}$$

Where $\bar{\theta} = (\bar{\lambda}, \bar{\phi})$ lies on the segment line between λ and ϕ, and μ_{max} is the largest eigenvalue of the matrix inside. The first inequality is from the Cuachy-Schawarz inequality, the second inequality holds since $M_n(\bar{\phi})'M_n(\bar{\phi})$ is non-negative definite, and the last inequality is from the spectral radius theorem. From Claim B.2, $\sup_{\phi\in\Theta}||S_n(\phi)||_\infty < \infty$ and $\sup_{\phi\in\Theta}||S_n(\phi)||_1 < \infty$. This implies $\sup_{\phi\in\Phi}||M_n(\bar{\phi})'M_n(\bar{\phi})||_\infty < \infty$. As shown in Amemiya (1975) $\partial\lambda F_{v*} = f_{v^*}\lambda L$ and $\frac{\partial}{\partial\lambda}f_{v*} = \frac{1}{\sigma_{v^*}}\lambda L f_{v^*} L$, with $L = (C_{ij}, a_i, a_j, \rho_{ij})$, given Assumption 6 the F is bounded away from zero and with the continuity of the density function this implies that the terms are $O_p(1)$. It follows that $(a'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))'(\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))a) = O_p(1)$ and $((b'\Psi(\bar{\lambda}))'\Psi(\bar{\lambda})b) = O_p(1)$.

The second term of the summation, $|a'(\Psi(\bar{\lambda})M_n(\bar{\phi})\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))b|$ is symmetric and therefore it obtains the same bound in probability.

The last term is $|\phi_1 - \phi_2||a'\Psi'(\bar{\lambda})\frac{\partial}{\partial\phi}M_n(\bar{\phi})\Psi(\bar{\lambda})b|$. Let us define $W_n = \frac{\partial}{\partial\phi}M_n(\phi)$, following the same argument we have

$$|\phi_1-\phi_2||a'\Psi'(\bar{\lambda})W_n(\bar{\phi})\Psi(\bar{\lambda})b| < |\phi_1-\phi_2||(a'\Psi(\bar{\lambda}))'(\Psi(\bar{\lambda}))a)^{1/2}((b'\Psi(\bar{\lambda}))'\Psi(\bar{\lambda})b)^{1/2}[\sup_{\phi\in\Phi}||W_n(\bar{\phi})'W_n(\bar{\phi})||_\infty]^{1/2}.$$

Given that $\sup_{\phi\in\Phi}||M_n(\bar{\phi})||_\infty < \infty$ and $\sup_{\phi\in\Phi}||M_n(\bar{\phi})||_1 < \infty$ we have that $\sup_{\phi\in\Phi}||W_n(\bar{\phi})'W_n(\bar{\phi})||_\infty < \infty$.

Finally we can write

$$\begin{aligned}
|a'\Psi(\lambda_1)'M_n(\phi_1)\Psi(\lambda_1)b - a'\Psi(\lambda_2)'M_n(\phi_2)\Psi(\lambda_2)b| &\leq |\lambda_1-\lambda_2|[|a'(\frac{\partial}{\partial\lambda}\Psi'(\bar{\lambda})M_n(\bar{\phi})\Psi(\bar{\lambda}))b| + |a'(\Psi(\bar{\lambda})M_n(\bar{\phi})\frac{\partial}{\partial\lambda}\Psi(\bar{\lambda}))b|]\\
&+ |\phi_1-\phi_2||a'\Psi'(\bar{\lambda})\frac{\partial}{\partial\phi}M_n(\bar{\phi})\Psi(\bar{\lambda})b|.\\
&\leq |\lambda_1-\lambda_2|O_p(1)\\
&+ |\phi_1-\phi_2|O_p(1).
\end{aligned}$$

$\sup_{|\phi_1-\phi 2|<\phi^*,|\lambda_1-\lambda 2|<\lambda^*} \frac{1}{n}|a'\Psi(\lambda_1)'M_n(\phi_1)\Psi(\lambda_1)b - a'\Psi(\lambda_2)'M_n(\phi_2)\Psi(\lambda_2)b| = O_p(max\{|\phi^*|,|\lambda^*|\})$, then ULLN follows.

(b) when $\zeta^{*'}(\theta) \neq \Psi(\lambda)$ we only have to show the stochastic equicontinuity of $\frac{1}{n}|a'\zeta^{*'}M_n(\phi)\zeta^* b|$, because θ enters $\zeta^{*'}(\theta)$ polynomially and the parameter space is compact. The proof is simply a particular case of *(a)* with the application of the mean-value theorem for one variable.

(c) We consider only the parameter τ as not constant. The case when ϕ is not constant is specular to the case (a). by the mean value theorem,

$$\begin{aligned}
&|a'P(\tau_1)'M_nP(\tau_1)b - a'P(\tau_2)'M_nP(\tau_2)b| \qquad (30)\\
&= |(\tau_1-\tau_2)a'(\frac{\partial}{\partial\tau}P'(\bar{\tau})M_nP(\bar{\tau}) + P(\bar{\tau})M_n\frac{\partial}{\partial\tau}P(\bar{\tau}))b|\\
&\leq |\tau_1-\tau_2|(|a'(\frac{\partial}{\partial\tau}P'(\bar{\tau})M_nP(\bar{\tau})b| + |a'P(\bar{\tau})M_n\frac{\partial}{\partial\tau}P(\bar{\tau}))b|)
\end{aligned}$$

Let us focus on the first term $|\tau_1-\tau_2||a'(\frac{\partial}{\partial\tau}P'(\bar{\theta})M_nP(\bar{\theta}))b|$. We have

$$\begin{aligned}
|\tau_1-\tau_2||a'(\frac{\partial}{\partial\tau}P'(\bar{\tau})'M_nP(\bar{\tau}))b| &\leq |\tau_1-\tau_2||(a'(\frac{\partial}{\partial\tau}P(\bar{\tau}))'(\frac{\partial}{\partial\tau}P(\bar{\tau}))a)^{1/2}((b'P(\bar{\tau}))'M_n'M_nP(\bar{\tau})b)^{1/2} \qquad (31)\\
&\leq |\tau_1-\tau_2||(a'(\frac{\partial}{\partial\tau}P(\bar{\tau}))'(\frac{\partial}{\partial\tau}P(\bar{\tau}))a)^{1/2}((b'P(\bar{\tau}))'P(\bar{\tau})b)^{1/2}\\
&\times [\mu_{max}M_n'M_n]^{1/2}\\
&\leq |\tau_1-\tau_2||(a'(\frac{\partial}{\partial\tau}P(\bar{\tau}))'(\frac{\partial}{\partial\tau}P(\bar{\tau}))a)^{1/2}((b'P(\bar{\tau}))'P(\bar{\tau})b)^{1/2}\\
&\times [||M_n'M_n||_\infty]^{1/2}.
\end{aligned}$$

Where $\bar{\tau}$ lies between τ_1 and τ_2, and μ_{max} is the largest eigenvalue of the matrix inside. The first inequality is from the Cuachy-Schawarz inequality, the second inequality holds since $M_n'M_n$ is non-negative definite, and the last inequality is from the spectral radius theorem. From Claim B.5, $||M_n'M_n||_\infty = O(\rho_c^{d_0}/h_n)$. As shown in Newey (2009) $\frac{\partial}{\partial\tau}P = O(K^3)$ for power series under Assumption 11. It follows that

$$|a'P(\tau_1)'M_nP(\tau_1)b - a'P(\tau_2)'M_nP(\tau_2)b| \leq |\tau_1 - \tau_2|((a'(\frac{\partial}{\partial\tau}P(\bar{\tau}))'(\frac{\partial}{\partial\tau}P(\bar{\tau}))a)^{1/2}((b'P(\bar{\tau}))'P(\bar{\tau})b))^{1/2} \times [||M_n'M_n||_\infty]^{1/2}. \leq |\tau_1 - \tau_2||O(K^4/h_n). \tag{32}$$

Where the inequality follows from Claim A.6 and Claim B.8. Thus,

$$\sup_{|\tau_1-\tau_2|<\delta^*}\frac{1}{n}|a'P(\tau_1)'M_nP(\tau_1)b - a'P(\tau_2)'M_nP(\tau_2)b| = O_p(\delta^*K^4/h_n).$$

By Assumption 12 $K^4/h_n = O(1)$, then ULLN follows.

The second term of the summation, $|a'P(\bar{\tau})M_n\frac{\partial}{\partial\tau}P(\bar{\tau}))b|$ is symmetric and therefore it obtains the same bound in probability.

Proof of Proposition 3. Under Assumptions 1,2, 4-8, suppose $sup_{i,n}||\zeta^*_{i,n}||_{4+\delta_\epsilon} < \infty$, for some $\delta_\epsilon > 0$ for the 2SP_IV or Assumptions 4-8 hold for the 2SSP_IV , Claims A.3, B.8 and B.9 the result follows from Claim A.5 and Proposition 3 in Qu and Lee (2015). □

Proof of Theorem 1. The estimator is

$$\hat{\kappa} - \kappa_0 = [(G_nY_n, X_n, \hat{\Psi}_n)'\hat{Q}_n(\hat{Q}_n'\hat{Q}_n)^{-1} \times \hat{Q}_n'(G_nY_n, X_n, \hat{\Psi}_n)]^{-1}(G_nY_n, X_n, \hat{\Psi}_n)' \times \hat{Q}_n(\hat{Q}_n'\hat{Q}_n)^{-1}\hat{Q}_n'\hat{\xi}_n. \tag{33}$$

Under Assumptions 1,2, 4-8, 13 and 14 , by applying Proposition 2 we have

$$\frac{1}{n}\hat{Q}_n'(G_nY_n, X_n, \hat{\Psi}_n) - \frac{1}{n}Q_n'(G_nY_n, X_n, \Psi) = o_p(1),$$

$$\frac{1}{n}\hat{Q}_n'\hat{Q}_n' - \frac{1}{n}Q_n'Q_n' = o_p(1),$$

$$\frac{1}{n}\hat{Q}_n'\hat{\xi}_n. - \frac{1}{n}Q_n'\xi_n. = o_p(1).$$

By applying proposition 1, $\hat{\kappa} - \kappa_0 \xrightarrow{p} a\, lim_{n\to\infty}\frac{1}{n}E(\gamma_0Q_n'(v_n^* - \Psi_n) + a\, lim_{n\to\infty}\frac{1}{n}E(Q_n'\varepsilon_n)$, where

$$a = \left(A_q'\left[lim_{n\to\infty}\left(\frac{E(Q_n'Q_n)}{n}\right)\right]^{-1}A_q\right)^{-1}A_q'\left[lim_{n\to\infty}\left(\frac{E(Q_n'Q_n)}{n}\right)\right]^{-1},$$

with $A_q = lim_{n\to\infty}\frac{1}{n}[E(Q_n'S_n)\beta_0X_n + E(Q_n'S_n\Psi_n)\gamma_0), E(Q_n')X_n, E(Q_n'\Psi_n))]$.

As $E(Q_n'\varepsilon) = 0$ and $E(Q_n'\xi_n) = 0$, we have $\hat{\kappa} - \kappa_0 \xrightarrow{p} 0$. Under given assumptions, since $\hat{\kappa} - \kappa_0$ can be written as a form of R_n in proposition 3, $\sqrt{n}(\hat{\kappa} - \kappa_0) \xrightarrow{d} N(0, \Sigma_{IV})$. In particular,

$(\hat{\kappa} - \kappa_0) = plim_{n\to\infty}\frac{1}{n}(J_n'P_{qn}J_n)^{-1}J_n'P_{qn}(\xi_n + \gamma(\frac{\partial}{\partial\lambda}\Psi_n))(\hat{\lambda} - \lambda)$, with $J_n = (S_n(\beta_0X_n + \gamma_0\Psi_n), X_n, \Psi_n)$,

$P_{qn} = Q_n(Q_n'Q_n)^{-1}Q_n'$ and $\lambda = (\alpha, \beta, A)$. Thus,

$$\Sigma_{IV} = plim_{n\to\infty}\frac{1}{n}(J_n'P_{qn}J_n)^{-1}J_n'P_{qn}\Sigma P_{qn}J_n(J_n'P_{qn}J_n)^{-1},$$

with $\Sigma = I_{n-1}\sigma_\xi^2 + (\gamma\frac{\partial}{\partial\lambda}\Psi_n)var(\lambda)(\frac{\partial}{\partial\lambda}\Psi_n'\gamma')$.

$\square$

Proof of Theorem 2. The estimator is

$$\begin{aligned}
\hat{\mu} - \mu_0 &= [(G_nY_n, X_n)'\hat{H}_n(\hat{H}_n'\hat{H}_n)^{-1} \\
&\times \hat{H}_n'(I - \hat{Q}_K)\hat{H}_n(\hat{H}_n'\hat{H}_n)^{-1} \\
&\times \hat{H}_n'(G_nY_n, X_n)]^{-1}(G_nY_n, X_n)'\hat{H}_n(\hat{H}_n'\hat{H}_n)^{-1} \\
&\times \hat{H}_n'(I - \hat{Q}_K)\hat{H}_n(\hat{H}_n'\hat{H}_n)^{-1}\hat{H}_n'(\varepsilon_n + \Psi_0(\tau)).
\end{aligned} \tag{34}$$

Under Assumptions 1-12, and 15 - 17, by applying Proposition 2 we have

$$\frac{h_n}{n}\hat{H}_n'(G_nY_n, X_n) - \frac{1}{n}H_n'(G_nY_n, X_n) = o_p(1),$$

$$\frac{h_n}{n}\hat{H}_n'\hat{H}_n' - \frac{1}{n}H_n'H_n' = o_p(1),$$

$$\frac{h_n}{n}\hat{H}_n'\hat{Q}_n\hat{H}_n. - \frac{1}{n}H_n'Q_nH_n = o_p(1).$$

By applying proposition 1, $\hat{\kappa} - \kappa_0 \xrightarrow{p} a\, lim_{n\to\infty}\frac{1}{n}E(H_n'\varepsilon_n) + a\, lim_{n\to\infty}\frac{1}{n}E(H_n'\Psi_0)$, where

$$\begin{aligned}
a &= \left(A_K'\left[lim_{n\to\infty}\left(\frac{E(H_n'H_n)}{n}\right)^{-1}B_K\, lim_{n\to\infty}\left(\frac{E(H_n'H_n)}{n}\right)^{-1}\right]A_K\right)^{-1} \\
&\times A_K\left[lim_{n\to\infty}\left(\frac{E(H_n'H_n)}{n}\right)^{-1}B_K\, lim_{n\to\infty}\left(\frac{E(H_n'H_n)}{n}\right)\right],
\end{aligned}$$

with $A_K = lim_{n\to\infty}\frac{1}{n}[E(H_n'S_n)\beta_0X_n), E(H_n'X_n)]$, and $B_K = lim_{n\to\infty}\left(\frac{E(H_n'(I-Q)H_n)}{n}\right)A_K'\left[lim_{n\to\infty}\left(\frac{E(Q_n'Q_n)}{n}\right)\right]^{-1}$.

As $E(H_n'\varepsilon) = 0$ and $E(H_n'\Psi_0) = 0$, we have $\hat{\mu} - \mu_0 \xrightarrow{p} 0$. Under given assumptions, since $\hat{\mu} - \mu_0$ can be written as a form of R_n in proposition 3, $\sqrt{n}(\hat{\kappa} - \kappa_0) \xrightarrow{d} N(0, \Sigma_{SPIV})$. In particular,

$\hat{\mu} - \mu_0 = plim_{n\to\infty}\frac{1}{n}(J_n'F_{Kn}Q_n^\perp F_{Kn}J_n)^{-1}J_n'F_{Kn}Q_n^\perp F_{Kn}(\varepsilon_n + \Psi_0)$, with $J_n = (S_n(\beta_0X_n + \gamma_0\Psi_0), X_n, P)$, $F_{Kn} = H_n(H_n'H_n)^{-1}H_n'$ and $Q_n^\perp = I - Q_n$.

Let $\tilde{\psi_i} = \psi_0(\hat{\tau}_i)$, $\psi_i = \psi_0(\tau_i)$, $\tilde{\psi}_{Ki} = \psi_K(\hat{\tau}_i)$, $\psi_{Ki} = \psi_K(\tau_i)$. Then

$$J_n'F_{Kn}Q_n^\perp F_{Kn}\Psi_0/\sqrt{n} = J_n'F_{Kn}Q_n^\perp F_{Kn}(\Psi - \tilde{\Psi})/\sqrt{n} + J_n'F_{Kn}Q_n^\perp F_{Kn}(\tilde{\Psi} - \tilde{\Psi}_K)/\sqrt{n}.$$

This follows from $F_{Kn}Q_n^\perp F_{Kn} = Q_n^\perp F_{Kn}Q_n^\perp$, see also Newey (2009) p.S226.

Let us start with the second term $J_n'F_{Kn}Q_n^\perp F_{Kn}(\tilde{\Psi} - \tilde{\Psi}_K)/\sqrt{n}$.

From Claims A.6 and A.7, by an expansion $\psi_0(\hat{\tau}_i) - \psi_K(\hat{\tau}_i)$ around τ_i we have

$$||\tilde{\psi} - \tilde{\psi_K} - \psi + \psi_K|| \leq \sqrt{n} \sup_{|\tau|\leq 1} \left| \frac{\partial \psi_0(\tau)}{\partial \tau} - \frac{\partial \psi_K(\tau)}{\partial \tau} \right| \max_{i\leq n} |\hat{\tau}_i - \tau_i| = O_p(K^{-s+1}) \xrightarrow{p} 0.$$

Therefore, $J_n' F_{Kn} Q_n^{\perp} F_{Kn}(\tilde{\Psi} - \tilde{\Psi}_K)/\sqrt{n} = J_n' F_{Kn} Q_n^{\perp} F_{Kn}(\Psi - \Psi_K)/\sqrt{n} + o_p(1)$.

From Claim A.7 we have $J_n' F_{Kn} Q_n^{\perp} F_{Kn}(\Psi - \Psi_K)/\sqrt{n} = o_p(1)$. By triangular inequality we have $J_n' F_{Kn} Q_n^{\perp} F_{Kn} \tilde{\Psi}/\sqrt{n} \xrightarrow{p} 0$.

Let us focus on the first term $J_n' F_{Kn} Q_n^{\perp} F_{Kn}(\Psi - \tilde{\Psi})/\sqrt{n}$. Let $\theta = (\alpha, \eta')'$, $\tau(m, \theta) = \tau(m(\lambda L), \theta)$ and $\psi_{\theta i} = \frac{\partial \psi_0(\tau(m,\theta_0))}{\partial \theta'}$. Then, by a second-order expansion and $\sqrt{n}-$consistency of $\hat{\theta}$

$$\begin{aligned}
J_n' F_{Kn} Q_n^{\perp} F_{Kn}(\Psi - \tilde{\Psi})/\sqrt{n} &= -[J_n' F_{Kn} Q_n^{\perp} F_{Kn} \Psi_\theta / n]\sqrt{n}(\hat{\theta} - \theta_0) \\
&= -[J_n' F_{Kn} Q_n^{\perp} F_{Kn}/n] Y \sqrt{n}(\hat{\lambda} - \lambda_0), \\
&= -[J_n' F_{Kn} Q_n^{\perp} F_{Kn}/n] Y z_i/\sqrt{n} + o_p(1),
\end{aligned}$$

where $Y = E(\frac{\partial \psi_0(m_i)}{\partial m_i} \frac{\partial m_i(L_i,\lambda)}{\partial \lambda})$

$$\Sigma_{SPIV} = plim_{n\to\infty} \frac{1}{n}(J_n' F_{Kn} Q_n^{\perp} F_{Kn} J_n)^{-1} J_n' F_{Kn} Q_n^{\perp} F_{Kn} \Sigma_2 J_n F_{Kn} Q_n^{\perp} F_{Kn} (J_n' F_{Kn} Q_n^{\perp} F_{Kn} J_n)^{-1},$$

with $\Sigma_2 = I_{n-1}\sigma_\varepsilon^2 + (\frac{\partial}{\partial \lambda}\Psi) var(\lambda) (\frac{\partial}{\partial \lambda}\Psi)'$.

$\square$

Appendix D: Tables

Table 1: Estimates from adjacency matrix with medium endogeneity varying endoegenous effects or sample size

Replications: 1000

		Parameter	2SLS Avg Point Estimation	2SLS Standard Deviation	2SLS MSE	2SP_IV Avg Point Estimation	2SP_IV Standard Deviation	2SP_IV MSE	2SSP_IV Avg Point Estimation	2SSP_IV Standard Deviation	2SSP_IV MSE
	n=50	$\hat{\phi}$	0.099	0.182	0.184	0.074	0.127	0.127	0.083	0.135	0.135
		$\hat{\beta}$	0.985	0.337	0.338	0.977	0.312	0.313	0.971	0.410	0.411
		$\hat{\alpha}$				-1.461	0.068	0.079	-1.812	0.048	0.316
		$\hat{\gamma}$				-0.076	0.020	0.031	-0.093	0.003	0.008
		$\hat{\iota}$				-0.285	0.050	0.053	-0.308	0.002	0.008
$\phi = 0.07$	n=100	$\hat{\phi}$	0.102	0.177	0.180	0.071	0.113	0.114	0.075	0.121	0.121
		$\hat{\beta}$	0.994	0.353	0.353	1.010	0.336	0.336	1.001	0.441	0.441
		$\hat{\alpha}$				-1.815	0.066	0.081	-1.453	0.048	0.319
		$\hat{\gamma}$				-0.093	0.022	0.032	-0.077	0.003	0.008
		$\hat{\iota}$				-0.308	0.050	0.051	-0.289	0.002	0.009
	n=500	$\hat{\phi}$	0.086	0.058	0.060	0.070	0.032	0.034	0.068	0.047	0.047
		$\hat{\beta}$	0.995	0.125	0.125	0.998	0.120	0.120	0.998	0.154	0.154
		$\hat{\alpha}$				-1.999	0.010	0.010	-2.373	0.061	0.378
		$\hat{\gamma}$				-0.097	0.003	0.004	-0.091	0.002	0.009
		$\hat{\iota}$				-0.296	0.008	0.009	-0.306	0.002	0.007
	n=50	$\hat{\phi}$	0.100	0.177	0.178	0.084	0.128	0.126	0.075	0.130	0.131
		$\hat{\beta}$	0.994	0.358	0.358	0.994	0.341	0.339	0.998	0.446	0.448
		$\hat{\alpha}$				-1.454	0.065	0.080	-1.811	0.049	0.315
		$\hat{\gamma}$				-0.079	0.021	0.030	-0.093	0.003	0.008
		$\hat{\iota}$				-0.291	0.050	0.050	-0.308	0.002	0.008
$\phi = 0.08$	n=100	$\hat{\phi}$	0.114	0.182	0.185	0.077	0.117	0.119	0.082	0.129	0.130
		$\hat{\beta}$	1.022	0.363	0.364	1.023	0.345	0.346	1.021	0.448	0.449
		$\hat{\alpha}$				-1.456	0.065	0.078	-1.814	0.048	0.318
		$\hat{\gamma}$				-0.078	0.022	0.031	-0.093	0.003	0.008
		$\hat{\iota}$				-0.288	0.049	0.050	-0.308	0.002	0.009
	n=500	$\hat{\phi}$	0.109	0.187	0.165	0.081	0.115	0.117	0.080	0.122	0.129
		$\hat{\beta}$	1.011	0.321	0.325	1.016	0.346	0.316	1.017	0.419	0.449
		$\hat{\alpha}$				-1.486	0.069	0.072	-1.414	0.041	0.318
		$\hat{\gamma}$				-0.088	0.025	0.029	-0.097	0.002	0.008
		$\hat{\iota}$				-0.222	0.046	0.051	-0.308	0.002	0.008
	n=50	$\hat{\phi}$	0.105	0.182	0.183	0.095	0.125	0.125	0.082	0.138	0.140
		$\hat{\beta}$	0.996	0.361	0.361	0.996	0.342	0.342	0.995	0.447	0.447
		$\hat{\alpha}$				-1.456	0.067	0.080	-1.812	0.048	0.316
		$\hat{\gamma}$				-0.078	0.021	0.031	-0.093	0.003	0.008
		$\hat{\iota}$				-0.289	0.051	0.052	-0.308	0.002	0.008
$\phi = 0.09$	n=100	$\hat{\phi}$	0.100	0.171	0.172	0.084	0.116	0.116	0.085	0.123	0.129
		$\hat{\beta}$	0.991	0.345	0.345	0.985	0.324	0.324	0.994	0.426	0.426
		$\hat{\alpha}$				-1.448	0.071	0.088	-1.814	0.048	0.317
		$\hat{\gamma}$				-0.078	0.021	0.030	-0.093	0.003	0.008
		$\hat{\iota}$				-0.292	0.052	0.053	-0.308	0.002	0.009
	n=500	$\hat{\phi}$	0.117	0.171	0.142	0.089	0.115	0.117	0.086	0.122	0.124
		$\hat{\beta}$	1.032	0.325	0.335	0.996	0.346	0.316	1.017	0.419	0.420
		$\hat{\alpha}$				-1.486	0.071	0.065	-1.414	0.046	0.317
		$\hat{\gamma}$				-0.088	0.022	0.030	-0.096	0.002	0.007
		$\hat{\iota}$				-0.292	0.052	0.052	-0.318	0.002	0.009
	n=50	$\hat{\phi}$	0.117	0.173	0.176	0.107	0.121	0.122	0.106	0.134	0.136
		$\hat{\beta}$	1.032	0.344	0.345	1.018	0.328	0.328	1.019	0.419	0.423
		$\hat{\alpha}$				-1.457	0.065	0.078	-1.813	0.048	0.317
		$\hat{\gamma}$				-0.078	0.021	0.030	-0.093	0.003	0.010
		$\hat{\iota}$				-0.287	0.049	0.051	-0.308	0.002	0.009
$\phi = 0.10$	n=100	$\hat{\phi}$	0.115	0.178	0.185	0.105	0.114	0.113	0.095	0.123	0.129
		$\hat{\beta}$	0.978	0.342	0.346	0.986	0.318	0.320	0.985	0.414	0.421
		$\hat{\alpha}$				-1.450	0.066	0.083	-1.712	0.049	0.316
		$\hat{\gamma}$				-0.076	0.021	0.032	-0.093	0.003	0.008
		$\hat{\iota}$				-0.290	0.050	0.051	-0.308	0.002	0.008
	n=500	$\hat{\phi}$	0.112	0.165	0.172	0.101	0.111	0.115	0.098	0.120	0.121
		$\hat{\beta}$	0.958	0.323	0.335	0.985	0.312	0.346	1.019	0.413	0.419
		$\hat{\alpha}$				-1.498	0.066	0.082	-1.652	0.045	0.313
		$\hat{\gamma}$				-0.087	0.019	0.030	-0.097	0.003	0.008
		$\hat{\iota}$				-0.310	0.033	0.052	-0.318	0.002	0.007

Notes: $\beta = 1.0$, $\alpha = -1.5$, $\gamma = -0.1$,and $\iota = -0.3$. $\sigma_{uv} = 0.3$. The number of elements in the approximating functions in 2SSP_IV is increasing in n.

Table 2: Estimates from adjacency matrix with different degree of endogeneity varying endoegenous effects

Replications: 1000

		Parameter	2SLS Avg Point Estimation	2SLS Standard Deviation	2SLS MSE	2SP_IV Avg Point Estimation	2SP_IV Standard Deviation	2SP_IV MSE	2SSP_IV Avg Point Estimation	2SSP_IV Standard Deviation	2SSP_IV MSE
		$\hat{\phi}$	0.0731	0.1714	0.1720	0.0718	0.1340	0.1345	0.0724	0.1389	0.1393
		$\hat{\beta}$	1.0024	0.3496	0.3496	0.9978	0.3445	0.3446	1.0023	0.4431	0.4432
	$\sigma_{uv} = 0.0$	$\hat{\alpha}$				-1.4565	0.0702	0.0826	-1.8115	0.0487	0.3153
		$\hat{\gamma}$				-0.0763	0.0214	0.0319	-0.0927	0.0029	0.0078
		$\hat{\iota}$				-0.2859	0.0534	0.0552	-0.3081	0.0024	0.0084
		$\hat{\phi}$	0.0865	0.1691	0.1692	0.0688	0.1370	0.1370	0.0696	0.1393	0.1397
		$\hat{\beta}$	1.0203	0.3509	0.3515	1.0129	0.3455	0.3458	1.0081	0.4464	0.4465
$\phi = 0.07$	$\sigma_{uv} = 0.15$	$\hat{\alpha}$				-1.4562	0.0631	0.0768	-1.8122	0.0485	0.3159
		$\hat{\gamma}$				-0.0791	0.0203	0.0291	-0.0926	0.0031	0.0080
		$\hat{\iota}$				-0.2881	0.0498	0.0512	-0.3081	0.0024	0.0085
		$\hat{\phi}$	0.0990	0.1817	0.1840	0.0705	0.1268	0.1273	0.0677	0.1347	0.1352
		$\hat{\beta}$	0.9854	0.3374	0.3377	0.9766	0.3119	0.3127	0.9708	0.4102	0.4112
	$\sigma_{uv} = 0.30$	$\hat{\alpha}$				-1.4606	0.0684	0.0789	-1.8123	0.0485	0.3160
		$\hat{\gamma}$				-0.0761	0.0203	0.0314	-0.0925	0.0032	0.0081
		$\hat{\iota}$				-0.2846	0.0504	0.0527	-0.3081	0.0024	0.0085
		$\hat{\phi}$	0.0821	0.1432	0.1437	0.0837	0.1228	0.1230	0.0763	0.1258	0.1259
		$\hat{\beta}$	0.9933	0.3409	0.3410	0.9961	0.3421	0.3421	0.9994	0.4516	0.4516
	$\sigma_{uv} = 0.0$	$\hat{\alpha}$				-1.4533	0.0662	0.0810	-1.8137	0.0481	0.3174
		$\hat{\gamma}$				-0.0777	0.0212	0.0308	-0.0928	0.0030	0.0078
		$\hat{\iota}$				-0.2912	0.0514	0.0522	-0.3082	0.0024	0.0085
		$\hat{\phi}$	0.0870	0.1536	0.1537	0.0825	0.1171	0.1173	0.0798	0.1305	0.1309
		$\hat{\beta}$	0.9935	0.3483	0.3483	0.9899	0.3448	0.3449	0.9950	0.4476	0.4476
$\phi = 0.08$	$\sigma_{uv} = 0.15$	$\hat{\alpha}$				-1.4588	0.0657	0.0776	-1.8148	0.0478	0.3184
		$\hat{\gamma}$				-0.0774	0.0204	0.0305	-0.0928	0.0030	0.0078
		$\hat{\iota}$				-0.2866	0.0499	0.0517	-0.3082	0.0024	0.0086
		$\hat{\phi}$	0.1000	0.1772	0.1783	0.0816	0.1256	0.1262	0.0765	0.1290	0.1295
		$\hat{\beta}$	0.9936	0.3584	0.3585	0.9941	0.3394	0.3394	0.9982	0.4459	0.4459
	$\sigma_{uv} = 0.30$	$\hat{\alpha}$				-1.4539	0.0648	0.0796	-1.8115	0.0487	0.3153
		$\hat{\gamma}$				-0.0788	0.0211	0.0299	-0.0927	0.0030	0.0079
		$\hat{\iota}$				-0.2907	0.0496	0.0504	-0.3081	0.0024	0.0084
		$\hat{\phi}$	0.0965	0.1432	0.1433	0.0916	0.1185	0.1186	0.0863	0.1229	0.1237
		$\hat{\beta}$	0.9892	0.3370	0.3372	0.9884	0.3367	0.3369	0.9937	0.4393	0.4393
	$\sigma_{uv} = 0.0$	$\hat{\alpha}$				-1.4615	0.0673	0.0775	-1.8132	0.0483	0.3169
		$\hat{\gamma}$				-0.0776	0.0201	0.0301	-0.0927	0.0031	0.0079
		$\hat{\iota}$				-0.2817	0.0495	0.0528	-0.3082	0.0024	0.0085
		$\hat{\phi}$	0.1120	0.1560	0.1575	0.0912	0.1175	0.1183	0.0849	0.1262	0.1263
		$\hat{\beta}$	0.9945	0.3573	0.3573	1.0001	0.3493	0.3493	1.0007	0.4445	0.4445
$\phi = 0.09$	$\sigma_{uv} = 0.15$	$\hat{\alpha}$				-1.4567	0.0631	0.0765	-1.8112	0.0488	0.3150
		$\hat{\gamma}$				-0.0756	0.0210	0.0322	-0.0925	0.0031	0.0081
		$\hat{\iota}$				-0.2865	0.0480	0.0499	-0.3081	0.0024	0.0084
		$\hat{\phi}$	0.1153	0.1821	0.1838	0.0897	0.1247	0.1249	0.0895	0.1383	0.1384
		$\hat{\beta}$	0.9965	0.3612	0.3612	0.9961	0.3415	0.3415	0.9953	0.4469	0.4469
	$\sigma_{uv} = 0.30$	$\hat{\alpha}$				-1.4559	0.0671	0.0804	-1.8123	0.0485	0.3160
		$\hat{\gamma}$				-0.0778	0.0209	0.0305	-0.0927	0.0030	0.0079
		$\hat{\iota}$				-0.2892	0.0508	0.0520	-0.3081	0.0024	0.0085
		$\hat{\phi}$	0.1001	0.1397	0.1397	0.1022	0.1145	0.1148	0.0997	0.1195	0.1198
		$\hat{\beta}$	1.0118	0.3515	0.3517	1.0087	0.3551	0.3552	1.0093	0.4655	0.4656
	$\sigma_{uv} = 0.0$	$\hat{\alpha}$				-1.4577	0.0664	0.0787	-1.8150	0.0477	0.3186
		$\hat{\gamma}$				-0.0778	0.0204	0.0301	-0.0927	0.0031	0.0079
		$\hat{\iota}$				-0.2869	0.0513	0.0530	-0.3083	0.0024	0.0086
		$\hat{\phi}$	0.1092	0.1542	0.1549	0.1011	0.1082	0.1084	0.1015	0.1204	0.1208
		$\hat{\beta}$	1.0117	0.3434	0.3436	1.0126	0.3347	0.3349	1.0046	0.4299	0.4299
$\phi = 0.10$	$\sigma_{uv} = 0.15$	$\hat{\alpha}$				-1.4537	0.0654	0.0801	-1.8125	0.0484	0.3162
		$\hat{\gamma}$				-0.0784	0.0210	0.0301	-0.0927	0.0030	0.0079
		$\hat{\iota}$				-0.2901	0.0487	0.0497	-0.3081	0.0024	0.0085
		$\hat{\phi}$	0.1167	0.1730	0.1738	0.1025	0.1213	0.1218	0.1009	0.1343	0.1343
		$\hat{\beta}$	1.0319	0.3439	0.3453	1.0176	0.3278	0.3282	1.0192	0.4189	0.4193
	$\sigma_{uv} = 0.30$	$\hat{\alpha}$				-1.4573	0.0652	0.0779	-1.8129	0.0483	0.3166
		$\hat{\gamma}$				-0.0780	0.0209	0.0304	-0.0927	0.0031	0.0079
		$\hat{\iota}$				-0.2866	0.0495	0.0513	-0.3081	0.0024	0.0085

Notes: $\beta = 1.0$, $\alpha = -1.5$, $\gamma = -0.1$,and $\iota = -0.3$. $n = 100$. The number of elements in the approximating functions in 2SSP_IV is 3 (K=3).

Table 3: Estimates from adjacency matrix with different bivariate distribution of the errors

Replications: 1000

	Distribution(dof)	Parameter	2SLS Avg Point Estimation	2SLS Standard Deviation	2SLS MSE	2SP_IV Avg Point Estimation	2SP_IV Standard Deviation	2SP_IV MSE	2SSP_IV Avg Point Estimation	2SSP_IV Standard Deviation	2SSP_IV MSE
		$\hat{\phi}$	0.1264	0.1315	0.1341	0.0950	0.0940	0.0942	0.0980	0.0925	0.0926
		$\hat{\beta}$	1.0135	0.3991	0.3994	1.0193	0.3792	0.3797	1.0143	0.4842	0.4844
$\phi = 0.1$	Student's T(7)	$\hat{\alpha}$				-1.3556	0.0529	0.1537	-1.8267	0.0423	0.3294
		$\hat{\gamma}$				-0.0666	0.0171	0.0375	-0.0930	0.0034	0.0078
		$\hat{\iota}$				-0.2046	0.0357	0.1018	-0.3088	0.0021	0.0091
		$\hat{\phi}$	0.1296	0.1107	0.1146	0.0743	0.0808	0.0848	0.1016	0.0743	0.0745
		$\hat{\beta}$	0.9939	0.5450	0.5450	1.0051	0.5067	0.5067	1.0075	0.6261	0.6261
$\phi = 0.1$	Student's T(5)	$\hat{\alpha}$				-1.2055	0.0496	0.2987	-1.8284	0.0412	0.3310
		$\hat{\gamma}$				-0.0544	0.0157	0.0483	-0.0930	0.0034	0.0078
		$\hat{\iota}$				-0.1505	0.0295	0.1523	-0.3089	0.0021	0.0092
		$\hat{\phi}$	0.1268	0.1086	0.1119	0.0760	0.0830	0.0864	0.1041	0.0733	0.0734
		$\hat{\beta}$	0.9758	0.5291	0.5297	0.9720	0.5102	0.5110	0.9839	0.6463	0.6465
$\phi = 0.1$	Student's T(3)	$\hat{\alpha}$				-1.2057	0.0509	0.2987	-1.8309	0.0393	0.3332
		$\hat{\gamma}$				-0.0523	0.0162	0.0504	-0.0932	0.0033	0.0076
		$\hat{\iota}$				-0.1522	0.0291	0.1507	-0.3090	0.0020	0.0093
		$\hat{\phi}$	0.1348	0.0157	0.0165	0.0698	0.0349	0.0389	0.0978	0.0284	0.0285
		$\hat{\beta}$	1.2832	3.5945	3.6057	1.2063	3.4936	3.4997	1.1626	0.7721	0.7727
$\phi = 0.1$	Student's T(1)	$\hat{\alpha}$				-0.7458	0.0514	0.7560	-1.8348	0.0359	0.3367
		$\hat{\gamma}$				-0.0277	0.0214	0.0754	-0.0930	0.0037	0.0079
		$\hat{\iota}$				-0.0812	0.0283	0.2207	-0.3092	0.0018	0.0094
		$\hat{\phi}$	0.2456	0.0226	0.1473	0.2011	0.0513	0.1134	0.1060	0.0323	0.0339
		$\hat{\beta}$	1.0260	0.3287	0.3298	1.0216	0.3204	0.3211	1.0129	0.4115	0.4117
$\phi = 0.1$	Weibull	$\hat{\alpha}$				-1.2367	0.0520	0.2684	-1.8283	0.0412	0.3309
		$\hat{\gamma}$				-0.0531	0.0168	0.0498	-0.0932	0.0033	0.0076
		$\hat{\iota}$				-0.1528	0.0299	0.1502	-0.3089	0.0021	0.0092
		$\hat{\phi}$	0.2203	0.0502	0.1303	0.1931	0.0401	0.1013	0.1030	0.0650	0.0665
		$\hat{\beta}$	1.0000	0.2193	0.2193	1.0052	0.2062	0.2063	0.9954	0.2955	0.2955
$\phi = 0.1$	Beta	$\hat{\alpha}$				-0.7719	0.1629	0.3169	-0.5998	0.0014	0.0998
		$\hat{\gamma}$				-0.3403	0.0540	0.2463	-0.0796	0.0023	0.0205
		$\hat{\iota}$				-1.3476	0.1653	1.0606	-0.3163	0.0077	0.0180

Notes: $\sigma_{uv} = 0.3$, $n = 100$, $\beta = 1.0$, $\gamma = -0.1$, $\iota = -0.3$, and $\alpha = -1.5$ for all distributions except from Beta distribution where $\alpha = -0.5$. The number of elements in the approximating functions in 2SSP_IV is 3 (K=3).