University of Kent

School of Economics Discussion Papers

Estimation of social interaction models using regularization

Guy Tchuente

July 2016

KDPE 1607

Estimation of social interaction models using regularization

Guy Tchuente*

University of Kent

July 2016

Abstract

In social interaction models, the identification of the network effect is based on either group size variation, structure of the network or the relative position in the network measured by the Bonacich centrality measure. These identification strategies imply the use of many instruments or instruments that are highly correlated. The use of highly correlated instruments may lead to the weak identification of the parameters while, in finite samples, the inclusion of an excessive number of moments increases the bias. This paper proposes regularized versions of the 2SLS and GMM as a solution to these problems. The regularization is based on three different methods: Tikhonov, Landweber Fridman, and Principal Components. The proposed estimators are consistent and asymptotically normal. A Monte Carlo study illustrates the relevance of the estimators and evaluates their finite sample performance.

Keywords: High-dimensional models, Social network, Identification, Spatial autoregressive model, GMM, 2SLS, regularization methods.

*School of Economics and MaGHiC, Email: g.tchuente@kent.ac.uk. Comments from Marine Carrasco and John Peirson are gratefully acknowledge. The author also thanks seminar participants of the School of Economics University of Kent, RES 2016 Econometrics Society European Meeting 2016, African Meeting of the Econometric society 2016 for their comments. Many thanks to Lui for kindly providing his code for bias corrected 2SLS and GMM.

Non-technical summary

This paper considers the estimation of social interaction models with network structures and the presence of endogenous, contextual, correlated and group fixed effects. In network models, an agent's behavior may be influenced by peers' choices (the endogenous effect), by peers' exogenous characteristics (the contextual effect), and/or by the common environment of the network (the correlated effect)(see Manski (1993) for a description of these models).

As discuss in Manski (1993) work on reflection problem in network model, identification and estimation of the endogenous interaction effect are of major interest in social interaction models. Following Manski,recent works have shown that identification of the parameters of the model is based on the structure or the size of the group in the network. For example, identification of the network effect can be achieved by using individuals' Bonacich (1987) centrality as an instrumental variables. However, the number such instruments increases with the number of groups; leading to the many instruments problem. Identification can also be achieved using the friend of a friend exogenous characteristics. Unfortunately, if the network is very dense, the identification is weakened. The variation in group size is another source of identification of the network effect. However, if the groups are very large the identification power is lowered. This paper uses high-dimensional estimation techniques, also know as regularization methods, to estimate network models. The regularization is proposed as a solution to the weak identification problem in network models.

The proposed regularized two stage least square and generalized method of moments based on three regularization methods help to deal with many moments or/and weak identification problems. We show that these estimators are consistent and asymptotically normal. Moreover, the regularized two stage least square estimators are asymptotically unbiased and achieve the asymptotic efficiency bound. The regularized estimators all involve the use of a regularization parameter. An optimal data-driven selection method for the regularization parameter is derived.

A Monte Carlo experiment shows that the regularized estimator performed well. The regularized two stage least square and generalized method of moments procedure substantially reduce the many instruments bias for both the two stage least square and generalized method of moments estimators, specifically in a large sample. Moreover, the qualities in term of bias and precision of the regularized estimator improves with the increase of the network density and the number of groups. These results show that regularization is a valuable solution to the potential weak identification problem existing in network models estimation.

1 Introduction

This paper considers the estimation of social interaction models with network structures and the presence of endogenous, contextual, correlated and group fixed effects. The model considered has the specification of a spatial autoregressive (SAR) model. In network models, an agent's behavior may be influenced by peers' choices (the endogenous effect), by peers' exogenous characteristics (the contextual effect), and/or by the common environment of the network (the correlated effect)(see Manski (1993) for a description of these models).

As discuss in Manski (1993) work on reflection problem in network model, identification and estimation of the endogenous interaction effect are of major interest in social interaction models. Following Manski (1993), Lee (2007) shows that both the endogenous and exogenous interaction effects can be identified if there is sufficient variation in group sizes. However, with large group sizes, identification can be weak in the sense that the estimator converge in distribution at low rates. Bramoulle et al. (2009) use the structure of the network to identify the network effect. Their identification strategy relies on the use of spatial lags of friends' (or the friends of the friends) characteristics as instruments. But, if the network is highly transitive (a friend of my friend is likely to be my friend) the identification is also weak. More recently, Liu and Lee (2010) have considered the estimation of a social network model where the endogenous effect is given by the aggregate choice of an agent's friends. They showed that different positions of the agents in a network captured by the Bonacich (1987) centrality measure can be used as additional instrumental variables (IV) to improve estimation efficiency. The number of such instruments depends on the number of groups and could be very large. Liu and Lee (2010) proposed 2SLS and GMM estimators. The proposed estimators have an asymptotic bias due to the presence of many instruments. As shown by Bekker (1994), the use of many instruments may be desirable to improve asymptotic efficiency. However, finite sample properties of instrumental variable estimators can be sensitive to the number of instruments.

In a linear model framework without network effect, Carrasco (2012) has proposed an estimation procedure that allows the use of many instruments, this number may be smaller or larger than the sample size, or even infinite. Moreover, Carrasco and Tchuente (2016) show that these methods can be used to improve identification in weak instrumental variable estimation.

The present paper proposes regularized two-stage least squares (2SLS) and generalized method of moments (GMM) estimators for network models with SAR representation. High dimensional reduction techniques are used to mitigate the finite sample bias of the 2SLS and GMM estimators coming from the use of many instruments or highly correlated instruments. The regularized 2SLS and GMM estimators are based on three ways to compute a regularized inverse of the (possibly infinite dimensional) covariance matrix of the instruments. The regularization methods are taken from the literature on inverse problem, see Kress (1999) and Carrasco, Florens, and Renault (2007). The first estimator is based on Tikhonov (ridge) regularization. The second estimator is based the iterative Landweber Fridman method. The third estimator is based on the principal components associated with the largest eigenvalues.

The regularized 2SLS and GMM estimators are consistent and asymptotically normal and unbias. The regularized 2SLS achieved the semiparametric efficiency bound. However, the consistency and asymptotic normality require more regularization compared to Carrasco (2012). The regularized GMM estimators for SAR models are also consistent, asymptotically normal and without asymptotic bias. Moreover, the same level of regularization as for the 2SLS is needed to achieve consistency. A Monte Carlo experiment shows that the regularized estimators performed well. In general, the quality of the regularized estimators improve with the density of the network.

The large existing literature on network models has focused its attention on two main issues, the identification and estimation of the network effect. In his seminal work, Manski (1993) showed that the linear-in-means specification suffers from the reflection problem so that endogenous and contextual effects cannot be separately identified. Lee (2007) and Bramoullé, Djebbari, and Fortin (2009) in a local-average network model propose identification strategies based on the difference in group size and group structure. Liu and Lee (2010) show that Bonacich (1987) centrality measure can be used as additional instruments to improve identification and estimation efficiency. Liu and Lee (2010) propose generalized method of moments (GMM) estimation approaches following Kelejian and Prucha (1998, 1999) who have proposed 2SLS and GMM approaches for the estimation of SAR models. The use of these moment methods usually implies the use of many moment conditions(see Donald and Newey (2001), Hansen, Hausman, and Newey (2008) and Hasselt (2010) for some recent developments in this area). We assume, in this paper, that there are many instruments and used a framework that enables us to have more instruments than sample size or even an infinite number of instruments. We, therefore, complement works done in models where the number of instruments exceeds the sample size. For instance, Kuersteiner (2012) considers a kernel weighted GMM estimator, while Okui (2011) uses shrinkage. Bai and Ng (2010) and Kapetanios and Marcellino (2010) also assume that the endogenous regressors depend on a small number of factors which are exogenous, and use estimated factors as instruments, they assume that the number of variable from which the factors are estimated can be larger than the sample size. Belloni, Chen, Chernozhukov, and Hansen (2012) propose an instrumental variable estimator under first stage sparsity assumption. Hansen and Kozbur (2014) propose a ridge regularized jackknife instrumental variable estimator in the presence of heteroskedasticity which does not require sparsity and provide tests with good sizes.

Another important direction of research in the IV estimation literature is on weak instruments or weak identification (see, e.g., Chao and Swanson (2005), Carrasco and Tchuente (2016)). In this paper, we assume that the concentration parameter grows at the same rate as the sample size. Hence, we restrict our attention to scenarios where instruments are stronger than assumed in the weak-instrument literature.

The paper is organized as follows. Section 2 presents the network model. Sections 3 discusses identification and estimation in network models. It proposes the regularized 2SLS and GMM approaches for the estimation of the model. The selection of the regularization parameter is discussed in Section 4. Monte Carlo evidence on the small sample performance of the proposed estimators is given in Section 5. Section 6 concludes.

2 The Model

The following social interaction model is considered.

$$Y_r = \lambda W_r Y_r + X_{1r}\beta_1 + W_r X_{2r}\beta_2 + \iota_{m_r}\gamma_r + u_r \tag{1}$$

with $u_r = \rho M_r u_r + \varepsilon_r$ and $r = 1...\bar{r}$ where $\bar{r}$ is the total number of groups, m_r is the number of individual in the group r.

$Y_r = (y_{1r}, ..., y_{m_r r})'$ is an m_r-dimensional vector, it represents the outcomes of interest. y_{ir} is the observation of the individual i in the group r. The total number of individuals in the sample is $n = \sum_{r=1}^{\bar{r}} m_r$.

W_r and M_r are $m_r \times m_r$ are sociomatrices of known constants. In principle, W_r and M_r may or may not be the same.

λ is a scalar, it captures the endogenous network effects, we assume that this effect is the same for all groups. Outcome of individuals influences those of their successors in the network graph.

X_{1r} and X_{2r} are respectively $m_r \times k_1$ and $m_r \times k_2$. They represents individuals exogenous characteristics. β_1 is the parameter measuring the dependence of individuals' outcomes on their own characteristics. The outcomes of individuals may also depend on the characteristics of their predecessors via the exogenous contextual effect measured by β_2. ι_{m_r} is an m_r-dimensional vector of ones and γ_r represents the unobserved group-specific effect. It is treated as a vector of unknown parameter that we are not going to estimate.

Aside from the group fixed effect, ρ captures unobservable correlated effects of individuals with their connections in the network. ε_r is the m_r-dimensional disturbance vector, ε_{ir} are iid mean 0 and variance σ^2 for all i and r. We define $X_r = (X_{1r}, W_r X_{2r})$.

For a sample with $\bar{r}$ groups, the data is stack up by defining $V = (V_1', ..., V_{\bar{r}}')'$ for $V = Y, X, \varepsilon$ or u.

We also define $W = D(W_1, W_2, ..., W_{\bar{r}})$ and $M = D(M_1, M_2, ..., M_{\bar{r}})$, $\iota = D(\iota_{m_1}, \iota_{m_2}, ..., \iota_{m_{\bar{r}}})$ where $D(A_1, .., A_K)$ is a block diagonal matrix in which the diagonal blocks are $m_k \times n_k$ matrices A_k's , $k = 1, ..., K$.

The full sample model is

$$Y = \lambda WY + X\beta + \iota\gamma + u \tag{2}$$

where $u = \rho M u + \varepsilon$.

We define $R(\rho) = (I - \rho M)$, Cochrane-Orcutt type transformation of the model is obtained by multiplying equation (2) by $R = R(\rho_0)$ where ρ_0 is the true value of the parameter ρ.

$$RY = \lambda RWY + RX\beta + R\iota\gamma + Ru$$

Which lead to the following equation

$$RY = \lambda RWY + RX\beta + R\iota\gamma + \varepsilon. \tag{3}$$

When the number of groups is large, we have the incidental parameter problem (see Neyman and Scott (1948), Lancaster (2000) for the consequences of the incidental parameter problem).

To eliminate the unobserved group heterogeneity, we define

$$J_r = I_{m_r} - (\iota_{m_r}, M_r\iota_{m_r})[(\iota_{m_r}, M_r\iota_{m_r})'(\iota_{m_r}, M_r\iota_{m_r})]^{-}(\iota_{m_r}, M_r\iota_{m_r})'$$

where A^{-} denotes a generalized inverse of a square matrix A. In general, J_r represents the projection of an m_r-dimensional vector on the space spanned by ι_{m_r} and $M_r\iota_{m_r}$ if they are linearly independent. Otherwise, $J_r = I_{m_r} - \frac{1}{m_r}\iota_{m_r}\iota_{m_r}'$, which is the deviation from group mean projector.

The matrix $J = D(J_1, J_2, ..., J_{\bar{r}})$, is then used to pre-multiplied the (3) to gives a model without the unobserved group effect parameters

$$JRY = \lambda JRWY + JRX\beta + J\varepsilon. \tag{4}$$

This is the structural equation and we are interested in the estimation of λ, β_1, β_2 and ρ.

We define $S(\lambda) = I - \lambda W$. Under the assumption that the model (2) represents an equilibrium equation and $S \equiv S(\lambda_0)$ is invertible at the true parameter value. The equilibrium vector Y is given by the reduced form equation.

$$Y = S^{-1}(X\beta + \iota\gamma) + S^{-1}R^{-1}\varepsilon \tag{5}$$

It follows that $WY = WS^{-1}(X\beta + \iota\gamma) + WS^{-1}R^{-1}\varepsilon$ and WY is correlated with ε Hence, in general, (4) cannot be consistently estimated by OLS. Moreover, this model may not be considered as a self-contained system where the transformed variable JRY can be expressed as a function of the exogenous variables and disturbances, and, hence, a partial likelihood type approach may not be feasible base only on (4).

In this paper, we consider the estimation of (4) using regularized 2SLS and regularized GMM.[1]

3 Identification and estimation of the network model

This section presents the identification and estimation of the network model parameters using regularization techniques. The following assumptions are needed.

Assumption 1. The elements of ε_{ir} are *i.i.d.* with zero mean, variance σ^2 and that a moment of order higher than the fourth exists.

Assumption 2. The sequences of matrices $\{W\}$, $\{M\}$, $\{S^{-1}\}$ and $\{R^{-1}\}$ are Uniformly Bounded (UB). And $Sup\|\lambda W\| < 1$.[2]

[1] An extension would be to estimate the same model using LIML (the least variance ratio (LVR)) of Carrasco and Tchuente (2015). In models with independent observations, the LIML estimator can also be derived based on the least variance ratio principle (see, Davidson and MacKinnon (1993)). The LVR estimator is not equivalent to the LIML estimator for the SAR model. This is analogous to the difference between the 2SLS and maximum likelihood estimators for the SAR model Lee (2004)

[2] Uniformly bounded in row (column) sums in absolute value of a sequence of square matrices $\{A\}$ will be abbreviated as UBR (UBC), and uniformly bounded in both row and column sums in absolute value as UB. A sequence of square matrices $\{A\}$, where $A = [A_{ij}]$, is said to be UBR (UBC) if the sequence of row sum matrix norm of A (column sum matrix norm of A) is bounded.

We take $\varepsilon(\rho_0, \delta) = JR(Y - Z\delta) = f(\delta_0 - \delta) + JRWS^{-1}R^{-1}\varepsilon(\lambda_0 - \lambda) + J\varepsilon$

with $f = JR[WS^{-1}(X\beta_0 + \iota\gamma_0), X]$ where λ_0, β_0 and γ_0 are true values of the parameters and $\delta = (\lambda, \beta')'$.

Under the Assumption 2 that $Sup\|\lambda W\| < 1$, the f can be approximated by a linear combination of $(WX, W^2X, ...)$, $(W\iota, W^2\iota, ...)$ and X. This is a typical case where the number of potential instruments is infinite.

Define $Q = J[Q_0, MQ_0]$ with $Q_0 = [WX, W^2X, ...W\iota, W^2\iota, ..., X]$ be the infinite dimensional set of instruments. we can also consider the case where only a finite number of instruments[3], let say $m_1 < n$, is used. For the finite number of instruments case, we define

$$Q_{m_1} = J[Q_{0m_1}, MQ_{0m_1}]$$

with $Q_{0m_1} = [WX, W^2X, ...W^{m_1}X, W\iota, W^2\iota, ..., W^{m_1}\iota, X]$.

As discussed in Liu and Lee (2010), δ is identified if $Q'_{m_1}f$ has full column rank $k+1$. This rank condition requires that f has a full rank $k+1$. Note that this is under the assumption that Q_{m_1} is full rank column (meaning no perfect collinearity between instruments).[4] If W_r does not have equal degrees in all its nodes and W_r is not row-normalized, then the centrality score of each individual in his group helps to identify δ. This is possible even if $\beta_0 = 0$. However, if W_r has constant row sums then $f = JR(WS^{-1}X\beta_0, X)$ and the identification can not hold for $\beta_0 = 0$. Under Assumption 3 δ is identified.[5]

The identification in the general case with infinite number of instruments is possible if the infinite number of row matrix $Q'f$ is full rank column. The identification is based on moment condition $E(Q'\varepsilon(\rho_0, \delta)) = 0$ ie $Q'f(\delta_0 - \delta) = 0$. For any sample size, n, $rank(Q) \leq n$. If we assume that $rank(QQ') = n$, which is not always true, then the rank condition requires only that f has a full rank $k+1$.[6] An the same identification conditions as in the finite dimensional case follow.

[3]The finite set of instrument can be the case when the network effect is very small such the $\lambda^m \to 0$ as $m \to \infty$ at a very fast rate.

[4]Section 3.1 discusses the consequences of near perfect collinearity on identification of the network effect.

[5]These identification results are from Liu and Lee (2010) our work generalized the result to infinite instruments.

[6]Section 3.2 proposes regularization tools that that can insure identification of δ in a regularized version of the orthogonal condition.

3.1 Weak identification and many instruments in a network model

Since Manski (1993), the identification problem in network model has been a major concern. Following his negative result on the ability to separately identify endogenous and exogenous interaction effects, in a linear-in-mean models, many studies have investigated structures in which identification is possible. The identification of the network effect is achieved through group size variation or by using the structure of the network. It is notable that, in all cases, we need to have additional information to overcome the reflection problem.

Lee (2007) uses variations in group sizes to identify both the endogenous and exogenous interaction effects. His identification relies on sufficient variation in group size. Unfortunately, with large group sizes, the identification can be weak in the sense that the estimates converge in distribution at low rates. Using Bramoulle et al. (2009) comments on Lee's identification with two groups of different size, we can show the a large group size implied near perfect collinearity between WX and W^2X.

Bramoulle et al. (2009) use the structure of the network to identify the network effect. Their work proposes a general framework that incorporates Lee's and Manski's setup as special cases. The identification strategy proposed in their work relies on the use of spatial lags of friends (or the friends of the friends) characteristics as instruments. The variables WX, W^2X and W^3X... are used as instruments for WY. The condition for identification is that I, W and W^2 (or I, W, W^2 ande W^3, in the presence of correlated effects) are linearly independent. Variation in group size ensures that I, W and W^2 are linearly independent. However, if the sizes of the group are large WX and W^2X are nearly linearly dependent leading to weak identification. Moreover, if the network is highly transitive (friend of my friend are likely to be my friend; $W \sim W^2$) the identification is also weak. In practice, as pointed out by Gibbons and Overman (2012), the used of WX, W^2X and W^3X... as instruments can lead to near perfect collinearity which implies weak identification.

Liu and Lee (2010) have, more recently, considered the estimation of a social network model. As in Bramoulle et al. (2009) they use the structure of the network to identify the network effect. In addition WX, W^2X and W^3X..., the Bonacich centrality across nodes in a network is used as an IV to identify network effects and improve estimation efficiency. The use of the Bonacich centrality measure usually leads to the use of many instruments. The 2SLS obtained with those instruments is biased because of the use of many instruments. Liu and Lee (2010) propose a bias corrected 2SLS to account for the many instruments bias.

In this paper, we use regularization techniques. These high dimensional reduction techniques enable the use of all instruments and deliver efficiency with better finite sample properties (see Carrasco (2012) and Carrasco and Tchuente (2015)). In this case, the asymptotic efficiency can be obtained by using many (or all potential) instruments. We use both the Bonacich centrality measure and WX, W^2X and W^3X..., as IVs and apply a high-dimensional technique to mitigate the problem of near perfect collinearity resulting from network structure or the bias of many instruments.

3.2 Regularization methods

The estimation of the parameters of interests can be achieved by using instrumental variables. We can use a finite number of instruments or all potential instruments. In the many instruments literature, with an increasing number of instruments, estimation is asymptotically more efficient. However, with a large number of instruments relative to sample size, we have the many instruments problem (see, e.g., Bekker (1994); Donald and Newey (2001); Han and Phillips (2006)). It is possible to use a fixed number of instrumental variables to avoid this problem. The 2SLS estimator with a fixed number of instrumental variables will be consistent and asymptotically normal but may not be efficient. In order to be able to use all potential instruments (Q), we use regularization tools.

We take π to be a positive measure on $\mathbb{N}$.[7] and we denote $l^2(\pi)$ as the Hilbert space of square summable sequence with respect to π in the real space. We define the covariance operator K of the instruments as

$$K : l^2(\pi) \rightarrow l^2(\pi)$$
$$(Kg)_j = \sum_{k\in\mathbb{N}} E(Q_{ji}Q_{ki}\pi_k)$$

Where Q_{ji} is the j^{th} column and i^{th} line of Q. Under the assumption that $|Q_{ji}Q_{ki}|$ for all j, k and i is bounded, K is a compact operator (see Carrasco, Florens, and Renault (2007) for a definition).

We consider λ_j and ϕ_j $j = 1, 2, ...$ to be respectively the eigenvalues (ordered in decreasing order) and the orthogonal eigenvector of K. The operator K can be estimated by K_n defined as:

$$K_n : l^2(\pi) \rightarrow l^2(\pi)$$
$$(K_n g)_j = \sum_{k\in\mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} Q_{ji}Q_{ki}\pi_k$$

[7] For a detailed discussion on the role and choice of π, see Carrasco (2012), Carrasco and Florens (2014) when π is a measure on $\mathbb{R}$. In the present case model, π can be for example $\pi_k = \frac{\lambda^k}{\sum_{k\in\mathbb{N}} \lambda^k}$ with $k \in \mathbb{N}$.

In the SAR model, the number of moment conditions can be infinite. Therefore, the inverse of K_n needs to be regularized because it is nearly singular. By definition (see Kress (1999) , page 269), a regularized inverse of an operator K is

$$R_\alpha : l^2(\pi) \rightarrow l^2(\pi)$$

such that $\lim_{\alpha \to 0} R_\alpha K\varphi = \varphi, \forall \varphi \in l^2(\pi)$.

We consider three different types of regularization schemes: Tikhonov (T), Landwerber Fridman (LF) and Principal Components (PC). They are defined as follows:

- **Principal Component (PC)**

  This method consists in using the first eigenfunctions:

  $$(K^\alpha)^{-1}r = \sum_{j=1}^{1/\alpha} \frac{1}{\lambda_j} \langle r, \phi_j \rangle \phi_j$$

  where $\frac{1}{\alpha}$ is some positive integer.[8] The use of PC in the first stage is equivalent to projecting on the first principal components of the set of IVs.

- **Tikhonov (T)**

  Also known as the ridge regularization
  $(K^\alpha)^{-1}r = (K^2 + \alpha I)^{-1}Kr$

  $$(K^\alpha)^{-1}r = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \alpha} \langle r, \phi_j \rangle \phi_j$$

  where $\alpha > 0$ and I is the identity operator.

- **Landweber Fridman (LF)**

  Let $0 < c < 1/\|K\|^2$ where $\|K\|$ is the largest eigenvalue of K (which can be estimated by the largest eigenvalue of K_n).

  $$(K^\alpha)^{-1}r = \sum_{j=1}^{\infty} \frac{[1 - (1 - c\lambda_j^2)^{\frac{1}{\alpha}}]}{\lambda_j} \langle r, \phi_j \rangle \phi_j.$$

  where $\frac{1}{\alpha}$ is some positive integer.

[8] $\langle ., . \rangle$ represents the scalar product in $l^2(\pi)$ and in $\mathbb{R}^n$ (depending on the context).

In the case of a finite number of moments, $P_{m_1} = Q_{m_1}(Q'_{m_1}Q_{m_1})^{-}Q'_{m_1}$ is the projection matrix on the space of instruments. The matrix $Q'_{m_1}Q_{m_1}$ may become nearly singular when m_1 gets large. Moreover, when $m_1 > n$, $Q'_{m_1}Q_{m_1}$ is singular. To cover these cases, we consider a regularized version of the inverse of the matrix $Q'_{m_1}Q_{m_1}$.

We take ψ_j to be the eigenvectors of the $n \times n$ matrix $Q_{m_1}Q'_{m_1}/n$ associated with eigenvalues λ_j. For any vector e, the regularized version of P_{m_1}, $P^{\alpha}_{m_1}$ is:

$$P^{\alpha}_{m_1}e = \frac{1}{n}\sum_{j=1}^{n} q(\alpha, \lambda_j^2)\langle e, \psi_j\rangle \psi_j$$

where for T: $q(\alpha, \lambda_j^2) = \dfrac{\lambda_j^2}{\lambda_j^2 + \alpha}$,
for LF: $q(\alpha, \lambda_j^2) = [1 - (1 - c\lambda_j^2)^{1/\alpha}]$,
for SC: $q(\alpha, \lambda_j^2) = I(\lambda_j^2 \geq \alpha)$, for PC $q(\alpha, \lambda_j^2) = I(j \leq 1/\alpha)$.

The network models suggest the use of an infinite number of instruments. This can be done based on Carrasco and Florens (2000) works. Following their approach, we define the counterpart of P^{α} for infinite number of instruments by

$$P^{\alpha} = G(K_n^{\alpha})^{-1}G^*$$

where $G : l^2(\pi) \to \mathbb{R}^n$ with

$$Gg = \left(\langle Q_1, g\rangle', \langle Q_2, g\rangle', ..., \langle Q_n, g\rangle'\right)'$$

and $G^* : \mathbb{R}^n \to l^2(\pi)$ with

$$G^*v = \frac{1}{n}\sum_{i=1}^{n} Q_i v_i$$

such that $K_n = G^*G$ and GG^* is an $n \times n$ matrix with typical element $\dfrac{\langle Q_i, Q_j\rangle}{n}$. Let $\hat{\phi}_j$, $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ... > 0$, $j = 1, 2, ...$ be the orthonormalized eigenvectors and eigenvalues of K_n and ψ_j the eigenfunctions of GG^*.

$G\hat{\phi}_j = \sqrt{\lambda_j}\psi_j$ and $G^*\psi_j = \sqrt{\lambda_j}\hat{\phi}_j$. Note that in this case for $e \in \mathbf{R}^n$, $P^{\alpha}e = \sum_{j=1}^{\infty} q(\alpha, \lambda_j^2)\langle e, \psi_j\rangle\psi_j$. And

$$\begin{aligned} v'P^{\alpha}w &= v'G(K_n^{\alpha})^{-1}G^*w \\ &= \left\langle (K_n^{\alpha})^{-1/2}\sum_{i=1}^{n} Q_i(.)\, v_i, (K_n^{\alpha})^{-1/2}\frac{1}{n}\sum_{i=1}^{n} Q_i(.)\, w_i \right\rangle. \end{aligned} \tag{6}$$

All the regularization techniques presented in this section depend on a regularization parameter α, the choice of this parameter is very important for small sample behavior of the estimator. In Section 4 we discuss the selection of the regularization parameter. The following section presents the regularized 2SLS for network model.

3.3 Regularized 2SLS estimators

This section proposes the regularized 2SLS using three regularized methods (Tikhonov, Landweber Fridman and Principal Component). They are presented in a unified framework covering a finite number of instrument and an infinite number of instruments. The main focus is the estimation of endogenous and contextual effects under the assumption of a preliminary estimator of the unobservable correlated effects of individuals with their connections in the network. The asymptotic properties are also derived.

Assumption 3. $H = \lim_{n\to\infty} \frac{1}{n} f'f$ is a finite nonsingular matrix.

Assumption 4. (i) The elements of X are uniformly bounded constants, X has the full rank k, and $\lim_{n\to\infty} \frac{1}{n} X'X$ exists and is nonsingular.
(ii) there is a $\omega \geq 1/2$ such that

$$\sum_{j=1}^{\infty} \frac{\langle E(Z(.,x_i)f_a(x_i)), \phi_j \rangle^2}{\lambda_j^{2\omega+1}} < \infty$$

Under Assumption 2, the operator K is a Hilbert-Schmidt operator; we assume that it has nonzero eigenvalues. Assumption 4 (ii) ensures that the use of regularization enables us of to have a good asymptotic approximation of the best instrument f.

Let $\varepsilon(\rho_0, \delta) = JR(Y - Z\delta)$ with $\delta = (\lambda, \beta')'$ and $Z = (WY, X)$. The estimation is based on moments corresponding to the orthogonality condition of Q and $J\varepsilon$ given by[9]

$$E(Q'\varepsilon(\rho_0, \delta)) = 0$$

Our identification results are conditional on ρ_0. We should then first have a preliminary estimator $\tilde{\rho}$ of ρ. We take $\tilde{R} = I - \tilde{\rho}M$ to be an estimate of R.

We consider $S_n(k) = \frac{1}{n}\sum_{i=1}^{n}(\check{Y}_i - \check{Z}_i\delta)Q_{ik}$ with $\check{Y} = \tilde{R}Y$ and $\check{Z} = \tilde{R}Z$.

[9] The set of instrumental variables is $Q = J[Q_0, MQ_0]$ with $Q_0 = [WX, W^2X, ...W\iota, W^2\iota, ..., X]$.

We denote $(K_n^\alpha)^{-1}$ the regularized inverse of K_n and $(K_n^\alpha)^{-1/2} = ((K_n^\alpha)^{-1})^{1/2}$.

The regularized 2SLS estimator of δ is defined as:

$$\hat{\delta}_{R2sls} = argmin\big\langle (K_n^\alpha)^{-1/2} S_n(.), (K_n^\alpha)^{-1/2} S_n(.)\big\rangle \tag{7}$$

Solving the minimization problem gives,

$$\hat{\delta}_{R2sls} = (Z'\tilde{R}'P^\alpha \tilde{R} Z)^{-1} Z'\tilde{R}'P^\alpha \tilde{R} Y. \tag{8}$$

Equation (8) defines the regularized 2SLS. The regularized 2SLS for SAR is closely related to the regularized 2SLS of Carrasco (2012) and the 2SLS of Liu and Lee (2010). It extends Carrasco (2012) by considering SAR models and is of the same form as Liu and Lee's (2010) 2SLS with the difference that the projection matrix P is replaced by its regularized counterpart P^α.

3.4 Consistency and asymptotic distributions of the regularized 2SLS

The following proposition shows consistency and asymptotic normality of the regularized 2SLS estimators.

Proposition 1 *Under Assumptions 1-4 , $\tilde{\rho} - \rho_0 = O_p(1/\sqrt{n})$ and $\alpha \to 0$. Then, the T, LF and SC estimators satisfy:*

1. *Consistency: $\hat{\delta}_{R2sls} \to \delta_0$ in probability as n and $\alpha\sqrt{n}$ go to infinity.*
2. *Asymptotic normality: $\sqrt{n}(\hat{\delta}_{R2sls} - \delta_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\varepsilon^2 H^{-1}\right)$ as n and $\alpha^2\sqrt{n}$ go to infinity.*

The convergence rate of the regularized 2SLS for SAR is different from those obtained without spatial correlation. For consistency in the SAR model, we need $\alpha\sqrt{n}$ to go to infinity. The Carrasco (2012) regularized 2SLS estimator is consistent with convergence rate $n\alpha^{\frac{1}{2}}$. The asymptotic normality is obtained if $\alpha^2\sqrt{n}$ goes to infinity which is also different from Carrasco (2012)'s asymptotic normality condition for 2SLS. The regularization parameter α is allowed to go to zero less faster than for the Carrasco (2012) in for consistency. Compare to Carrasco (2012), more regularization is needed in order to achieve appropriate asymptotic behavior. The reinforcement of these conditions is certainly due to t regularization taking into account the spatial representation of the data.

In Liu and Lee (2010), the 2SLS estimator has a bias due to the increasing number of instrumental variables. Interestingly, the regularized 2SLS for SAR models is well centered, under the assumption that $\alpha\sqrt{n}$ go to infinity.

The bias of the 2SLS in Liu and Lee (2010) is of the form

$$\sqrt{n}b_{2sls} = \sigma^2 tr(P^\alpha RWS^{-1}R^{-1})(Z'RP^\alpha RZ)^{-1}e_1.$$

Using Lemma 1 and 2 in appendix, we show that the 2SLS bias is of order $\sqrt{n}b_{2sls} = O_p(\frac{1}{\alpha\sqrt{n}})$ which goes to zero as $\alpha\sqrt{n}$ goes to infinity.

4 Selection of the Regularization Parameter

This section discusses the selection of the regularized parameter for network models. We first derive an approximation of the mean square error using Nagar-type expansion. The dominant term of the mean square error is estimated and the selected regularization parameter is the one achieving the minimum of this term.

4.1 Approximation of the Mean Square Error (MSE)

The following proposition provides an approximation of the MSE.

Proposition 2 *If Assumptions 1 to 4 hold, $\tilde{\rho} - \rho_0 = O_p(1/\sqrt{n})$ and $n\alpha \to \infty$ for LF, PC and T regularized 2SLS estimators, we have*

$$n(\hat{\delta}_{R2sls} - \delta_0)(\hat{\delta}_{R2sls} - \delta_0)' = Q(\alpha) + \hat{R}(\alpha),$$

$$E(Q(\alpha)|X) = \sigma_\varepsilon^2 H^{-1} + S(\alpha), \tag{9}$$

$$r(\alpha)/tr(S(\alpha)) = o_p(1),$$

with $r(\alpha) = E(\hat{R}(\alpha)|X)$

$$S(\alpha) = \sigma_\varepsilon^2 H^{-1} \left[\frac{f'(1-P^\alpha)^2 f}{n} + \sigma_\varepsilon^2 \frac{1}{n} \left(\sum_j q_j \right)^2 e_1 \iota' D' D \iota e_1' \right] H^{-1}.$$

For LF, SC, $S(\alpha) = O_p\left(\frac{1}{n\alpha^2} + \alpha^\omega\right)$ and for T, $S(\alpha) = O_p\left(\frac{1}{n\alpha^2} + \alpha^{\min(\omega,2)}\right)$, with $D = JRWS^{-1}R^{-1}$.

The relevant, for the selection of α, dominant term $S(\alpha)$ will be minimized to achieve the smallest MSE. $S(\alpha)$ account for a trade-off between the bias and variance. When α goes to zero, the bias term increases while the variance term decreases. The approximation of the regularized 2SLS estimator is similar to Carrasco (2012) regularized 2SLS. However, the expression of the MSE is more complicated because of the spatial correlation.

4.2 Estimation of the MSE

The aim of this section is to find the regularized parameter that minimizes the conditional MSE of $\bar{\gamma}'\hat{\delta}_{2sls}$ for some arbitrary $k+1\times 1$ vector $\bar{\gamma}$. This conditional MSE is:

$$\begin{aligned} MSE &= E[\bar{\gamma}'(\hat{\delta}_{2sls}-\delta_0)(\hat{\delta}_{2sls}-\delta_0)'\bar{\gamma}|X] \\ &\sim \bar{\gamma}'S(\alpha)\bar{\gamma} \\ &\equiv S_{\bar{\gamma}}(\alpha). \end{aligned}$$

$S_{\bar{\gamma}}(\alpha)$ involves the function f which is unknown. We need to replace $S_{\bar{\gamma}}$ by an estimate.

Stacking the observations, the reduced form equation can be rewritten as

$$RZ = f + v.$$

This expression involves $n\times(k+1)$ matrices. We can reduce the dimension by post-multiplying by $H^{-1}\bar{\gamma}$:

$$RZH^{-1}\bar{\gamma} = fH^{-1}\bar{\gamma} + vH^{-1}\bar{\gamma} \Leftrightarrow RZ_{\bar{\gamma}} = f_{\bar{\gamma}} + v_{\bar{\gamma}} \tag{10}$$

where $v_{\bar{\gamma}i} = v_i'H^{-1}\bar{\gamma}$ is a scalar.

We take $\tilde{\delta}$ to be a preliminary estimator of δ obtained, for instance, from a finite number of instruments. And denote $\tilde{\rho}$ as a preliminary estimator of ρ obtained by the method of moments as follow:

$$\tilde{\rho} = armin\tilde{g}(\rho)'\tilde{g}(\rho)$$

where $\tilde{g}(\rho) = [M_1\tilde{\varepsilon}(\rho), M_2\tilde{\varepsilon}(\rho), M_3\tilde{\varepsilon}(\rho)]'\tilde{\varepsilon}(\rho)$,

$$M_1 = JWJ - tr(JWJ)I/tr(J),$$

$$M_2 = JMJ - tr(JMJ)I/tr(J),$$

$$M_3 = JMWJ - tr(JMWJ)I/tr(J),$$

$$\tilde{\varepsilon}(\rho) = JR(\rho)(Y - Z'\tilde{\delta}).$$

$\tilde{\delta} = [Z'Q_1(Q_1'Q_1)^{-1}Q_1'Z]^{-1}Z'Q_1(Q_1'Q_1)^{-1}Q_1'Y$ with Q_1 a single instrument. We obtain the residual $\hat{\varepsilon}(\rho) = JR(\tilde{\rho})(Y - Z'\tilde{\delta})$.

Let us denote $\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}(\rho)'\hat{\varepsilon}(\rho)/n$, $\hat{v}_{\bar{\gamma}} = (I - P^\alpha)R(\tilde{\rho})Z\tilde{H}^{-1}\bar{\gamma}$ where $\tilde{H}$ is a consistent estimate of H, $\tilde{v}_{\bar{\gamma}} = (I - P^{\tilde{\alpha}})R(\tilde{\rho})Z\tilde{H}^{-1}\bar{\gamma}$ and, $\hat{\sigma}_{v_{\bar{\gamma}}}^2 = \tilde{v}_{\bar{\gamma}}'\tilde{v}_{\bar{\gamma}}/n$

We consider the following goodness-of-fit criteria:

Mallows C_p (Mallows (1973))

$$\hat{\varpi}^m(\alpha) = \frac{\hat{v}_{\bar{\gamma}}'\hat{v}_{\bar{\gamma}}}{n} + 2\hat{\sigma}_{v_{\bar{\gamma}}}^2\frac{tr(P^\alpha)}{n}.$$

Generalized cross-validation (Craven and Wahba (1979))

$$\hat{\varpi}^{cv}(\alpha) = \frac{1}{n}\frac{\hat{v}_{\bar{\gamma}}'\hat{v}_{\bar{\gamma}}}{\left(1 - \frac{tr(P^\alpha)}{n}\right)^2}.$$

Leave-one-out cross-validation (Stone (1974))

$$\hat{\varpi}^{lcv}(\alpha) = \frac{1}{n}\sum_{i=1}^{n}(\tilde{RZ}_{\bar{\gamma}_i} - \hat{f}_{\bar{\gamma}_{-i}}^\alpha)^2,$$

where $\tilde{RZ}_{\bar{\gamma}} = W\tilde{H}^{-1}\bar{\gamma}$, $\tilde{RZ}_{\bar{\gamma}_i}$ is the i^{th} element of $\tilde{RZ}_{\bar{\gamma}}$ and $\hat{f}_{\bar{\gamma}_{-i}}^\alpha = P_{-i}^\alpha\tilde{RZ}_{\bar{\gamma}_{-i}}$. The $n \times (n-1)$ matrix P_{-i}^α is such that the $P_{-i}^\alpha = T(K_{n-i}^\alpha)T_{-i}^*$ are obtained by suppressing the i^{th} observation from the sample. $\tilde{RZ}_{\bar{\gamma}_{-i}}$ is the $(n-1)\times 1$ vector constructed by suppressing the i^{th} observation of $\tilde{W}_{\bar{\gamma}}$.

Using (9), $S_{\bar{\gamma}}(\alpha)$ can be rewritten as

$$S_{\bar{\gamma}}(\alpha) = \sigma_\varepsilon^2\left[\frac{f_{\bar{\gamma}}'(I - P^\alpha)^2 f_{\bar{\gamma}}}{n} + \sigma_\varepsilon^2\frac{1}{n}\left(\sum_j q_j\right)^2 e_{1\bar{\gamma}}\iota'D'D\iota e_{1\bar{\gamma}}'\right]$$

Using Li's results on C_p or cross-validation procedures, note that $\hat{\varpi}(\alpha)$ approximates

$$\varpi(\alpha) = \frac{f_{\bar{\gamma}}'(I - P^\alpha)^2 f_{\bar{\gamma}}}{n} + \sigma_{v\bar{\gamma}}^2\frac{tr\left((P^\alpha)^2\right)}{n}.$$

Therefore, $S_\gamma(\alpha)$ is estimated by the following equation.

$$\hat{S}_{\bar{\gamma}}(\alpha) = \hat{\sigma}_\varepsilon^2\left[\hat{\varpi}(\alpha) - \hat{\sigma}_{v_{\bar{\gamma}}}^2\frac{tr\left((P^\alpha)^2\right)}{n} + \hat{\sigma}_\varepsilon^2\frac{1}{n}\left(tr(P^\alpha)\right)^2 e_{1\bar{\gamma}}\iota'\tilde{D}'\tilde{D}\iota e_{1\bar{\gamma}}'\right]$$

where $\tilde{D}$ is a consistent estimator of D.

Our selection procedure is very close to Carrasco (2012), the optimality of the selection procedure can be established using the results of Li (1986) and Li (1987).

The regularized 2SLS and the selection of the regularization parameters are based on a preliminary estimator of ρ. This means that if we are not able to estimate correctly ρ the estimation of δ could be biased in an unpredictable direction.

The following section introduces the regularized GMM to jointly estimate ρ and δ.

4.3 Regularized GMM estimator

The regularized 2SLS can be generalized to the GMM with additional quadratic moment equations. The use of quadratic moments for the estimation of SAR models has been proposed in Kelejian and Prucha (1999) and Liu and Lee (2010). Identification of δ and ρ follows the same strategy as in Liu and Lee (2010). The use of quadratic moments helps in the identification of all the parameters of the model. But, the challenge with the use of GMM is that the derivation of the approximation of the MSE is very difficult. [10] However, we use the same regularization parameter obtained from the data-driven procedure for the regularized 2SLS.

The moments are $g_1(\theta) = Q'\varepsilon(\theta)$ with $\theta = (\rho, \delta)$ and

$\varepsilon(\theta) = JR(\rho)(Y - Z\delta) = f(\rho)(\delta_0 - \delta) + JR(\rho)WS^{-1}R^{-1}\varepsilon(\lambda_0 - \lambda) + JR(\rho)R^{-1}\varepsilon$ where $f(\rho) = JR(\rho)E(Z)$.

The additional quadratic moments are

$g_2(\theta) = [U_1\varepsilon(\theta), ..., U_q\varepsilon(\theta)]'\varepsilon(\theta)$, where U_j are constant square matrices such that $tr(JU_j) = 0$.

The number of quadratic moments is fixed (q). An example of U_j matrix is for any $n \times n$ constant matrix U, define M as $M = U - tr(JU)I/tr(J)$, then $tr(JM) = 0$. For notational simplicity, we take U_j to replace JU_jJ.

The vector of combined linear and quadratic empirical moments for the GMM estimation is given by $g(\theta) = [g_1(\theta)', g_2'(\theta)]'$. For analytic tractability, we impose uniform boundedness on the quadratic matrices U_j's.

Assumption 5. The sequence of matrices $\{U_j\}$ with $tr(JU_j) = 0$ are UB for $j = 1, ..., q$.

Assumption 3'. $\lim_{n\to\infty} \frac{1}{n} f(\rho)' f(\rho)$ is a finite nonsingular matrix for any ρ such that $R(\rho)$ is nonsingular.

[10] As a solution in a recent work, Kotchoni (2012) uses a bootstrap procedure to select the regularization parameter.

Under assumption 3', δ_0 is identified. Knowing δ_0, ρ_0 can be identified based on the quadratic moment conditions.

Assumption 6.
$lim_{n\rightarrow\infty}\frac{1}{n}tr(U_jMR^{-1}+U_j'MR^{-1})\neq 0$ for some j, and

$$lim_{n\rightarrow\infty}\frac{1}{n}[tr(U_1MR^{-1}+U_1'MR^{-1}),...,tr(U_qMR^{-1}+U_q'MR^{-1})]$$

is linearly independent of

$$lim_{n\rightarrow\infty}\frac{1}{n}[tr(R^{-1'}M'U_1MR^{-1}+R^{-1'}M'U_1'MR^{-1}),...,tr(R^{-1'}M'U_qMR^{-1}+R^{-1'}M'U_q'MR^{-1})].$$

Assumptions 5 and 6 are sufficient for the identification condition for ρ_0 via the unique solution of $E(g_2(\theta)|\delta=\delta_0)=0$ for a large enough sample size.

For any $n\times n$ matrix $A=[a_{ij}]$, we denote $vec_D(A)=(a_{11},...,a_{nn})'$. Let μ_3 and μ_4 denote, respectively, the third and fourth moments of the error term ε_i. Let also $\phi=[vec_D(U_1),...,vec_D(U_q)]$ and $\bar{g}_2(\theta)=\frac{\mu_3}{\sigma_0^2}\phi'P^{\alpha}\varepsilon(\theta)-g_2(\theta)$

The optimal GMM objective function can be treated as a linear combination of the objective functions of the 2SLS and the optimal GMM based on moments $\bar{g}_2(\theta)$ (see Liu and Lee (2010) for the proof). Following the same arguments, the regularized GMM with an infinite number of instruments is the sum of the objective functions of the regularized 2SLS and the optimal GMM based on moments $\bar{g}_2(\theta)$ which has a fixed dimension. The advantage of such a representation is that it uses the same regularization operators in both estimation procedures.

The regularized GMM estimator of θ is defined as:

$$\hat{\theta}_{rgmm}=argmin\ \sigma^{-2}\langle(K_n^{\alpha})^{-1/2}\bar{S}_n(.),(K_n^{\alpha})^{-1/2}\bar{S}_n(.)\rangle+\frac{1}{n}\bar{g}_2(\theta)'V\bar{g}_2(\theta) \tag{11}$$

with $V=[Var(\bar{g}_2(\theta))]^{-1}$ while $\bar{S}_n(k)=\frac{1}{n}\sum_{i=1}^{n}(\dot{Y}_i-\dot{Z}_i\delta)Q_{ik}$ where $\dot{Y}=R(\rho)Y$ and $\dot{Z}=R(\rho)Z$.

Lemma 5 in appendix shows that $V^{-1}=\frac{\mu_3^2}{\sigma^2}\phi'P^2\phi+(\mu_4-3\sigma^4)\phi'\phi+\sigma^4\Gamma-2\frac{\mu_3^2}{\sigma^2}\phi'P\phi$
with $\Gamma=\frac{1}{2}[vec(U_1+U_1'),...vec(U_q+U_q')]'[vec(U_1+U_1'),...vec(U_q+U_q')]$

Assumption 7. $lim_{n\rightarrow\infty}nV$ exists and is a nonsingular matrix.

Assumption 8. The true parameter θ_0 is in the interior of the compact parameter space Θ.

Proposition 3 *Under Assumptions 1-2, 3', 4-8 , with σ_ε^2, μ_3 and μ_4 replaced by their consistent initial estimators then the feasible optimal regularized GMM estimators for T, LF and PC satisfy $\hat{\theta}_{rgmm} \to \theta_0$ in probability as n and $\alpha\sqrt{n}$ go to infinity.*

Interestingly, the regularized GMM are consistent and converge with the same rate as the 2SLS. The following proposition result gives the asymptotic distribution of the feasible regularized GMM estimators.

Proposition 4 *Under Assumptions 1-2, 3', 4-8 , with σ_ε^2, μ_3 and μ_4 replaced by $\sqrt{n}$-consistent initial estimators, the feasible optimal regularized GMM estimator for T, LF and PC satisfied $\sqrt{n}(\hat{\theta}_{rgmm} - \theta_0) \stackrel{d}{\to} \mathcal{N}\left(0, [\sigma_\varepsilon^{-2} D(0,H) + plim \bar{D}_2' V \bar{D}_2]^{-1}\right)$ as n and $\alpha^2\sqrt{n}$ go to infinity; with*

$$\bar{D}_2 = D_2 - \frac{\mu_3}{\sigma_\varepsilon^2}(0, \phi' f)$$

where $D_2 = E(\frac{\partial g_2(\theta)}{\partial \theta'}) = -\sigma_\varepsilon^2 \begin{pmatrix} tr[(U_1+U_1')MR^{-1}] & tr[(U_1+U_1')RWS^{-1}R^{-1}] & 0 \\ tr[(U_2+U_2')MR^{-1}] & tr[(U_2+U_2')RWS^{-1}R^{-1}] & 0 \\ . & . & . \\ . & . & . \\ . & . & . \\ tr[(U_q+U_q')MR^{-1}] & tr[(U_q+U_q')RWS^{-1}R^{-1}] & 0 \end{pmatrix}.$

The regularized GMM is well centred, under the assumption that $\alpha^2\sqrt{n}$ go to infinity. This results were expected given that the regularized 2SLS estimators do not suffer from many instruments asymptotic bias. The regularized GMM proposed in this paper have the advantage that they can be computed with the same regularization parameter as in the regularized 2SLS estimators.

5 Monte Carlo simulations

To investigate the finite sample performance of the regularized 2SLS and GMM estimators, we conduct a simulation study based on the following model.

$$Y = \lambda_0 WY + X\beta_{10} + WX\beta_{20} + \iota\alpha_0 + u$$

with $u = \rho_0 Mu + \varepsilon$.

There are four samples with different numbers of groups $\bar{r}$ and group sizes m_r. The first sample contains 30 groups with equal group sizes of $m_r = 10$. The second sample contains 60 groups with equal group sizes of $m_r = 10$. To study the effect of group sizes, we also consider, respectively, 30 and 60 groups with equal group sizes of $m_r = 15$. For each group, the sociomatrix W_r is generated as follows. First, for the i^{th} row of W_r $(i = 1, ..., m_r)$, k_{ri} is generated uniformly at random from the set of integers [0, 1, 2, 3] , [0, 1, ..., 6] or [0, 1, ..., 8]. Allowing for a difference in the maximum number of friends helps to study the effect of the density of the network on the estimator.

The sociomatrix W_r is constructed as follows:

-Set the $(i+1)th, ..., (i+k_{ri})th$ elements of the i^{th} row of W_r to be ones and the rest of elements in that row to be zeros, if $i + k_{ri} \leq m_r$;

-Otherwise, the entries of ones will be wrapped around.

-In the case of $k_{ri} = 0$, the i^{th} row of W_r will have all zeros. M is the row-normalized W.

$X \sim \mathcal{N}(0, I)$, $\alpha_{0r} \sim \mathcal{N}(0, 0.01)$ $\varepsilon_{r,i} \sim \mathcal{N}(0, 1)$. The data are generated with $\beta_{10} = \beta_{20} = 0.2$ $\lambda_0 = \rho_0 = 0.1$.

The estimation methods considered are:

- 2SLS with few instruments, the set of instrument is given by $Q_1 = J[X, WX, MX, MWX]$,
- 2SLS with many instruments where the instrument set is $Q_2 = [Q_1, JW\iota]$.
- the regularized 2SLS estimator T 2SLS (Tikhonov), LF 2SLS (Landweber Fridman), PC 2SLS (Principal component) with many instruments $\tilde{Q}_2$. $\tilde{Q}_2$ is a matrix of instruments with Q_2's instruments normalized to unit variance.[11]
- Liu and Lee (2010) Bias-corrected 2SLS with many instruments
- Optimal GMM with $g(\theta) = [Q_1, U_1\varepsilon(\theta), U_2\varepsilon(\theta)]'\varepsilon(\theta)$ for few moments, with $U_1 = JMR^{-1}J - tr(JMR^{-1}J)I/tr(J)$ and
  $U_2 = JRWS^{-1}R^{-1}J - tr(JRWS^{-1}R^{-1}J)I/tr(J)$
- Optimal GMM with $g(\theta) = [Q_2, U_1\varepsilon(\theta), U_2\varepsilon(\theta)]'\varepsilon(\theta)$ for many moments, with $U_1 = JMR^{-1}J - tr(JMR^{-1}J)I/tr(J)$ and
  $U_2 = JRWS^{-1}R^{-1}J - tr(JRWS^{-1}R^{-1}J)I/tr(J)$
- Liu and Lee (2010) Bias-corrected GMM, Bias-corrected GMM with many instruments
- the regularized GMM estimators TGMM (Tikhonov), LFGMM (Landweber Fridman), PCGMM (Principal component) with many instruments $\tilde{Q}_2$.

[11] As pointed out by Newey (2013) the choice of identity as regularization matrix for the Tikhonov regularization method does not take into account difference in location and scale.

For all 2SLS estimator, a preliminary estimator of ρ is obtained by the method of moments, $\tilde{\rho} = armin\tilde{g}(\rho)'\tilde{g}(\rho)$ where $\tilde{g}(\rho) = [M_1\tilde{\varepsilon}(\rho), M_2\tilde{\varepsilon}(\rho), M_3\tilde{\varepsilon}(\rho)]'\tilde{\varepsilon}(\rho)$,

$$M_1 = JWJ - tr(JWJ)I/tr(J),$$

$$M_2 = JMJ - tr(JMJ)I/tr(J),$$

$$M_3 = JMWJ - tr(JMWJ)I/tr(J),$$

$$\tilde{\varepsilon}(\rho) = JR(\rho)(Y - Z'\tilde{\delta})$$

and

$$\tilde{\delta} = [Z'Q_1(Q_1'Q_1)^{-1}Q_1'Z]^{-1}Z'Q_1(Q_1'Q_1)^{-1}Q_1'Y.$$

Before presenting the simulations results, it important to note that the data generating process in this experiment exhibit a very low transitivity level. Moreover, the reduced form model is sparse (for example, when the maximum number of friend is 3, $W^q = 0$ for $q > 4$). The instruments coming from the relative position in the network are independent of each other. This means that high dimensional reduction techniques should not going very effective in summarizing the information.

Mean, standard deviation (SD) and root mean square errors (RMSE) of the empirical distributions of the estimates are reported.

The simulation results are summarized as follows.

1. The used on additional linear moment conditions reduce SDs in 2SLS estimators of λ_0 and β_{20} and GMM estimators of λ_0 , β_{20} and ρ_0. The 2SLS and GMM (large iv) have smaller standard deviation compare to the 2SLS and GMM (finite iv).

2. The additional instruments in Q_2 introduce biases into 2SLS estimators of λ_0 and β_{20} and GMM estimators of λ_0 , β_{20} and ρ_0. The 2SLS and GMM (finite iv) have a mean value of estimators closer to the true value of the parameter than 2SLS and GMM (large iv).

3. The regularized 2SLS and GMM procedures substantially reduce the many instruments' bias for both the 2SLS and GMM estimators, specifically in large samples. The bias-correction estimators are similar to regularized estimators in term of bias correction for large sample with a relatively more dense network. But, in a small sample, the bias of the bias-corrected estimator is smaller than those of the regularized estimators. Relative to the 2SLS with many instruments, the regularized 2SLS estimators reduce the many instruments bias and have comparable standard deviations.

Table 1: Simulation results for maximum number of connections 3 (1/2)

$m = 10$	$g = 30$			
	$\lambda_0 = 0.1$	$\beta_{10} = 0.2$	$\beta_{20} = 0.2$	$\rho_0 = 0.1$
2SLS (finite iv)	0.098(0.207)[0.207]	0.200(0.071)[0.071]	0.208(0.071)[0.071]	0.128(0.231)[0.233]
2SLS (large iv)	0.015(0.100)[0.131]	0.190(0.068)[0.068]	0.220(0.060)[0.063]	-
Bias-corrected 2SLS	0.106(0.131)[0.131]	0.198(0.069)[0.069]	0.205(0.063)[0.064]	-
T 2SLS	0.040(0.110)[0.125]	0.187(0.079)[0.080]	0.216(0.065)[0.066]	-
LF 2SLS	0.052(0.121)[0.130]	0.188(0.083)[0.084]	0.215(0.067)[0.068]	-
PC 2SLS	0.052(0.121)[0.130]	0.188(0.083)[0.084]	0.215(0.067)[0.068]	-
GMM (finite iv)	0.097(0.150)[0.150]	0.198(0.070)[0.070]	0.206(0.065)[0.065]	0.116(0.227)[0.227]
GMM (large iv)	0.075(0.078)[0.082]	0.195(0.069)[0.069]	0.208(0.059)[0.060]	0.051(0.139)[0.147]
Bias-corrected GMM	0.095(0.095)[0.096]	0.197(0.069)[0.069]	0.205(0.061)[0.061]	0.098(0.169)[0.169]
TGMM	0.085(0.097)[0.099]	0.198(0.080)[0.080]	0.208(0.066)[0.066]	0.090(0.166)[0.167]
LFGMM	0.088(0.136)[0.137]	0.204(0.110)[0.110]	0.207(0.074)[0.074]	0.121(0.232)[0.233]
PCGMM	0.096(0.183)[0.183]	0.243(0.555)[0.556]	0.208(0.128)[0.129]	0.115(0.251)[0.251]
	$g = 60$			
2SLS (finite iv)	0.104(0.136)[0.136]	0.203(0.047)[0.047]	0.204(0.049)[0.049]	0.116(0.177)[0.178]
2SLS (large iv)	0.032(0.081)[0.105]	0.196(0.046)[0.046]	0.217(0.043)[0.046]	-
Bias-corrected 2SLS	0.108(0.099)[0.099]	0.202(0.047)[0.047]	0.204(0.045)[0.045]	-
T 2SLS	0.055(0.088)[0.099]	0.193(0.051)[0.052]	0.213(0.046)[0.048]	-
LF 2SLS	0.064(0.095)[0.101]	0.193(0.053)[0.054]	0.212(0.048)[0.049]	-
PC 2SLS	0.064(0.095)[0.101]	0.193(0.053)[0.054]	0.212(0.048)[0.049]	-
GMM (finite iv)	0.092(0.106)[0.106]	0.201(0.047)[0.047]	0.205(0.045)[0.045]	0.119(0.184)[0.185]
GMM (large iv)	0.078(0.058)[0.062]	0.199(0.047)[0.047]	0.207(0.041)[0.042]	0.066(0.090)[0.096]
Bias-corrected GMM	0.092(0.071)[0.071]	0.201(0.047)[0.047]	0.204(0.043)[0.043]	0.101(0.109)[0.109]
TGMM	0.086(0.077)[0.078]	0.200(0.053)[0.053]	0.206(0.046)[0.047]	0.100(0.125)[0.125]
LFGMM	0.088(0.100)[0.101]	0.203(0.075)[0.075]	0.204(0.049)[0.050]	0.118(0.172)[0.173]
PCGMM	0.088(0.162)[0.163]	0.192(0.516)[0.516]	0.201(0.109)[0.109]	0.112(0.221)[0.222]

Mean (SD) [RMSE]

Table 2: Simulation results for maximum number of connections 3 (2/2)

$m = 15$	$g = 30$			
	$\lambda_0 = 0.1$	$\beta_{10} = 0.2$	$\beta_{20} = 0.2$	$\rho_0 = 0.1$
2SLS (finite iv)	0.098(0.155)[0.155]	0.203(0.052)[0.052]	0.202(0.055)[0.055]	0.115(0.204)[0.205]
2SLS (large iv)	0.069(0.094)[0.099]	0.200(0.052)[0.052]	0.209(0.044)[0.045]	-
Bias-corrected 2SLS	0.101(0.105)[0.105]	0.202(0.052)[0.052]	0.202(0.047)[0.047]	-
T 2SLS	0.086(0.106)[0.107]	0.199(0.059)[0.059]	0.207(0.049)[0.050]	-
LF 2SLS	0.089(0.114)[0.115]	0.200(0.062)[0.062]	0.207(0.053)[0.054]	-
PC 2SLS	0.089(0.114)[0.115]	0.200(0.062)[0.062]	0.207(0.053)[0.054]	-
GMM (finite iv)	0.092(0.122)[0.122]	0.202(0.051)[0.051]	0.203(0.050)[0.050]	0.113(0.171)[0.171]
GMM (large iv)	0.091(0.072)[0.073]	0.201(0.051)[0.051]	0.202(0.043)[0.043]	0.090(0.116)[0.116]
Bias-corrected GMM	0.096(0.078)[0.078]	0.201(0.051)[0.051]	0.201(0.044)[0.044]	0.100(0.123)[0.123]
TGMM	0.094(0.091)[0.092]	0.201(0.061)[0.061]	0.203(0.049)[0.049]	0.105(0.139)[0.139]
LFGMM	0.098(0.123)[0.123]	0.210(0.096)[0.097]	0.201(0.060)[0.060]	0.106(0.177)[0.177]
PCGMM	0.087(0.177)[0.177]	0.202(0.399)[0.399]	0.207(0.179)[0.179]	0.092(0.179)[0.179]
	$g = 60$			
2SLS (finite iv)	0.093(0.103)[0.103]	0.198(0.037)[0.037]	0.200(0.039)[0.039]	0.109(0.143)[0.143]
2SLS (large iv)	0.066(0.061)[0.070]	0.197(0.038)[0.038]	0.206(0.034)[0.035]	-
Bias-corrected 2SLS	0.096(0.072)[0.072]	0.198(0.038)[0.038]	0.199(0.036)[0.036]	-
T 2SLS	0.079(0.069)[0.072]	0.196(0.043)[0.043]	0.204(0.037)[0.037]	-
LF 2SLS	0.082(0.074)[0.076]	0.196(0.046)[0.046]	0.203(0.040)[0.040]	-
PC 2SLS	0.082(0.074)[0.076]	0.196(0.046)[0.046]	0.203(0.040)[0.040]	-
GMM (finite iv)	0.087(0.079)[0.080]	0.198(0.038)[0.038]	0.201(0.036)[0.036]	0.118(0.134)[0.135]
GMM (large iv)	0.090(0.050)[0.050]	0.198(0.038)[0.038]	0.199(0.033)[0.033]	0.085(0.085)[0.087]
Bias-corrected GMM	0.095(0.053)[0.053]	0.198(0.038)[0.038]	0.198(0.034)[0.034]	0.096(0.089)[0.089]
TGMM	0.090(0.062)[0.062]	0.197(0.044)[0.044]	0.200(0.036)[0.036]	0.105(0.102)[0.102]
LFGMM	0.091(0.086)[0.087]	0.200(0.067)[0.067]	0.199(0.042)[0.042]	0.111(0.140)[0.141]
PCGMM	0.078(0.153)[0.155]	0.185(0.331)[0.331]	0.211(0.157)[0.158]	0.111(0.172)[0.173]

Mean (SD) [RMSE]

Table 3: Simulation results for maximum number of connections 6 (1/2)

$m = 10$	$g = 30$			
	$\lambda_0 = 0.1$	$\beta_{10} = 0.2$	$\beta_{20} = 0.2$	$\rho_0 = 0.1$
2SLS (finite iv)	0.102(0.118)[0.118]	0.196(0.074)[0.074]	0.206(0.049)[0.049]	0.103(0.187)[0.188]
2SLS (large iv)	0.052(0.056)[0.074]	0.183(0.069)[0.071]	0.206(0.047)[0.047]	-
Bias-corrected 2SLS	0.109(0.078)[0.078]	0.196(0.072)[0.072]	0.202(0.048)[0.048]	-
T 2SLS	0.065(0.063)[0.073]	0.172(0.079)[0.083]	0.203(0.051)[0.051]	-
LF 2SLS	0.071(0.069)[0.075]	0.167(0.085)[0.091]	0.203(0.053)[0.053]	-
PC 2SLS	0.071(0.069)[0.075]	0.167(0.085)[0.091]	0.203(0.053)[0.053]	-
GMM (finite iv)	0.101(0.101)[0.101]	0.195(0.073)[0.073]	0.204(0.049)[0.049]	0.076(0.204)[0.205]
GMM (large iv)	0.078(0.050)[0.055]	0.189(0.071)[0.071]	0.204(0.047)[0.048]	-0.047(0.145)[0.207]
Bias-corrected GMM	0.090(0.057)[0.058]	0.192(0.071)[0.072]	0.204(0.048)[0.048]	0.073(0.180)[0.182]
TGMM	0.087(0.063)[0.064]	0.184(0.083)[0.085]	0.203(0.051)[0.051]	0.030(0.178)[0.191]
LFGMM	0.092(0.104)[0.104]	0.185(0.147)[0.147]	0.202(0.057)[0.057]	0.067(0.209)[0.212]
PCGMM	0.077(0.162)[0.164]	0.161(0.396)[0.398]	0.210(0.145)[0.146]	0.064(0.238)[0.241]
	$g = 60$			
2SLS (finite iv)	0.099(0.090)[0.090]	0.204(0.051)[0.051]	0.207(0.034)[0.035]	0.118(0.158)[0.159]
2SLS (large iv)	0.053(0.038)[0.060]	0.193(0.047)[0.048]	0.209(0.032)[0.034]	-
Bias-corrected 2SLS	0.099(0.064)[0.064]	0.203(0.049)[0.049]	0.205(0.034)[0.034]	-
T 2SLS	0.066(0.044)[0.056]	0.184(0.054)[0.056]	0.205(0.035)[0.035]	-
LF 2SLS	0.072(0.049)[0.057]	0.180(0.058)[0.061]	0.204(0.036)[0.036]	-
PC 2SLS	0.072(0.049)[0.057]	0.180(0.058)[0.061]	0.204(0.036)[0.036]	-
GMM (finite iv)	0.097(0.082)[0.082]	0.203(0.050)[0.050]	0.207(0.034)[0.035]	0.094(0.209)[0.209]
GMM (large iv)	0.077(0.035)[0.042]	0.199(0.048)[0.048]	0.207(0.033)[0.034]	-0.031(0.112)[0.172]
Bias-corrected GMM	0.089(0.042)[0.043]	0.201(0.049)[0.049]	0.206(0.033)[0.034]	0.073(0.141)[0.144]
TGMM	0.088(0.046)[0.048]	0.198(0.058)[0.058]	0.205(0.035)[0.036]	0.051(0.138)[0.146]
LFGMM	0.092(0.083)[0.084]	0.200(0.104)[0.104]	0.204(0.038)[0.038]	0.097(0.196)[0.196]
PCGMM	0.077(0.115)[0.118]	0.186(0.291)[0.291]	0.210(0.130)[0.130]	0.094(0.223)[0.223]

Mean (SD) [RMSE]

Table 4: Simulation results for maximum number of connections 6 (2/2)

$m = 15$	$g = 30$			
	$\lambda_0 = 0.1$	$\beta_{10} = 0.2$	$\beta_{20} = 0.2$	$\rho_0 = 0.1$
2SLS (finite iv)	0.104(0.068)[0.069]	0.205(0.053)[0.053]	0.203(0.038)[0.038]	0.116(0.144)[0.145]
2SLS (large iv)	0.081(0.043)[0.047]	0.199(0.052)[0.052]	0.207(0.035)[0.036]	-
Bias-corrected 2SLS	0.100(0.065)[0.065]	0.203(0.053)[0.053]	0.204(0.037)[0.037]	-
T 2SLS	0.092(0.048)[0.049]	0.198(0.060)[0.060]	0.207(0.039)[0.040]	-
LF 2SLS	0.095(0.053)[0.053]	0.198(0.062)[0.062]	0.207(0.041)[0.041]	-
PC 2SLS	0.095(0.053)[0.053]	0.198(0.062)[0.062]	0.207(0.041)[0.041]	-
GMM (finite iv)	0.098(0.060)[0.060]	0.203(0.052)[0.053]	0.204(0.037)[0.037]	0.117(0.189)[0.190]
GMM (large iv)	0.090(0.038)[0.039]	0.201(0.052)[0.052]	0.204(0.035)[0.035]	0.076(0.103)[0.105]
Bias-corrected GMM	0.098(0.042)[0.042]	0.202(0.052)[0.052]	0.203(0.035)[0.035]	0.090(0.109)[0.109]
TGMM	0.095(0.046)[0.046]	0.201(0.062)[0.062]	0.205(0.039)[0.040]	0.096(0.139)[0.140]
LFGMM	0.095(0.065)[0.065]	0.195(0.102)[0.103]	0.204(0.043)[0.044]	0.121(0.207)[0.208]
PCGMM	0.096(0.105)[0.105]	0.192(0.355)[0.356]	0.206(0.098)[0.099]	0.137(0.226)[0.229]
	$g = 60$			
2SLS (finite iv)	0.103(0.050)[0.050]	0.200(0.039)[0.039]	0.200(0.026)[0.026]	0.108(0.100)[0.100]
2SLS (large iv)	0.086(0.030)[0.033]	0.196(0.039)[0.039]	0.202(0.024)[0.025]	-
Bias-corrected 2SLS	0.105(0.035)[0.035]	0.200(0.039)[0.039]	0.199(0.025)[0.025]	-
T 2SLS	0.094(0.035)[0.035]	0.193(0.043)[0.043]	0.201(0.027)[0.027]	-
LF 2SLS	0.097(0.038)[0.039]	0.193(0.044)[0.044]	0.201(0.028)[0.028]	-
PC 2SLS	0.097(0.038)[0.039]	0.193(0.044)[0.044]	0.201(0.028)[0.028]	-
GMM (finite iv)	0.100(0.045)[0.045]	0.199(0.038)[0.038]	0.200(0.026)[0.026]	0.104(0.125)[0.126]
GMM (large iv)	0.094(0.027)[0.028]	0.198(0.038)[0.038]	0.200(0.024)[0.024]	0.077(0.070)[0.074]
Bias-corrected GMM	0.102(0.031)[0.031]	0.199(0.038)[0.038]	0.199(0.024)[0.024]	0.089(0.074)[0.075]
TGMM	0.098(0.035)[0.035]	0.196(0.045)[0.045]	0.200(0.027)[0.027]	0.093(0.083)[0.084]
LFGMM	0.099(0.053)[0.053]	0.194(0.080)[0.080]	0.199(0.030)[0.030]	0.107(0.129)[0.129]
PCGMM	0.102(0.095)[0.095]	0.203(0.352)[0.352]	0.198(0.079)[0.079]	0.121(0.170)[0.172]

Mean (SD) [RMSE]

Table 5: Simulation results for maximum number of connections 8 (1/2)

$m = 10$	$g = 30$			
	$\lambda_0 = 0.1$	$\beta_{10} = 0.2$	$\beta_{20} = 0.2$	$\rho_0 = 0.1$
2SLS (finite iv)	0.092(0.108)[0.108]	0.191(0.073)[0.074]	0.204(0.047)[0.047]	0.111(0.211)[0.211]
2SLS (large iv)	0.064(0.043)[0.056]	0.188(0.069)[0.070]	0.206(0.045)[0.046]	-
Bias-corrected 2SLS	0.099(0.062)[0.062]	0.194(0.071)[0.071]	0.203(0.048)[0.048]	-
T 2SLS	0.073(0.049)[0.056]	0.180(0.083)[0.086]	0.201(0.049)[0.049]	-
LF 2SLS	0.077(0.054)[0.058]	0.177(0.093)[0.096]	0.200(0.051)[0.051]	-
PC 2SLS	0.077(0.054)[0.058]	0.177(0.093)[0.096]	0.200(0.051)[0.051]	-
GMM (finite iv)	0.093(0.093)[0.094]	0.191(0.073)[0.073]	0.204(0.048)[0.048]	0.078(0.261)[0.262]
GMM (large iv)	0.080(0.040)[0.045]	0.190(0.073)[0.073]	0.204(0.047)[0.047]	-0.110(0.184)[0.279]
Bias-corrected GMM	0.088(0.044)[0.045]	0.191(0.073)[0.073]	0.203(0.047)[0.047]	0.064(0.225)[0.228]
TGMM	0.086(0.050)[0.052]	0.187(0.086)[0.087]	0.202(0.050)[0.050]	0.001(0.231)[0.252]
LFGMM	0.087(0.091)[0.092]	0.184(0.128)[0.129]	0.200(0.059)[0.059]	0.059(0.295)[0.297]
PCGMM	0.073(0.143)[0.146]	0.183(0.412)[0.412]	0.200(0.118)[0.118]	0.068(0.277)[0.278]
	g=60			
2SLS (finite iv)	0.096(0.065)[0.065]	0.202(0.048)[0.048]	0.204(0.033)[0.033]	0.113(0.162)[0.162]
2SLS (large iv)	0.071(0.028)[0.040]	0.198(0.047)[0.047]	0.207(0.032)[0.032]	-
Bias-corrected 2SLS	0.102(0.039)[0.039]	0.202(0.048)[0.048]	0.202(0.033)[0.033]	-
T 2SLS	0.080(0.032)[0.037]	0.194(0.056)[0.057]	0.203(0.034)[0.034]	-
LF 2SLS	0.084(0.036)[0.039]	0.192(0.063)[0.063]	0.201(0.035)[0.035]	-
PC 2SLS	0.084(0.036)[0.039]	0.192(0.063)[0.063]	0.201(0.035)[0.035]	-
GMM (finite iv)	0.095(0.062)[0.062]	0.201(0.049)[0.049]	0.204(0.033)[0.033]	0.096(0.181)[0.181]
GMM (large iv)	0.084(0.027)[0.031]	0.199(0.049)[0.049]	0.205(0.033)[0.033]	-0.062(0.119)[0.201]
Bias-corrected GMM	0.092(0.029)[0.030]	0.201(0.049)[0.049]	0.203(0.033)[0.033]	0.073(0.148)[0.151]
TGMM	0.091(0.032)[0.033]	0.199(0.058)[0.058]	0.203(0.035)[0.035]	0.040(0.151)[0.162]
LFGMM	0.096(0.065)[0.065]	0.200(0.088)[0.088]	0.202(0.039)[0.039]	0.073(0.192)[0.193]
PCGMM	0.081(0.109)[0.111]	0.190(0.469)[0.469]	0.204(0.118)[0.118]	0.080(0.207)[0.208]

Mean (SD) [RMSE]

Table 6: Simulation results for maximum number of connections 8 (2/2)

$m = 15$	$g = 30$			
	$\lambda_0 = 0.1$	$\beta_{10} = 0.2$	$\beta_{20} = 0.2$	$\rho_0 = 0.1$
2SLS (finite iv)	0.102(0.052)[0.052]	0.203(0.053)[0.053]	0.203(0.033)[0.033]	0.112(0.136)[0.137]
2SLS (large iv)	0.087(0.028)[0.031]	0.198(0.051)[0.051]	0.204(0.032)[0.032]	-
Bias-corrected 2SLS	0.101(0.036)[0.036]	0.202(0.052)[0.052]	0.202(0.032)[0.032]	-
T 2SLS	0.093(0.032)[0.033]	0.195(0.058)[0.058]	0.204(0.034)[0.034]	-
LF 2SLS	0.095(0.036)[0.036]	0.193(0.061)[0.061]	0.204(0.035)[0.035]	-
PC 2SLS	0.095(0.036)[0.036]	0.193(0.061)[0.061]	0.204(0.035)[0.035]	-
GMM (finite iv)	0.100(0.048)[0.048]	0.202(0.053)[0.053]	0.203(0.033)[0.033]	0.096(0.164)[0.164]
GMM (large iv)	0.091(0.027)[0.028]	0.199(0.051)[0.051]	0.203(0.032)[0.032]	0.057(0.099)[0.108]
Bias-corrected GMM	0.097(0.030)[0.030]	0.201(0.051)[0.051]	0.202(0.032)[0.032]	0.084(0.108)[0.109]
TGMM	0.096(0.033)[0.033]	0.198(0.060)[0.060]	0.203(0.034)[0.034]	0.076(0.115)[0.117]
LFGMM	0.100(0.054)[0.054]	0.196(0.114)[0.114]	0.202(0.036)[0.036]	0.087(0.153)[0.153]
PCGMM	0.099(0.094)[0.094]	0.195(0.261)[0.261]	0.209(0.108)[0.108]	0.110(0.179)[0.179]
	$g = 60$			
2SLS (finite iv)	0.103(0.038)[0.038]	0.199(0.039)[0.039]	0.199(0.022)[0.022]	0.105(0.097)[0.097]
2SLS (large iv)	0.091(0.020)[0.022]	0.195(0.038)[0.039]	0.200(0.021)[0.021]	-
Bias-corrected 2SLS	0.104(0.026)[0.026]	0.199(0.039)[0.039]	0.198(0.021)[0.021]	-
T 2SLS	0.096(0.023)[0.023]	0.191(0.043)[0.043]	0.200(0.022)[0.022]	-
LF 2SLS	0.098(0.026)[0.026]	0.190(0.044)[0.045]	0.200(0.023)[0.023]	-
PC 2SLS	0.098(0.026)[0.026]	0.190(0.044)[0.045]	0.200(0.023)[0.023]	-
GMM (finite iv)	0.101(0.035)[0.035]	0.199(0.039)[0.039]	0.199(0.021)[0.021]	0.099(0.158)[0.158]
GMM (large iv)	0.095(0.019)[0.020]	0.197(0.038)[0.039]	0.200(0.021)[0.021]	0.062(0.072)[0.081]
Bias-corrected GMM	0.101(0.021)[0.021]	0.199(0.039)[0.039]	0.199(0.021)[0.021]	0.083(0.078)[0.080]
TGMM	0.099(0.024)[0.024]	0.195(0.044)[0.045]	0.199(0.022)[0.022]	0.083(0.088)[0.090]
LFGMM	0.101(0.043)[0.043]	0.192(0.085)[0.086]	0.198(0.025)[0.025]	0.101(0.125)[0.125]
PCGMM	0.099(0.080)[0.080]	0.178(0.224)[0.225]	0.197(0.078)[0.078]	0.114(0.157)[0.158]

Mean (SD) [RMSE]

4. The regularized GMM estimators reduce the bias in the estimation of ρ_0 relative to bias corrected GMM and GMM with a large number of instruments. However, the precision of the regularized GMM estimators of ρ is not as good as the bias correction.[12]

5. The performance of the regularized estimators increases with the higher density of the network and the larger number of groups. The behavior of the regularized estimator with respect to the network density suggests that the regularized estimators are good candidates to improve the asymptotic behavior of the estimator of network effect when the level of transitivity in the groups is very high.

6 Conclusion

This paper uses the regularization methods for the estimation of network models. The regularization is proposed as a solution to the weak identification problem in network models. Identification of the network effect can be achieved by using individuals' Bonacich (1987) centrality as an instrumental variables. But, the number such instruments increased with the number of groups; leading to the many instruments problem. Identification can also be achieved using the friend of a friend exogenous characteristics. However, if the network is very dense or the group size is very large, the identification is weaken. The proposed regularized 2SLS and GMM based on three regularization methods help to deal with many moments and weak identification problems. These estimators are consistent and asymptotically normal. The regularized 2SLS estimators achieve the asymptotic efficiency bound. We derive an optimal data-driven selection method for the regularization parameter. A Monte Carlo experiment shows that the regularized estimator performed well. The regularized 2SLS and GMM procedures substantially reduce the many instruments bias for both the 2SLS and GMM estimators, specifically in a large sample. Moreover, the qualities in term of bias and precision of the regularized estimator improves with the increase of the network density and the number of groups. These results show that regularization is a valuable solution to the potential weak identification problem existing in network models estimation.

[12] Note that, with some moment selection method applied to the same type of models, the problem of precision of the estimator is observed, for example, in Liu and Lee (2013) the decile range of the C2LS-op and 2SLS-op seems to be the largest.

References

Bai, J., and S. Ng (2010): "Instrumental Variable Estimation in a Data Rich Environment," *Econometric Theory*, 26, 1577–1606.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62(3), 657–81.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80(6), 2369–2429.

Bonacich, P. (1987): "Power and centrality: A family of measures," *American journal of sociology*, pp. 1170–1182.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of econometrics*, 150(1), 41–55.

Carrasco, M. (2012): "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170(2), 383–398.

Carrasco, M., and J.-P. Florens (2000): "Generalization Of Gmm To A Continuum Of Moment Conditions," *Econometric Theory*, 16(06), 797–834.

——— (2014): "On the asymptotic efficiency of GMM," *Econometric Theory*, 30(02), 372–406.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, vol. 6 of *Handbook of Econometrics*, chap. 77. Elsevier.

Carrasco, M., and G. Tchuente (2015): "Regularized LIML for many instruments," *Journal of Econometrics*, 186(2), 427–442.

——— (2016): "Efficient estimation with many weak instruments using regularization techniques," *Econometric Reviews*, pp. 1–29.

Chao, J. C., and N. R. Swanson (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73(5), 1673–1692.

Craven, P., and G. Wahba (1979): "Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of the generalized cross-validation," *Numer. Math.*, 31, 377–403.

Davidson, R., and J. G. MacKinnon (1993): *Estimation and Inference in Econometrics*, no. 9780195060119 in OUP Catalogue. Oxford University Press.

Donald, S. G., and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69(5), 1161–91.

Gibbons, S., and H. G. Overman (2012): "Mostly pointless spatial econometrics?*," *Journal of Regional Science*, 52(2), 172–191.

Han, C., and P. C. B. Phillips (2006): "GMM with Many Moment Conditions," *Econometrica*, 74(1), 147–192.

Hansen, C., J. Hausman, and W. Newey (2008): "Estimation With Many Instrumental Variables," *Journal of Business & Economic Statistics*, 26, 398–422.

Hansen, C., and D. Kozbur (2014): "Instrumental variables estimation with many weak instruments using regularized JIVE," *Journal of Econometrics*, 182(2), 290–308.

Hasselt, M. v. (2010): "Many instruments asymptotic approximations under nonnormal error distributions," *Econometric Theory*, 26(02), 633–645.

Kapetanios, G., and M. Marcellino (2010): "Factor-GMM estimation with large sets of possibly weak instruments," *Computational Statistics and Data Analysis*, 54, 2655–2675.

Kelejian, H., and I. R. Prucha (2001): "On the asymptotic distribution of the Moran< i> I</i> test statistic with applications," *Journal of Econometrics*, 104(2), 219–257.

Kelejian, H. H., and I. R. Prucha (1999): "A generalized moments estimator for the autoregressive parameter in a spatial model," *International economic review*, 40(2), 509–533.

Kotchoni, R. (2012): "Applications of the characteristic function-based continuum GMM in finance," *Computational Statistics & Data Analysis*, 56(11), 3599–3622.

Kress, R. (1999): *Linear Integral Equations.* Springer.

Kuersteiner, G. (2012): "Kernel-weighted GMM estimators for linear time series models," *Journal of Econometrics*, 170, 399–421.

Lancaster, T. (2000): "The incidental parameter problem since 1948," *Journal of econometrics*, 95(2), 391–413.

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(2), 333–374.

Li, K.-C. (1986): "Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing," *The Annals of Statistics*, 14, 1101–1112.

——— (1987): "Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete Index Set," *The Annals of Statistics*, 15, 958–975.

Liu, X., and L.-f. Lee (2010): "GMM estimation of social interaction models with centrality," *Journal of Econometrics*, 159(1), 99–115.

Liu, X., and L.-F. Lee (2013): "Two-stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments," *Econometric Reviews*, 32(5-6), 734–753.

Mallows, C. L. (1973): "Some Comments on Cp," *Technometrics*, 15, 661–675.

Manski, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 60(3), 531–542.

Newey, W. K. (2013): "Nonparametric instrumental variables estimation," *The American Economic Review*, 103(3), 550–556.

Neyman, J., and E. L. Scott (1948): "Consistent estimates based on partially consistent observations," *Econometrica: Journal of the Econometric Society*, pp. 1–32.

Okui, R. (2011): "Instrumental variable estimation in the presence of many moment conditions," *Journal of Econometrics*, 165, 70–86.

Stone, C. J. (1974): "Cross-validatory choice and assessment of statistical predictions," *Journal of the Royal Statistical Society*, 36, 111–147.

A Appendix: Summary of notations

To avoid heavy notations let us use

$P = P^{\alpha}$, $q_j = q(\lambda_j^2, \alpha)$

$tr(A)$ is the trace of matrix A

e_j is the j^{th} unit (column) vector

$e_f = \dfrac{1}{n} f'(I-P)f,$

$e_{2f} = \dfrac{1}{n} f'(I-P)^2 f,$

$\Delta_f = tr(e_f)$ and $\Delta_{2f} = tr(e_{2f})$

$\Gamma = \dfrac{1}{2}[vec(U_1 + U_1'), ...vec(U_q + U_q')]'[vec(U_1 + U_1'), ...vec(U_q + U_q')]$

$\phi = [vec_D(U_1), ..., vec_D(U_q)]$

$\bar{g}_2(\theta) = \dfrac{\mu_3}{\sigma_0^2} \phi' P^{\alpha} \varepsilon(\theta) - g_2(\theta)$

$g_2(\theta) = [U_1 \varepsilon(\theta), ..., U_q \varepsilon(\theta)]' \varepsilon(\theta)$, where U_j are constant square matrices such that $tr(JU_j) = 0$.

B Appendix: Some Lemmas

Lemma 1:

(i) $tr(P) = \sum_j q_j = O(1/\alpha)$ and $tr(P^2) = \sum_j q_j^2 = o((\sum_j q_j)^2)$.

(ii) Suppose that $\{A\}$ is a sequence of $n \times n$ UB matrices. For $B = PA$, $tr(B) = o((\sum_j q_j)^2)$, $tr(B^2) = o((\sum_j q_j)^2)$, and $\sum_i B_{ii}^2 = o((\sum_j q_j)^2)$, where B_{ii}'s are diagonal elements of B.

Proof of Lemma 1

(i) proof is in Carrasco (2012) Lemma 4 (i).

(ii) By eigenvalue decomposition, $AA' = \Pi \Delta \Pi'$, where Π is an orthonormal matrix and Δ is the eigenvalue matrix. It follows that $PAA'P \leq \lambda_{max} P^2$ with λ_{max} the largest eigenvalue. It follows that $tr(PAA'P) \leq \lambda_{max} tr(P^2) = o_p((\sum_j q_j)^2)$. By Cauchy-Schwartz inequality $tr(B) \leq [tr(P^2)]^{1/2} [tr(PAA'P)]^{1/2} = o_p((\sum_j q_j)^2)$. Also by Cauchy-Schwartz inequality $tr(B) \leq tr(BB') = tr(PAA'P) = o((\sum_j q_j)^2)$

Lemma 2: Let C and D be two UB $n \times n$ matrix sequences.

(i) $C'PD = O(n/\alpha)$

(ii) $\varepsilon' C' P D \varepsilon = O_p(1/\alpha^2)$ and $C' P D \varepsilon = O_p(\sqrt{n}/\alpha)$

Proof of Lemma 2

(i) By Cauchy-Schwartz inequality $|e_i' C' P^{\alpha} D e_j \leq \sqrt{e_i' C' C e_i}\sqrt{e_j' D' P^2 D e_j} = O(n/\alpha)$ which implies $C' P D = O(n/\alpha)$

(ii) $E|\varepsilon' C' P D \varepsilon| \leq \sqrt{E(\varepsilon' C' P^2 C \varepsilon)}\sqrt{E(\varepsilon' D' P^2 D \varepsilon)} = \sigma^2 \sqrt{tr(C' P^2 C)}\sqrt{tr(D' P^2 D)} = O(\frac{1}{\alpha^2})$ by Markov inequality $\varepsilon' C' P D \varepsilon = O_p(\frac{1}{\alpha^2})$

By Cauchy-Schwartz inequality $|e_j' C' P D \varepsilon| \leq \sqrt{e_j' C' C e_j}\sqrt{\varepsilon' D' P^2 D \varepsilon} = O_p(\sqrt{n}/\alpha)$,thus $C' P D \varepsilon = O_p(\sqrt{n}/\alpha)$

Lemma 3: Suppose $\tilde{\rho}$ is a consistent estimator of ρ_0 and $\tilde{R} = R(\tilde{\rho})$.
Then, $\frac{1}{n} Z' \tilde{R}' P \tilde{R} Z = \frac{1}{n} Z' R' P R Z + O_p[(\tilde{\rho} - \rho_0)/\alpha]$ and
$\frac{1}{n} Z' \tilde{R}' P \tilde{R} R^{-1} \varepsilon = \frac{1}{n} Z' R P \varepsilon + O_p[(\tilde{\rho} - \rho_0)/(\alpha\sqrt{n}(1 + \alpha\sqrt{n}))]$.

Proof of Lemma 3

$\tilde{R} = R - (\tilde{\rho} - \rho_0) M$, thus

$$
\begin{aligned}
Z' \tilde{R}' P \tilde{R} Z/n &= Z' R' P R Z/n \\
&- (\tilde{\rho} - \rho_0) Z' M' P R Z/n - (\tilde{\rho} - \rho_0) Z' R' P M Z/n \\
&+ (\tilde{\rho} - \rho_0)^2 Z' M' P M Z/n
\end{aligned}
$$

Let show that $Z' R' P M Z/n = O_p(1/\alpha)$ and $Z' M' P M Z/n = O_p(1/\alpha)$
Note that $Z = [W S^{-1}(X\beta_0 + \iota\gamma_0), X] + W S^{-1} R^{-1} \varepsilon e_1'$.

Under assumption 3, $Z' R' P M Z/n = O(1/\alpha) + O_p(1/\sqrt{n}\alpha) + O_p(1/n\alpha^2) = O_p(1/\alpha)$ and $Z' M' P M Z/n = O_p(1/\alpha)$ by lemma 2 (i).

$$
\begin{aligned}
Z' \tilde{R}' P \tilde{R} \varepsilon/n &= Z' R P \varepsilon/n \\
&- (\tilde{\rho} - \rho_0) Z' M' P \varepsilon/n - (\tilde{\rho} - \rho_0) Z' R' P M R^{-1} \varepsilon/n \\
&+ (\tilde{\rho} - \rho_0)^2 Z' M' P M R^{-1} \varepsilon/n
\end{aligned}
$$

I used the same argument as in the previous case under assumption 3,

$Z' R' P M R^{-1} \varepsilon/n = O_p(1/\sqrt{n}\alpha + 1/n\alpha^2) = O_p[1/\alpha\sqrt{n}(1 + 1/\alpha\sqrt{n})]$, $Z' M' P \varepsilon/n = O_p[1/\alpha\sqrt{n}(1 + 1/\alpha\sqrt{n})]$ and $Z' M' P \varepsilon/n = O_p[1/\alpha\sqrt{n}(1 + 1/\alpha\sqrt{n})]$ by lemma 2 (ii).

Lemma 4: If Assumptions 1-4, are satisfied and $\alpha \to 0$ then:

(i)$Z'RPRZ/n = H + o_p(1)$ if $\alpha\sqrt{n} \to \infty$

(ii)$Z'RP\varepsilon/\sqrt{n} = f'\varepsilon/\sqrt{n} + o_p(1)$ if $\alpha^2\sqrt{n} \to \infty$

Proof of Lemma 4

Let $v = JRWS^{-1}R^{-1}\varepsilon$, $JRZ = f + ve_1'$

(i) $\frac{1}{n}Z'RPRZ = \frac{1}{n}f'f - \frac{1}{n}f'(I-P)f + \frac{1}{n}e_1v'Pve_1' + \frac{1}{n}f'Pve_1' + \frac{1}{n}e_1v'Pf$

Let $e_f = \frac{1}{n}f'(I-P)f$, $e_{2f} = \frac{1}{n}f'(I-P)^2f$, $\Delta_f = tr(e_f)$ and $\Delta_{2f} = tr(e_{2f})$ By Cauchy-Schwartz inequality $\frac{1}{n}|e_i'f'(I-P)fe_j| \leq \frac{1}{n}\sqrt{e_i'f'fe_i}\sqrt{e_j'f'(I-P)^2fe_j} = O(\sqrt{\Delta_{2f}})$

From Carrasco (2012) Lemma 5 (i) $\Delta_{2f} = \begin{cases} O_p(\alpha^{\omega}) \; for \; LF \; and \; SC \\ O_p(\alpha^{min(\omega,2)}) \; for \; T \end{cases}$ Thus, $\Delta_{2f} = o_p(1)$

By Lemma 2 (ii) $\frac{1}{n}e_1v'Pve_1' + \frac{1}{n}f'Pve_1' + \frac{1}{n}e_1v'Pf = O_p(\frac{1}{n\alpha^2} + \frac{1}{\alpha\sqrt{n}}) = o_p(1)$

(ii) $Z'RP\varepsilon/\sqrt{n} = f'\varepsilon/\sqrt{n} - f'(I-P)\varepsilon/\sqrt{n} + e_1v'P\varepsilon/\sqrt{n}$

By Lemma 5 (ii) of Carrasco (2012) $f'(I-P)\varepsilon/\sqrt{n} = O_p(\sqrt{\Delta_{2f}})$ and by Lemma 2 (ii) $e_1v'P\varepsilon/\sqrt{n} = O_p(1/\alpha^2\sqrt{n})$

Lemma 5: If Assumptions 1-4, are satisfied.

(i) $V^{-1} = \frac{\mu_3^2}{\sigma^2}\phi'P^2\phi + (\mu_4 - 3\sigma^4)\phi'\phi + \sigma^4\Gamma - 2\frac{\mu_3^2}{\sigma^2}\phi'P\phi$

with $\Gamma = \frac{1}{2}[vec(U_1 + U_1'), ...vec(U_q + U_q')]'[vec(U_1 + U_1'), ...vec(U_q + U_q')]$

(ii) $E(\bar{g}_2(\theta)g_1(\theta)') = \mu_3\phi'PQ - \mu_3\phi'Q = o(1)$

Proof of Lemma 5

Note that $V = [Var(\bar{g}_2(\theta))]^{-1}$

$$\begin{aligned} V^{-1} = Var(\bar{g}_2(\theta)) &= E[\frac{\mu_3^2}{\sigma^4}\phi'P\varepsilon(\theta)\varepsilon(\theta)'P\phi] - 2E[\frac{\mu_3}{\sigma^2}\phi'P\varepsilon(\theta)g_2(\theta)'] \\ &+ E[g_2(\theta)g_2(\theta)'] \end{aligned}$$

It can be shown that

$$\begin{aligned} E[\frac{\mu_3^2}{\sigma^4}\phi'P\varepsilon(\theta)\varepsilon(\theta)'P\phi] &= E[\frac{\mu_3^2}{\sigma^4}\phi'PJ\varepsilon\varepsilon'JP\phi] \\ &= \frac{\mu_3}{\sigma^2}\phi'P^2\phi \end{aligned}$$

Then the second term is

$$
\begin{aligned}
E[\frac{\mu_3}{\sigma^2}\phi' P\varepsilon(\theta)g_2(\theta)] &= \frac{\mu_3}{\sigma^2}E[\phi' PJ\varepsilon\varepsilon' J[U_1J\varepsilon,...,U_qJ\varepsilon]] \\
&= \frac{\mu_3}{\sigma^2}E[\phi' P\varepsilon[\varepsilon' JU_1J\varepsilon,...,\varepsilon' JU_qJ\varepsilon]] \\
&= \frac{\mu_3}{\sigma^2}E[\phi' P\varepsilon[tr(\varepsilon' JU_1J\varepsilon),...,tr(\varepsilon' JU_qJ\varepsilon)]] \\
&= \frac{\mu_3^2}{\sigma^2}\sum_{ij}\phi_i' P_{ij}[tr(e_j' JU_1Je_j),...,tr(e_j' JU_qJe_j)] \\
&= \frac{\mu_3^2}{\sigma^2}\phi' P\phi
\end{aligned}
$$

With U_jJ replace by JU_jJ for all j. And the last term

$$
\begin{aligned}
E[g_2(\theta)g_2(\theta)'] &= \mu_4\phi'\phi + \sigma^4[\gamma - 3\phi'\phi] \\
&= (\mu_4 - 3\sigma^4)\phi'\phi + \sigma^4\Gamma
\end{aligned}
$$

with $\Gamma = \frac{1}{2}[vec(U_1+U_1'),...vec(U_q+U_q')]'[vec(U_1+U_1'),...vec(U_q+U_q')]$

C Appendix: Proofs of propositions

Proof of Proposition 1

The regularized 2SLS estimator satisfies $\hat{\delta}_{R2sls} - \delta_0 = (Z'\tilde{R}'P\tilde{R}Z)^{-1}Z'\tilde{R}'P\tilde{R}R^{-1}\varepsilon$.
$Z'\tilde{R}'P\tilde{R}Z/n = O_p(1) + O_n(1/\alpha\sqrt{n})$ by Lemma 3 and 4.
$\tilde{R}'P\tilde{R}R^{-1}\varepsilon/n = O_p(1/\sqrt{n}) + O_p[(1/(n\alpha(1+\alpha\sqrt{n}))$ by lemma 3 and 4.

Then $\hat{\delta}_{R2sls} - \delta_0 = o_p(1)$ as $\alpha\sqrt{n} \to \infty$ and $\alpha \to 0$ This prove the consistency of the regularized 2SLS for SAR with many instruments.

$$
\sqrt{n}(\hat{\delta}_{R2sls} - \delta_0) = (Z'\tilde{R}'P\tilde{R}Z/n)^{-1}[Z'\tilde{R}'P\tilde{R}R^{-1}\varepsilon/\sqrt{n}].
$$

Using Lemma 3 , 4 and Slutzky theorem

$$
\sqrt{n}(\hat{\delta}_{R2sls} - \delta_0) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\varepsilon^2 H^{-1}\right)
$$

if $\alpha^2\sqrt{n} \to \infty$ and $\alpha \to 0$

Proof of Proposition 2

Let us considers the MSE of the estimation of the parameters.

$$n(\hat{\delta}_{R2sls} - \delta_0)(\hat{\delta}_{R2sls} - \delta_0) = \hat{H}^{-1}\hat{h}\hat{h}'\hat{H}^{-1}$$

with $\hat{H} = \dfrac{Z'\tilde{R}'P\tilde{R}Z}{n}$ and $\hat{h} = \dfrac{Z'\tilde{R}'P\tilde{R}Y}{\sqrt{n}}$. Our objective is to approximate the MSE. To achieve this aim, we are going to use a Nagar-type approximation in order to be able to concentrate on the largest part of the MSE. By Lemma 3,

$$\begin{aligned} \hat{H} &= Z'RPRZ/n \\ &- (\tilde{\rho} - \rho_0)Z'M'PRZ/n - (\tilde{\rho} - \rho_0)Z'R'PMZ/n \\ &+ (\tilde{\rho} - \rho_0)^2 Z'M'PMZ/n \end{aligned}$$

And $\hat{H} = Z'RPRZ/n + O_p((\tilde{\rho} - \rho_0)/\alpha)$. By Lemma 4, we have that

$$\hat{H} = \frac{1}{n}f'f - \frac{1}{n}f'(I-P)f + \frac{1}{n}e_1 v'Pve_1' + \frac{1}{n}f'Pve_1' + \frac{1}{n}e_1 v'Pf + O_p((\tilde{\rho} - \rho_0)/\alpha)$$

Let us define $T^H = T_1^H + T_2^H + T_3^H$ with
$T_1^H = -\frac{1}{n}f'(I-P)f$, $T_2^H = \frac{1}{n}e_1 v'Pve_1'$ and $T_3^H = \frac{1}{n}f'Pve_1' + \frac{1}{n}e_1 v'Pf + O_p((\tilde{\rho} - \rho_0)/\alpha)$. Such that

$$\begin{aligned} \hat{H} &= \frac{1}{n}f'f + T_1^H + T_2^H + T_3^H \\ &= H + T_1^H + T_2^H + T_3^H + o_p(1) \\ &= H + T^H + o_p(1). \end{aligned}$$

Following similar arguments, we have

$$\hat{h} = f'\varepsilon/\sqrt{n} - f'(I-P)\varepsilon/\sqrt{n} + e_1 v'P\varepsilon/\sqrt{n} + O_p[(\tilde{\rho} - \rho_0)/(\alpha(1 + \alpha\sqrt{n}))]$$

Let us also define, $T^h = T_1^h + T_2^h$ with
$T_1^h = -f'(I-P)\varepsilon/\sqrt{n}$ and $T_2^h = e_1 v'P\varepsilon/\sqrt{n} + O_p[(\tilde{\rho} - \rho_0)/(\alpha(1 + \alpha\sqrt{n}))]$.

We therefore have

$$\begin{aligned} \hat{h} &= f'\varepsilon/\sqrt{n} + T_1^h + T_2^h \\ &= h + T_1^h + T_2^h + o_p(1) \\ &= h + T^h + o_p(1). \end{aligned}$$

Using Nagar-type enpension on $\hat{H}^{-1}$

$$n(\hat{\delta}_{R2sls}-\delta_0)(\hat{\delta}_{R2sls}-\delta_0) = H^{-1}[I-T^H H^{-1}][hh'+hT^h+T^h h'+T^h T^{h'}][I-H^{-1}T^H]H^{-1}+o_p(1)$$

Let us define $A(\alpha)=[I-T^H H^{-1}]\Im(\alpha)[I-H^{-1}T^H]$ with $\Im(\alpha)=[hh'+hT^h+T^h h'+T^h T^{h'}]$.

Therefore, $A(\alpha)=\Im(\alpha)+T^H H^{-1}\Im(\alpha)H^{-1}T^H - T^H H^{-1}\Im(\alpha)-\Im(\alpha)H^{-1}T^H$

$$\begin{aligned} E[\Im(\alpha)|X] &= \sigma^2[H-2e_f+\frac{1}{n}f'Pve_1'+\frac{1}{n}e_1v'Pf+e_{2f}] \\ &- E[\frac{1}{n}f'(I-P)\varepsilon\varepsilon'Pve_1'+\frac{1}{n}e_1v'P\varepsilon\varepsilon'(I-P)f|X] \\ &+ E[\frac{1}{n}e_1v'P\varepsilon\varepsilon'Pve_1'|X] \end{aligned}$$

$E(T^H H^{-1}\Im(\alpha)|X)=-\sigma^2 e_f+o_p(1)$ and
$E(\Im(\alpha)H^{-1}T^H|X)=-\sigma^2 e_f+o_p(1)$

$$\begin{aligned} E(T^H H^{-1}\Im(\alpha)H^{-1}T^H|X) &= \sigma^2 H O_p([\frac{1}{n\alpha^2}+\frac{1}{\alpha\sqrt{n}}+\Delta_f]^2) \\ &= O_p([\frac{1}{n\alpha^2}+\frac{1}{\alpha\sqrt{n}}+\Delta_f]^2) \end{aligned}$$

We have

$$\begin{aligned} E(A(\alpha)|X) &= \sigma^2 H+\sigma^2 e_{2f}+E[\frac{1}{n}e_1v'P\varepsilon\varepsilon'Pve_1'|X] \\ &- E[\frac{1}{n}f'(I-P)\varepsilon\varepsilon'Pve_1'+\frac{1}{n}e_1v'P\varepsilon\varepsilon'(I-P)f|X] \\ &+ \frac{1}{n}f'Pve_1'+\frac{1}{n}e_1v'Pf+O_p([\frac{1}{n\alpha^2}+\frac{1}{\alpha\sqrt{n}}+\Delta_f]^2) \end{aligned}$$

By Lemma 5 (viii) of Carrasco (2012), we have

$$E[\frac{1}{n}f'(I-P)\varepsilon\varepsilon'Pve_1'+\frac{1}{n}e_1v'P\varepsilon\varepsilon'(I-P)f|X]=O_p(\sqrt{\Delta_{2f}}/\sqrt{\alpha n})$$

and $\frac{1}{n}e_1v'(P-P^2)f=O_p(\sqrt{\Delta_{2f}}/\sqrt{\alpha n})$.

By Lemma 5 (iii) of Carrasco (2012), $\frac{1}{n}f'Pve_1'+\frac{1}{n}e_1v'Pf=O_p(\frac{1}{n\alpha})$.

And, by Lemma 5 (iv) of Carrasco (2012),

$$E[\frac{1}{n}e_1v'P\varepsilon\varepsilon'Pve_1'/n|X]=\frac{1}{n}(\sum_j q_j)^2\sigma^4 e_1\iota'D'D\iota e_1'+o_p((\sum_j q_j)^2/n)$$

with $D=JRWS^{-1}R^{-1}$.

We can conclude that

$$n(\hat{\delta}_{R2sls}-\delta_0)(\hat{\delta}_{R2sls}-\delta_0)=Q(\alpha)+\hat{R}(\alpha)$$

with $E[Q(\alpha)|X]=H^{-1}\sigma^2+H^{-1}\left[\sigma^2 e_{2f}+\frac{1}{n}(\sum_j q_j)^2\sigma^4 e_1\iota' D' D\iota e_1'\right]H^{-1}$ and

$$r(\alpha)=E(\hat{R}(\alpha)|X)=o_p((\sum_j q_j)^2/n)+O_p([\frac{1}{n\alpha^2}+\frac{1}{\alpha\sqrt{n}}+\Delta_f]^2+\frac{1}{n\alpha}+\frac{\Delta_{2f}}{\sqrt{\alpha n}})$$

$S(\alpha)=H^{-1}\left[\sigma^2 e_{2f}+\frac{1}{n}(\sum_j q_j)^2\sigma^4 e_1\iota' D' D\iota e_1'\right]H^{-1}$

It can be noticed that $r(\alpha)/tr(S(\alpha))=o_p(1)$, our argument are similar to those used in Carrasco (2012). Which means that $S(\alpha)$ is the dominant part of the mean squared error of the estimation of the model using regularized 2SLS.

Proof of Proposition 3

First, we show that the minimizer of $\sigma^{-2}\langle(K_n^{\alpha})^{-1/2}\bar{S}_n(.),(K_n^{\alpha})^{-1/2}\bar{S}_n(.)\rangle+\frac{1}{n}\bar{g}_2(\theta)'V\bar{g}_2(\theta)$ is a consistent estimator of θ_0.

$$\langle(K_n^{\alpha})^{-1/2}\bar{S}_n(.),(K_n^{\alpha})^{-1/2}\bar{S}_n(.)\rangle=\frac{1}{n}\varepsilon'(\theta)P\varepsilon(\theta)$$

with $\varepsilon(\theta)=JR(\rho)(Y-Z\delta)$, $R(\rho)=R+(\rho_0-\rho)M$ and $S(\lambda)=S+(\lambda_0-\lambda)W$.

$$\varepsilon(\theta)=v(\theta)+h(\theta)$$

$v(\theta)=f(\delta_0-\delta)+(\rho_0-\rho)F(\delta_0-\delta)$ where $F=JME(Z)$
and $h(\theta)=(\lambda_0-\lambda)JRWS^{-1}R^{-1}\varepsilon+J\varepsilon+(\rho_0-\rho)JMR^{-1}\varepsilon+(\lambda_0-\lambda)(\rho_0-\rho)JMWS^{-1}R^{-1}\varepsilon$ thus $h(\theta)=JR(\rho)S(\lambda)S^{-1}R^{-1}\varepsilon$.

$$\begin{aligned}\frac{1}{n}\varepsilon'(\theta)P\varepsilon(\theta) &= \frac{1}{n}v(\theta)'(\theta)Pv(\theta)\\ &+ \frac{1}{n}v(\theta)'Ph(\theta)+\frac{1}{n}h(\theta)'Pv(\theta)\\ &+ \frac{1}{n}h(\theta)'Ph(\theta)\end{aligned}$$

$$\begin{aligned}
\frac{1}{n}v(\theta)'(\theta)Pv(\theta) &= \frac{1}{n}(\delta_0-\delta)'f'Pf(\delta_0-\delta)\\
&+ \frac{1}{n}(\rho_0-\rho)(\delta_0-\delta)'f'PF(\delta_0-\delta)+\frac{1}{n}(\rho_0-\rho)(\delta_0-\delta)'F'Pf(\delta_0-\delta)\\
&+ \frac{1}{n}(\rho_0-\rho)^2(\delta_0-\delta)'F'PF(\delta_0-\delta)\\
&= \frac{1}{n}(\delta_0-\delta)'f(\rho)'Pf(\rho)(\delta_0-\delta)\\
&= (\delta_0-\delta)'\left[\frac{1}{n}f(\rho)'f(\rho)+\frac{1}{n}f(\rho)'(I-P)f(\rho)\right](\delta_0-\delta)
\end{aligned}$$

By Carrasco (2012) Lemma 5 (i) we can conclude that

$$\frac{1}{n}v(\theta)'(\theta)Pv(\theta)=(\delta_0-\delta)'\left[lim_{n\to\infty}\frac{1}{n}f(\rho)'f(\rho)+\right](\delta_0-\delta)+o_p(1)$$

and by Lemma 2 (ii) $\frac{1}{n}v(\theta)'Ph(\theta)+\frac{1}{n}h(\theta)'Pv(\theta)=O_p(1/\alpha\sqrt{n})$ and $\frac{1}{n}h(\theta)'Ph(\theta)=O_p(1/n\alpha^2)$ thus

$$\frac{1}{n}\varepsilon'(\theta)P\varepsilon(\theta)=(\delta_0-\delta)'\left[lim_{n\to\infty}\frac{1}{n}f(\rho)'f(\rho)+\right](\delta_0-\delta)+o_p(1) \tag{12}$$

as n and $\alpha\sqrt{n}$ go to infinity.

Under assumption 3' $\lim_{n\to\infty}\frac{1}{n}f(\rho)'f(\rho)$ is a finite nonsingular matrix for any ρ such that $R(\rho)$ is nonsingular. Therefore,

$(\delta_0-\delta)'\left[lim_{n\to\infty}\frac{1}{n}f(\rho)'f(\rho)+\right](\delta_0-\delta)\geq 0$ with equality if $\delta=\delta_0$.

The second part of the regularized GMM function minimized the optimal GMM function of the moments $\bar{g}_2(\theta)=\frac{\mu_3}{\sigma_0^2}\phi'P^\alpha\varepsilon(\theta)-g_2(\theta)$ using similar arguments,

$$\begin{aligned}
\frac{1}{n}\phi'P^\alpha\varepsilon(\theta) &= \frac{1}{n}\phi'P^\alpha v(\theta)+\frac{1}{n}\phi'P^\alpha h(\theta)\\
&= \frac{1}{n}\phi'P^\alpha f(\rho)(\delta_0-\delta)+o_p(1)\\
&= lim_{n\to\infty}\frac{1}{n}\phi'P^\alpha f(\rho)(\delta_0-\delta)+o_p(1).
\end{aligned}$$

as n and $\alpha\sqrt{n}$ go to infinity. Hence, $\frac{1}{n}\phi'P^\alpha\varepsilon(\theta)$ has a limit of zero at $\delta=\delta_0$. Because the number of quadratic is fixed, by a similar argument as in the proof of Proposition 4 of Liu and Lee (2010) $\frac{1}{n}g_2(\theta)-\frac{1}{n}E[g_2(\theta)]=o_p(1)$ uniformly in θ in any bounded set Θ. The consistency follows from the uniform convergence in probability and the identification uniqueness of the limiting function.

This prove that the of $\sigma^{-2}\langle(K_n^\alpha)^{-1/2}\bar{S}_n(.),(K_n^\alpha)^{-1/2}\bar{S}_n(.)\rangle+\frac{1}{n}\bar{g}_2(\theta)'V\bar{g}_2(\theta)$ is a consistent estimator of θ_0.

It remains to show that
$(\tilde{\sigma}^{-2}-\sigma^{-2})\frac{1}{n}\varepsilon'(\theta)P\varepsilon(\theta)+\frac{1}{n}\varepsilon(\theta)'P\phi\left[\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV\right]\frac{1}{n}\phi'P\varepsilon(\theta)-\frac{2}{n}\varepsilon(\theta)'P\phi\left[\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV\right]\frac{1}{n}g_2(\theta)+\frac{1}{n}g_2(\theta)'\left[\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV\right]\frac{1}{n}g_2(\theta)=o_p(1)$ uniformly in $\theta\in\Theta$.

Note that $\frac{1}{n}\varepsilon'(\theta)P\varepsilon(\theta)=O_p(1)$ uniformly in $\theta\in\Theta$.

$$\begin{aligned}\frac{1}{n}\tilde{V}^{-1}-\frac{1}{n}V^{-1} &= (\frac{\tilde{\mu}_3^2}{\tilde{\sigma}^2}-\frac{\mu_3^2}{\sigma^2})\frac{1}{n}\phi'P^2\phi+[(\tilde{\mu}_4-3\tilde{\sigma}^4)-(\mu_4-3\sigma^4)]\frac{1}{n}\phi'\phi\\ &+ (\tilde{\sigma}^4-\sigma^4)\frac{1}{n}\Gamma-2(\frac{\tilde{\mu}_3^2}{\tilde{\sigma}^2}-\frac{\mu_3^2}{\sigma^2})\frac{1}{n}\phi'P\phi\\ &= o_p(1)\end{aligned}$$

By Lemma 2 (i) and the fact that U_j are UB. It follows that $n\tilde{V}-nV=o_p(1)$ by the continuous mapping theorem.

$$\|\frac{1}{n}\varepsilon(\theta)'P\phi\left[(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV)\right]\frac{1}{n}\phi'P\varepsilon(\theta)\|\leq\|\frac{1}{n}\varepsilon(\theta)'P\phi\|^2.\|(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV)\|$$

as $\frac{1}{n}\|\varepsilon(\theta)'P\phi\|=O_p(1)$ uniformly in $\theta\in\Theta$. Hence, $\|\frac{1}{n}\varepsilon(\theta)'P\phi\|^2.\|(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV)\|=o_p(1)$ By a similar argument in the proof of Proposition 4 Liu and Lee (2010), $\frac{1}{n}\|g_2(\theta)\|=O_p(1)$ uniformly in $\theta\in\Theta$.

$$\|\frac{1}{n}\varepsilon(\theta)'P\phi\left[(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV)\right]\frac{1}{n}g_2(\theta)\|\leq\frac{1}{n}\|\varepsilon(\theta)'P\phi\|.\|(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV)\|.\frac{1}{n}\|g_2(\theta)\|=o_p(1)$$

and

$$\|\frac{1}{n}g_2(\theta)'\left[(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV)\right]\frac{1}{n}g_2(\theta)\|\leq\|(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV)\|.\|\frac{1}{n}\|g_2(\theta)\|^2=o_p(1)$$

uniformly in $\theta\in\Theta$. It follows that
$(\tilde{\sigma}^{-2}-\sigma^{-2})\frac{1}{n}\varepsilon'(\theta)P\varepsilon(\theta)+\frac{1}{n}\varepsilon(\theta)'P\phi\left[\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV\right]\frac{1}{n}\phi'P\varepsilon(\theta)-\frac{2}{n}\varepsilon(\theta)'P\phi\left[\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV\right]\frac{1}{n}g_2(\theta)+\frac{1}{n}g_2(\theta)'\left[\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{V}-\frac{\mu_3}{\sigma^2}nV\right]\frac{1}{n}g_2(\theta)=o_p(1)$ uniformly in $\theta\in\Theta$. This prove the consistency of regularized GMM.

Proof of Proposition 4

Let $\mathbf{H}(\theta)=\tilde{\sigma}^{-2}\langle(K_n^\alpha)^{-1/2}\bar{S}_n(.),(K_n^\alpha)^{-1/2}\bar{S}_n(.)\rangle+\frac{1}{n}\bar{g}_2(\theta)'\tilde{V}\bar{g}_2(\theta)$ be the objective function of the regularized GMM. By the Taylor expansion of $\frac{\partial\mathbf{H}(\hat{\theta})}{\partial\theta}=0$ at θ_0,

$$\sqrt{n}(\hat{\theta}-\theta_0) = -\left[\frac{\partial^2\mathbf{H}(\bar{\theta})}{\partial\theta\partial'\theta}\right]^{-1}\times\sqrt{n}\frac{\partial\mathbf{H}(\theta_0)}{\partial\theta}$$

where $\bar{\theta}$ is between $\hat{\theta}$ and θ_0.

$\mathbf{H}(\theta) = \tilde{\sigma}^{-2}\frac{1}{n}\varepsilon(\theta)'P\varepsilon(\theta) + \frac{1}{n}\bar{g}_2(\theta)'\tilde{V}\bar{g}_2(\theta)$

Hence, $\frac{1}{2}\frac{\partial \mathbf{H}(\theta)}{\partial \theta} = \sigma_0^{-2}\frac{1}{n}\frac{\partial \varepsilon(\theta)'}{\partial \theta}P\varepsilon(\theta) + \frac{1}{n}\frac{\partial \bar{g}_2(\theta)'}{\partial \theta}\tilde{V}\bar{g}_2(\theta)$

$$\sqrt{n}\frac{\partial \mathbf{H}(\theta_0)}{\partial \theta} = 2\left[\tilde{\sigma}^{-2}\frac{1}{\sqrt{n}}\frac{\partial \varepsilon(\theta_0)'}{\partial \theta}P\varepsilon(\theta_0) + \frac{1}{\sqrt{n}}\frac{\partial \bar{g}_2(\theta_0)'}{\partial \theta}\tilde{V}\bar{g}_2(\theta_0)\right]$$

$$\begin{aligned} -\sigma_0^2\frac{1}{\sqrt{n}}\frac{\partial \varepsilon(\theta_0)'}{\partial \theta}P\varepsilon(\theta_0) &= \sigma_0^{-2}\frac{1}{\sqrt{n}}\left[MR^{-1}\varepsilon, RZ\right]'P\varepsilon \\ &= \sigma_0^{-2}\frac{1}{\sqrt{n}}\left[MR^{-1}\varepsilon, f + ve_1'\right]'P\varepsilon \\ &= \sigma_0^{-2}\frac{1}{\sqrt{n}}(0, f'\varepsilon) + O_p(1/\alpha^2\sqrt{n}) + o_p(1) \end{aligned}$$

And It can be shown that

$$\frac{1}{\sqrt{n}}\frac{\partial \bar{g}_2(\theta_0)'}{\partial \theta}\tilde{V}\bar{g}_2(\theta_0) = \frac{\partial \bar{g}_2(\theta_0)'}{\partial \theta}\tilde{V}\frac{1}{\sqrt{n}}\bar{g}_2(\theta_0)$$

Let $D_2 = E(\frac{\partial g_2(\theta)}{\partial \theta'}) = -\sigma_0^2\begin{pmatrix} tr[(U_1+U_1')MR^{-1}] & tr[(U_1+U_1')RWS^{-1}R^{-1}] & 0 \\ tr[(U_2+U_2')MR^{-1}] & tr[(U_2+U_2')RWS^{-1}R^{-1}] & 0 \\ . & . & . \\ . & . & . \\ . & . & . \\ tr[(U_q+U_q')MR^{-1}] & tr[(U_q+U_q')RWS^{-1}R^{-1}] & 0 \end{pmatrix}$ and

$$\bar{D}_2 = D_2 - \frac{\mu_3}{\sigma_0^2}(0, \phi' f)$$

It can be shown that $-\frac{1}{n}\frac{\partial \bar{g}_2(\theta_0)'}{\partial \theta} = \frac{1}{n}\bar{D}_2' + O_p(1/\alpha\sqrt{n}) + o_p(1)$

Using central limit theorem of Kelejian and Prucha (2001) as in Liu and Lee (2010).

$$\frac{1}{\sqrt{n}}\frac{\partial \bar{g}_2(\theta_0)'}{\partial \theta}\tilde{V}\bar{g}_2(\theta_0) \xrightarrow{d} \mathcal{N}\left(0, plim\bar{D}_2'V\bar{D}_2\right)$$

Moreover, by Lemma 5 $\frac{1}{\sqrt{n}}(0, f'\varepsilon)$ and $\frac{1}{\sqrt{n}}\bar{g}_2(\theta_0)$ are asymptotically uncorrelated.

Thus, $\sqrt{n}\frac{\partial \mathbf{H}(\theta_0)}{\partial \theta} \xrightarrow{d} 2\mathcal{N}\left(0, \sigma_0^{-2}D(0, H) + plim\bar{D}_2'V\bar{D}_2\right)$

The hessian is

$$\frac{\partial^2 \mathbf{H}(\bar{\theta})}{\partial \theta \partial' \theta} = 2\left[\tilde{\sigma}^{-2}\frac{1}{n}\frac{\partial \varepsilon(\bar{\theta})'}{\partial \theta}P\frac{\partial \varepsilon(\bar{\theta})}{\partial \theta'} + \frac{1}{n}\frac{\partial \bar{g}_2(\bar{\theta})'}{\partial \theta}\tilde{V}\frac{\partial \bar{g}_2(\bar{\theta})}{\partial \theta'} + \tilde{\sigma}^{-2}\frac{1}{n}\frac{\partial^2 \varepsilon(\bar{\theta})'}{\partial \theta \partial \theta'}P\varepsilon(\bar{\theta}) + \frac{1}{n}\frac{\partial^2 \bar{g}_2(\bar{\theta})'}{\partial \theta \partial \theta'}\tilde{V}\bar{g}_2(\bar{\theta})\right]$$

Under assumptions of proposition 3, $\bar{\theta} \xrightarrow{p} \theta_0$. In addition, $\frac{1}{n}\frac{\partial^2 \bar{g}_2(\bar{\theta})'}{\partial\theta\partial\theta'} = O_p(1)$ and $\frac{1}{n}\frac{\partial^2 \varepsilon(\bar{\theta})'}{\partial\theta\partial\theta'} = O_p(1)$ uniformly for $\theta \in \Theta$.

Thus, $\frac{1}{n}\frac{\partial^2 \varepsilon(\bar{\theta})'}{\partial\theta\partial\theta'}P\varepsilon(\bar{\theta}) = \frac{1}{n}\frac{\partial^2 \varepsilon(\bar{\theta})'}{\partial\theta\partial\theta'}PJR(\bar{\rho})((Y - Z\bar{\delta})) = O_p(\frac{1}{\alpha\sqrt{n}})$ and $\frac{1}{n}\frac{\partial^2 \bar{g}_2(\bar{\theta})'}{\partial\theta\partial\theta'}\tilde{V}\bar{g}_2(\bar{\theta}) = o_p(1)$

$$
\begin{aligned}
\frac{1}{n}\frac{\partial\varepsilon(\bar{\theta})'}{\partial\theta}P\frac{\partial\varepsilon(\bar{\theta})}{\partial\theta'} &= \frac{1}{n}[M(Y - Z\bar{\delta}), R(\bar{\rho})Z]'P[M(Y - Z\bar{\delta}), R(\bar{\rho})Z] \\
&= \frac{1}{n}[M(Y - Z\delta_0), RZ]'P[M(Y - Z\delta_0), RZ] + o_p(1) \\
&= \frac{1}{n}[MR^{-1}\varepsilon, f + ve_1']'P[MR^{-1}\varepsilon, f + ve_1'] + o_p(1) \\
&= \frac{1}{n}\begin{bmatrix} ([MR^{-1}\varepsilon)'PMR^{-1}\varepsilon & ([MR^{-1}\varepsilon)'P(f + ve_1') \\ (f + ve_1')'PMR^{-1}\varepsilon & (f + ve_1')'P(f + ve_1') \end{bmatrix} + o_p(1) \\
&= D(0, H) + O_p(1/\alpha^2 n) + o_p(1)
\end{aligned}
$$

And $\frac{1}{n}\frac{\partial\bar{g}_2(\bar{\theta})'}{\partial\theta}\tilde{V}\frac{\partial\bar{g}_2(\bar{\theta})}{\partial\theta'} = plim\bar{D}_2'V\bar{D}_2 + O_p(1/\alpha^2 n) + o_p(1)$.

Using the fact that $\tilde{\sigma}^{-2}$ is a consistent estimator of σ_0^{-2} and $n\tilde{V} - nV = o_p(1)$

$$
\frac{\partial^2 \mathbf{H}(\bar{\theta})}{\partial\theta\partial'\theta} = 2[\sigma_0^{-2}D(0, H) + plim\bar{D}_2'V\bar{D}_2] + o_p(1)
$$

The result follows by Slutzky theorem.

Recent Kent Discussion Papers in Economics

16/06: 'The Post-crisis Slump in Europe: A Business Cycle Accounting Analysis', Florian Gerth and Keisuke Otsu

16/05: 'The Revenue Implication of Trade Liberalisation in Sub-Saharan Africa: Some new evidence', Lanre Kassim

16/04: 'The rise of the service economy and the real return on capital', Miguel León-Ledesma and Alessio Moro

16/03: 'Is there a mission drift in microfinance? Some new empirical evidence from Uganda', Francis Awuku Darko

16/02: 'Early Marriage, Social Networks and the Transmission of Norms', Niaz Asadullah and Zaki Wahhaj

16/01: 'Intra-household Resource Allocation and Familial Ties', Harounan Kazianga and Zaki Wahhaj

15/21: 'Endogenous divorce and human capital production', Amanda Gosling and María D. C. García-Alonso

15/20: 'A Theory of Child Marriage', Zaki Wahhaj

15/19: 'A fast algorithm for finding the confidence set of large collections of models', Sylvain Barde

15/18: 'Trend Dominance in Macroeconomic Fluctuations', Katsuyuki Shibayama

15/17: 'Efficient estimation with many weak instruments using regularization techniques', Marine Carrasco and Guy Tchuente

15/16: 'High school human capital portfolio and college outcomes', Guy Tchuente

15/15: 'Regularized LIML for many instruments, Marine Carrasco and Guy Tchuente

15/14: 'Agglomeration Economies and Productivity Growth: U.S. Cities, 1880-1930', Alexander Klein and Nicholas Crafts

15/13: 'Microcredit with Voluntary Contributions and Zero Interest Rate - Evidence from Pakistan', Mahreen Mahmud

15/12: 'Act Now: The Effects of the 2008 Spanish Disability Reform', Matthew J. Hill, Jose Silva and Judit Vall

15/11: 'Testing for Level Shifts in Fractionally Integrated Processes: a State Space Approach', Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris