Two Stage Least Squares Estimation of Spatial Autoregressive Models with Endogenous Regressors and Many Instruments*

Xiaodong Liu
Department of Economics
University of Colorado at Boulder
Boulder, CO 90309, USA

Lung-fei Lee
Department of Economics
Ohio State University
Columbus, OH 43210, USA

First version: August 2009
Current version: June 2011

Abstract

This paper considers the IV estimation of spatial autoregressive models with endogenous regressors in the presence of many instruments. To improve asymptotic efficiency, it may be desirable to use many valid instruments. However, finite sample properties of IV estimators can be sensitive to the number of instruments. For a spatial model with endogenous regressors, this paper derives the asymptotic distribution of the 2SLS estimator when the number of instruments grows with the sample size, and suggests a bias-correction procedure based on the leading-order many-instrument bias. The paper also gives the Nagar-type approximate MSEs of the 2SLS estimator and the bias-corrected 2SLS estimator, which can be minimized to choose instruments as in Donald and Newey (2001). A limited Monte Carlo experiment is carried out to study the finite sample performance of the instrument selection procedure.

JEL classification: C13, C21

Key words: spatial autoregressive model, 2SLS, endogenous regressors, instrument selection

*We thank the editor, associate editor and three referees for helpful comments. The paper was presented at the 2009 international symposium on econometric theory and applications, Kyoto, Japan, the 19th annual meeting of the midwest econometrics group, West Lafayette, IN. We thank participants at the meetings for valuable comments.

1 Introduction

This paper considers the instrumental variable (IV) estimation of spatial autoregressive (SAR) models with endogenous regressors in the presence of many instruments. The paper derives the asymptotic distribution of the two-stage least squares (2SLS) estimator when the number of instruments grows with the sample size, and suggests a bias-correction procedure based on the leading-order many-instrument bias. To improve asymptotic efficiency, it may be desirable to use many valid instruments. However, finite sample properties of IV estimators can be sensitive to the number of instruments (see, e.g., Morimune, 1983; Bekker, 1994). Donald and Newey (2001) suggest minimizing the Nagar-type approximation (Nagar, 1959) of the mean square error (MSE) to choose among valid instruments. We derive the approximate MSEs of the 2SLS estimator and the bias-corrected 2SLS (C2SLS) estimator for the SAR model. The approximate MSEs are more complicated than those in Donald and Newey (2001) due to the sample observations being spatially correlated.

We consider the following SAR model with endogenous regressors:

$$y_n = \lambda_0 W_n y_n + Z_{2n}\gamma_0 + \epsilon_n. \tag{1}$$

In this model, n is the total number of spatial units, y_n is an n-dimensional vector of dependent variables, W_n is an $n \times n$ spatial weights matrix of known constants with a zero diagonal, Z_{2n} is an $n \times m$ matrix of explanatory variables that are possibly correlated with ϵ_n, and the disturbances $\epsilon_{n1}, \cdots, \epsilon_{nn}$ of the n-dimensional vector ϵ_n are i.i.d. $(0, \sigma_\epsilon^2)$. Let X_n be a matrix of exogenous variables. The nonparametric reduced-form relationship between Z_{2n} and X_n is given by $Z_{2n} = \bar{Z}_{2n} + U_n$, where $\bar{Z}_{2n} = \mathrm{E}(Z_{2n}) = f(X_n)$. We assume u_{ni}'s in $U_n = (u'_{n1}, \cdots, u'_{nn})'$ are i.i.d. such that $E(u_{ni}) = 0$ and $\mathrm{E}(u'_{ni}u_{ni}) = \Sigma_u$. The potential correlation between Z_{2n} and ϵ_n is captured by $\mathrm{E}(u_{ni}\epsilon_{ni}) = \sigma_{u\epsilon}$.[1] Let $Z_n = (W_n y_n, Z_{2n})$ and $\delta_0 = (\lambda_0, \gamma_0')'$. (1) can be rewritten more compactly as $y_n = Z_n\delta_0 + \epsilon_n$.

The model (1) represents an equilibrium equation, so $(I_n - \lambda_0 W_n)$ is assumed to be invertible. The equilibrium vector y_n is given by $y_n = (I_n - \lambda_0 W_n)^{-1}(Z_{2n}\gamma_0 + \epsilon_n)$. Let $G_n(\lambda) = W_n(I_n - \lambda W_n)^{-1}$ and $G_n = G_n(\lambda_0)$. It follows that $W_n y_n = G_n(Z_{2n}\gamma_0 + \epsilon_n)$. Let $G_n^{(p)} = (W_n, \cdots, W_n^{p+1})$ and $\Psi_q(X_n)$ be a known $n \times q$ matrix that depends on X_n.[2] The set of instruments is given by an $n \times K$

[1] When the jth regressor in Z_{2n} is exogenous, we have $\sigma_{u\epsilon,j} = 0$.
[2] For example, the functions in $\Psi_q(X_n)$ may be a series of base functions in the approximation of $\bar{Z}_{2n}$.

matrix $Q_K = [G_n^{(p)}(I_{p+1} \otimes \Psi_q(X_n)), \Psi_q(X_n)]$ for $K \geq (m+1)$, where $\otimes$ denotes the Kronecker product operator.[3] Let $P_K = Q_K(Q_K'Q_K)^- Q_K'$, where A^- denotes a generalized inverse. The 2SLS estimator considered is $\hat{\delta}_{2sls,n} = (Z_n' P_K Z_n)^{-1} Z_n' P_K y_n$.

Let F_n and V_n denote the deterministic and stochastic parts of Z_n respectively, such that $Z_n = F_n + V_n$ where $F_n = [G_n \bar{Z}_{2n}\gamma_0, \bar{Z}_{2n}]$ and $V_n = [G_n(U_n\gamma_0 + \epsilon_n), U_n]$. The asymptotic variance of the 2SLS estimator will decrease as linear combinations of Q_K can approximate F_n more closely. The efficiency lower bound for IV estimators can be attained if a linear combination of Q_K equals F_n (Chamberlain, 1987). Under some regularity conditions, G_n can be approximated by a linear combination of $G_n^{(p)}$ as p goes to infinity.[4] To achieve asymptotic efficiency, we need not only to choose $\Psi_q(X_n)$, likely with a large number of columns q, so that $\bar{Z}_{2n}$ can be closely approximated by linear combinations of $\Psi_q(X_n)$ as in Donald and Newey (2001), but also to choose a big p so that G_n can be closely approximated by a linear combination of $G_n^{(p)}$. On the other hand, when the number of instruments increases too fast relative to the sample size, the 2SLS estimator can be asymptotically biased of a certain order (see, e.g., Morimune, 1983). Hence, the MSE, which takes both bias and variance into account, appears to be an appropriate criterion to choose the instrument set.

The current paper makes an interesting connection between spatial models with endogenous regressors and models that are characterized by the availability of many instruments. Spatial models with endogenous regressors have been considered by Kelejian and Prucha (2004) in the case of simultaneous systems of spatially interrelated equations and by Kelejian and Prucha (2007) in the case of a general spatial regression model with spatially correlated error terms and unknown heteroskedasticity in the innovations. Kelejian and Prucha (2007) presented a general SAR model $y_n = X_n\beta_0 + \lambda_0 W_n y_n + Y_n\gamma_0 + u_n$, where Y_n denotes the matrix of endogenous regressors. As indicated in Kelejian and Prucha (2007), this equation may represent an equation in a linear system of equations such that $\mathrm{E}(Y_n) = \sum_{s=0}^{\infty} W_n^s \underline{X}_n \Pi_s$, where $\underline{X}_n$ is the matrix of all exogenous variables in the system. In that case, one may have many linear moment conditions as $W_n^s \underline{X}_n$ for all s can be used as instruments. Similarly, if some elements of Y_n are generated by a nonlinear model (Kelejian and Prucha, 2007, p143), without a specific nonlinear functional form for the conditional mean of

[3] Following Donald and Newey (2001), K serves as both the number of instruments and the index of the instrument set. More generally, it is possible to specify a different index for the instrument set and use the MSE criterion introduced later to choose the instrument set. (Newey, 2007)

[4] See Lemma 2.1 in the next section.

Y_n, we might also end up with many instruments for estimation. The focus of Kelejian and Prucha (2007) is on the robust estimation with a general pattern of spatial correlation and heteroskedasticity, and a 2SLS estimator with a predetermined number of instruments is proposed to estimate the general model. The issue raised from the availability of many instruments and the issue of how to choose the optimal number of instruments have not been considered in that paper.

Since Bekker's (1994) seminal work, the study of many-instrument asymptotics, where the number of instruments increases with the sample size, has attracted a lot of attention in the IV estimation literature. Some recent developments in this area include selection of instruments and moment conditions (e.g., Donald and Newey, 2001), various types of asymptotic sequences, bias-correction (e.g., Hansen et al., 2008), and the effects of non-normality on the asymptotic distribution (e.g., Bekker and van der Ploeg, 2005; van Hasselt, 2010). In particular, Donald and Newey (2001) have considered a case where the number of instruments grows with, but at a slower rate than, the sample size and suggested minimizing the Nagar-type approximation (Nagar, 1959) of the mean square error (MSE) to choose among valid instruments. This literature of many instrument estimation has focused on the case with independent observations.

In this paper, we also consider the case where the number of instruments increase with the sample size. The many-instrument asymptotics allows us to study high-order properties of the proposed 2SLS estimator and the bias-corrected 2SLS (C2SLS) estimator for the SAR model. In particular, it facilitates the derivation of the Nagar-type approximate MSE which can be used as a criterion function to choose instruments as in Donald and Newey (2001). The theoretical innovation of this paper is that we allow for spatial correlation in the dependent variable, which complicates matters as compared to models with only independent observations. The main differences will be discussed later in this paper.

The rest of the paper is organized as follows. Section 2 derives the asymptotic distribution of the 2SLS estimator and proposes a bias-correction procedure to adjust for the leading-order bias due to many instruments. Section 3 gives the approximate MSEs of the 2SLS and C2SLS estimators, which can be used as criterion functions for choosing instruments. The estimation of the criterion functions is discussed in Section 4. Section 5 presents some Monte-Carlo evidence on the finite sample performance of the instrument selection procedure. Section 6 briefly concludes. The notations are collected in Section 7 for easy reference. A technical appendix with all the proofs is available on the web page http://spot.colorado.edu/~xiaodong/.

2 The 2SLS Estimation and Bias Correction

To derive the asymptotic properties of the 2SLS estimator, we assume the following regularity conditions. Henceforth, uniform boundedness in row (column) sums in absolute value of a sequence of square matrices $\{A_n\}$ will be abbreviated as UBR (UBC), and uniform boundedness in both row and column sums in absolute value as UB.[5] Let $\|A\| = \sqrt{\text{tr}(A'A)}$ denote the Frobenius (Euclidean) norm for a matrix A.

Assumption 1 $\{\epsilon_{ni}, u_{ni}\}$ are i.i.d. with zero mean, $\text{E}(\epsilon_{ni}^2) = \sigma_\epsilon^2$, $\text{E}(u_{ni}' u_{ni}) = \Sigma_u$, and $\text{E}(u_{ni}\epsilon_{ni}) = \sigma_{u\epsilon}$. $\text{E}(|\epsilon_{ni}|^4)$, $\text{E}(||u_{ni}||^4)$ and $\text{E}(||u_{ni}\epsilon_{ni}||^2)$ are bounded, uniformly in n.

Assumption 2 The sequences of matrices $\{W_n\}$ and $\{(I_n - \lambda_0 W_n)^{-1}\}$ are UB.

Assumption 3 (i) $\bar{H} = \lim_{n\to\infty} H_n$, where $H_n = \frac{1}{n}F_n' F_n$, is a finite nonsingular $(m+1)$ square matrix; (ii) for each K there exists π_K such that $\frac{1}{n}||F_n - Q_K \pi_K||^2 \to 0$ as $n, K \to \infty$.

Assumption 1 imposes restrictions on the moments of the error terms. The uniform boundedness of $\{W_n\}$ and $\{(I_n - \lambda_0 W_n)^{-1}\}$ in Assumption 2 limits spatial dependence among the units to a tractable degree and is originated in Kelejian and Prucha (1999). Assumption 3 (i) gives a sufficient identification condition for δ_0. Assumption 3 (ii) concerns approximation of the unknown reduced-form F_n. If $\sup_n ||\lambda_0 W_n||_\infty < 1$, where $||A_n||_\infty = \max_i \sum_{j=1}^n |A_{n,ij}|$ is the row sum matrix norm, $G_n = W_n(I_n - \lambda_0 W_n)^{-1} = \sum_{j=0}^\infty \lambda_0^j W_n^{j+1} = \sum_{j=0}^p \lambda_0^j W_n^{j+1} + (\lambda_0 W_n)^{p+1} G_n$. With the vector of coefficients $\lambda_0^{(p)} = (1, \lambda_0, \cdots, \lambda_0^p)'$, G_n can be approximated by $G_n^{(p)}(\lambda_0^{(p)} \otimes I_n)$ such that $||G_n - G_n^{(p)}(\lambda_0^{(p)} \otimes I_n)||_\infty \leq ||\lambda_0 W_n||_\infty^{p+1} ||G_n||_\infty = o(1)$ as $p \to \infty$.[6] Assumption 3 (ii) is implied by a condition in the following lemma.

Lemma 2.1 *Suppose* $\sup_n ||\lambda_0 W_n||_\infty < 1$ *and, for each* q*, there exists* π_q^0 *such that* $||\bar{Z}_{2n} - \Psi_q(X_n)\pi_q^0||_\infty \to 0$ *as* $n, q \to \infty$*. Then, Assumption 3* (ii) *is satisfied for* $Q_K = [G_n^{(p)}(I_{p+1} \otimes \Psi_q(X_n)), \Psi_q(X_n)]$ *where* $p, q \to \infty$ *as* $n \to \infty$*.*

[5] A sequence of square matrices $\{A_n\}$, where $A_n = [A_{n,ij}]$, is said to be UBR (UBC) if the sequence of row sum matrix norm $||A_n||_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^n |A_{n,ij}|$ (column sum matrix norm $||A_n||_1 = \max_{j=1,\cdots,n} \sum_{i=1}^n |A_{n,ij}|$) is bounded. (Horn and Johnson, 1985)

[6] If $\sup_n ||\lambda_0 W_n||_\infty < 1$, p can be small as it can have the order $O(\ln(n))$. This is so, because, by taking $p = \ln(n)$, $||\lambda_0 W_n||_\infty^p < \exp(-a\ln(n)) = 1/n^a$ for some $a > 0$. So K may be mainly determined by q. In Lemma 2.1, we provided some primitive conditions on W_n and q, which would be sufficient for Assumption 3 (ii) on K to hold. In order to make the model more general, we did not impose other restrictions on the functional relation between p, q and K.

Assumption 4 (i) The elements of Q_K are uniformly bounded constants and $\lim_{n\to\infty} \frac{1}{n} Q_K' Q_K$ exists and is nonsingular for each K; (ii) the elements of $\bar{Z}_{2n}$ are uniformly bounded constants; and (iii) $\max_i P_{K,ii} \to 0$ and $\max_i |M_{K,ii}| \to 0$ as $n, K \to \infty$, where $M_K = P_K G_n$.

The uniform boundedness conditions in Assumption 4 (i) and (ii) are for analytic tractability. The first part of Assumption 4 (iii) is originated in Donald and Newey (2001), and the second part is a generalized requirement which takes into account of spatial interactions. They imposes restrictions on the rate at which K increases with n.[7] The following proposition provides the asymptotic distribution of the 2SLS with many instruments.

Proposition 1 *Under Assumptions 1-4, if $K/n \to 0$, then $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0 - b_{2sls,n}) \xrightarrow{d} N(0, \sigma_\epsilon^2 \bar{H}^{-1})$, where $b_{2sls,n} = (Z_n' P_K Z_n)^{-1}[\mathrm{tr}(M_K)(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2), K\sigma_{u\epsilon}]' = O_p(K/n)$.*

When the number of instruments grows at a rate slower than the sample size, the 2SLS estimator is consistent and asymptotically normal. However, the asymptotic distribution of the 2SLS estimator may not center around the true parameter value due to the presence of many instruments. The following corollary summarizes the asymptotic properties of the 2SLS for different divergent rates of K relative to n.

Corollary 2 *Under Assumptions 1-4,* (i) *if $K^2/n \to 0$, $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0) \xrightarrow{d} N(0, \sigma_\epsilon^2 \bar{H}^{-1})$;* (ii) *if $K^2/n \to c < \infty$ and $c \neq 0$, $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0) \xrightarrow{d} N(b_{2sls}, \sigma_\epsilon^2 \bar{H}^{-1})$, where $b_{2sls} = \lim_{n\to\infty} \sqrt{n} b_{2sls,n} = \bar{H}^{-1} \lim_{n\to\infty}[\mathrm{tr}(M_K)(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2), K\sigma_{u\epsilon}]'/\sqrt{n}$; and* (iii) *if $K^2/n \to \infty$ but $K^{1+\eta}/n \to 0$ for $0 < \eta < 1$, $K^\eta(\hat{\delta}_{2sls,n} - \delta_0) \xrightarrow{p} 0$.*

The 2SLS estimator with K instruments given in Q_K is efficient when $K^2/n \to 0$, as the variance matrix $\sigma_\epsilon^2 \bar{H}^{-1}$ attains the efficiency lower bound for the class of IV estimators. On the other hand, the condition that $K/n \to 0$ in Proposition 1 is crucial for the IV estimator to be consistent. The following corollary shows the inconsistency of the 2SLS estimator if K/n converges to a nonzero constant.

Proposition 3 *Under Assumptions 1-4, if $K/n \to c \neq 0$, then $\hat{\delta}_{2sls,n} - \delta_0 - b_{2sls,n} \xrightarrow{p} 0$, where $b_{2sls,n}$ might converge to a nonzero constant.*

[7] Since P_K is an $n \times n$ projector with $\mathrm{tr}(P_K) = K$, the average diagonal element of P_K is $\bar{P}_{K,ii} = \frac{1}{n}\mathrm{tr}(P_K) = K/n$. Thus, the first part of Assumption 4 (iii) imposes some restrictions on the rate K in terms of n.

To understand this result, let's look at the first-order condition of the 2SLS, $\frac{1}{n}Z_n'P_K(y_n - Z_n\hat{\delta}_{2sls,n}) = 0$. At δ_0, $\mathrm{E}[\frac{1}{n}Z_n'P_K(y_n - Z_n\delta_0)] = \frac{1}{n}\mathrm{E}(V_n'P_K\epsilon_n) = \frac{1}{n}[\mathrm{tr}(M_K)(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2), K\sigma_{u\epsilon}]' = O(K/n)$, which may not converge to zero when the number of instruments grows at the same rate of the sample size.

The many-instrument bias of the 2SLS estimator can be corrected by the estimated leading-order bias $b_{2sls,n}$ given in Proposition 1. Let $\tilde{\delta}_n = (\tilde{\lambda}_n, \tilde{\gamma}_n')'$ be a $\sqrt{n}$-consistent 2SLS estimator based on a fixed number of instruments Q_0. The leading-order bias $b_{2sls,n}$ can be estimated by $\hat{b}_{2sls,n} = (Z_n'P_KZ_n)^{-1}[\mathrm{tr}(P_K\tilde{G}_n)(\tilde{\sigma}_{u\epsilon}\tilde{\gamma}_n + \tilde{\sigma}_\epsilon^2), K\tilde{\sigma}_{u\epsilon}]'$, where $\tilde{G}_n = G_n(\tilde{\lambda}_n)$, $\tilde{\sigma}_{u\epsilon}' = \frac{1}{n}Z_{2n}'(y_n - Z_n\tilde{\delta}_n)$, and $\tilde{\sigma}_\epsilon^2 = \frac{1}{n}(y_n - Z_n\tilde{\delta}_n)'(y_n - Z_n\tilde{\delta}_n)$. The bias-corrected 2SLS (C2SLS) estimator is given by $\hat{\delta}_{c2sls,n} = \hat{\delta}_{2sls,n} - \hat{b}_{2sls,n}$.

Proposition 4 *Under Assumptions 1-4, if $K/n \to 0$, then $\sqrt{n}(\hat{\delta}_{c2sls,n} - \delta_0) \xrightarrow{d} N(0, \sigma_\epsilon^2\bar{H}^{-1})$.*

The C2SLS estimator is properly centered and asymptotically efficient as long as the number of instruments increases slower than the sample size.

3 The MSE

Following Donald and Newey (2001), to derive the approximate MSE, we will need to find a decomposition

$$n(\hat{\delta}_n - \delta_0)(\hat{\delta}_n - \delta_0)' = \hat{L}(K) + \hat{r}(K), \tag{2}$$

where $\mathrm{E}[\hat{L}(K)] = \sigma_\epsilon^2 H_n^{-1} + S(K) + T(K)$. Here, $S(K)$ is the dominant term that depends on K in the MSE of the dominant terms for the estimator, and $T(K)$ and $\hat{r}(K)$ are remainder terms that diminish faster than $S(K)$, such that $[\hat{r}(K) + T(K)]/\mathrm{tr}(S(K)) = o_p(1)$ as $K, n \to \infty$.

To be intuitively motivated, as the 2SLS and C2SLS estimators can be presented in the form $\sqrt{n}(\hat{\delta}_n - \delta_0) = \hat{H}_n^{-1}\hat{h}_n$, we consider an expansion such that $\sqrt{n}(\hat{\delta}_n - \delta_0) = H_n^{-1}\hat{\tau} + \hat{Z}$, where $\hat{\tau} = \hat{h}_n - (\hat{H}_n - H_n)H_n^{-1}h_n$, $\hat{Z} = H_n^{-1}(H_n - \hat{H}_n)(\hat{H}_n^{-1} - H_n^{-1})h_n + (\hat{H}_n^{-1} - H_n^{-1})(\hat{h}_n - h_n)$, and $h_n = F_n'\epsilon_n/\sqrt{n}$. $\hat{\tau}$ and $\hat{Z}$ are, respectively, linear and quadratic in the differences of estimates and true values. $\hat{L}(K)$ in (2) is the leading component of $H_n^{-1}\hat{\tau}\hat{\tau}'H_n^{-1}$, and $\sigma_\epsilon^2 H_n^{-1} + S(K)$ is the dominant component of $\mathrm{E}[\hat{L}(K)]$.

Assumption 5 (i) $\lim_{K\to\infty}\frac{1}{K}\mathrm{tr}(M_K) \neq 0$, (ii) $\sum_i |M_{K,ii}| = O(K)$.

Assumption 5 (i) is for analytic simplicity. Assumption 5 (ii), together with the second part of

Assumption 4 (iii), simplifies the presentation of the approximate MSE for the C2SLS estimator.[8] Proposition 5 gives the approximate MSE for the 2SLS estimator.[9]

Proposition 5 *Under Assumptions 1-4 and 5* (i)*, if $K^2/n \to 0$, $\sigma_{u\epsilon} \neq 0$, $\mu_3 = \mathrm{E}(\epsilon_{ni}^3) = 0$ and $E(\epsilon_{ni}^2 u_{ni}) = 0$, then equation (2) for the 2SLS estimator is satisfied with*

$$S(K) = \frac{1}{n} H_n^{-1} [\mathrm{E}(V_n' P_K \epsilon_n) \mathrm{E}(\epsilon_n' P_K V_n) + \sigma_\epsilon^2 F_n' (I_n - P_K) F_n] H_n^{-1}, \tag{3}$$

where $\mathrm{E}(V_n' P_K \epsilon_n) = [\mathrm{tr}(M_K)(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2), K\sigma_{u\epsilon}]'$.

The first term in (3) is approximately the MSE of the random vector $H_n^{-1} V_n' P_K \epsilon_n / \sqrt{n}$. The dominant component of the MSE of this term comes from its expectation. This term is due to many-instrument bias and increases with the number of instruments K. The second term in $S(K)$ comes from approximating the reduced from F_n by linear combinations of the columns in the IV matrix Q_K. This is a variance term and will be smaller when Q_K, likely with a large K, provides a better approximation to F_n. Hence, for the 2SLS estimator, minimization of the MSE criterion will account for a trade-off between the bias and variance. For the 2SLS estimator, the approximate MSE has a similar interpretation as that in Donald and Newey (2001). But the expression of the MSE is more complicated due to spatial correlation. In addition to the projector P_K, which plays the important role in Donald and Newey (2001), the term $M_K = P_K G_n$, where the matrix G_n captures the spatial dependence, also plays an important role in the MSE.

The following proposition gives the approximate MSE for the C2SLS estimator. Let

$$\Pi_1(K) = \begin{bmatrix} \mathrm{tr}(M_K' M_K)(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2)^2 + \mathrm{tr}(M_K^2)\sigma_\epsilon^2(\gamma_0'\Sigma_u\gamma_0 + 2\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2) & * \\ \mathrm{tr}(M_K)[(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2)\sigma_{u\epsilon}' + \sigma_\epsilon^2(\Sigma_u\gamma_0 + \sigma_{u\epsilon}')] & K(\sigma_{u\epsilon}'\sigma_{u\epsilon} + \sigma_\epsilon^2\Sigma_u) \end{bmatrix},$$

[8] Without this assumption, the approximate MSE for the C2SLS estimator will have an additional term, which can be estimated along with $S(K)$ given in Proposition 6. We have conducted Monte Carlo experiments based on the approximate MSE with and without the estimated additional term. The simulation results are robust.

[9] If $\mu_3 \neq 0$ and $E(\epsilon_{ni}^2 u_{ni}) \neq 0$, the MSE will have an additional term

$$\frac{1}{n} H_n^{-1} [F_n' vec_D(M_K) E(\epsilon_{ni}^2 u_{ni})\gamma_0 + F_n' vec_D(M_K)\mu_3, F_n' vec_D(P_K) E(\epsilon_{ni}^2 u_{ni})]^s H_n^{-1},$$

where $A^s = A + A'$ and $vec_D(A)$ is a column vector formed by the diagonal elements of a square matrix A. This term can be estimated along with $S(K)$ in (3). We have not included it in the instrument selection criterion function for simplicity. In Donald and Newey (2001), this simplicity has also been adopted for their suggested MSE criteria.

and

$$\Pi_2(K) = \begin{bmatrix} \Pi_{2,11} & * \\ [\frac{K}{n}\mathrm{tr}(G_n) - \mathrm{tr}(M_K)]\sigma_\epsilon^2(\Sigma_u\gamma_0 + \sigma_{u\epsilon}') & 0 \end{bmatrix},$$

where $\Pi_{2,11} = 2[\frac{1}{n}\mathrm{tr}(M_K)\mathrm{tr}(G_n) - \mathrm{tr}(M_K' M_K)]\sigma_\epsilon^2(\gamma_0'\Sigma_u\gamma_0 + 2\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2) + 2[\frac{1}{n}\mathrm{tr}(M_K)\mathrm{tr}(G_n) - \mathrm{tr}(P_K G_n^2)]\sigma_\epsilon^2(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2)$.

Proposition 6 *Under Assumptions 1-5, if $K/n \to 0$, $\sigma_{u\epsilon} \neq 0$, $\mu_3 = 0$ and $E(\epsilon_{ni}^2 u_{ni}) = 0$, then equation (2) for the C2SLS estimator is satisfied with*

$$S(K) = \frac{1}{n}H_n^{-1}[\Pi_1(K) + \Pi_2(K) + \sigma_\epsilon^2 F_n'(I_n - P_K)F_n]H_n^{-1}. \tag{4}$$

For the C2SLS, the term $\frac{1}{n}H_n^{-1}\Pi_1(K)H_n^{-1}$ in (4) comes from the dominant component in the variance of $H_n^{-1}[V_n' P_K \epsilon_n - \mathrm{E}(V_n' P_K \epsilon_n)]/\sqrt{n}$.[10] The second term $\frac{1}{n}H_n^{-1}\Pi_2(K)H_n^{-1}$ in $S(K)$ is due to the estimated leading order bias. And the last term is identical to that in (3) and comes from approximation error. Thus, for the C2SLS estimator, minimization of the approximate MSE will account for a trade-off between variance terms. For the C2SLS estimator, the approximate MSE derived in this paper is more complicated and has one more term $\frac{1}{n}H_n^{-1}\Pi_2(K)H_n^{-1}$ than the one in Donald and Newey (2001). Note that, if we had independent observations instead of spatially interdependent observations, in $\Pi_2(K)$, G_n will be replaced an identity matrix and $M_K = P_K G_n$ will be replaced by the projection matrix P_K, and, then, we would have $\Pi_2(K) = 0$.

4 The Optimal Choice of K

The optimal K can be obtained by minimizing the MSE criteria $S_\xi(K) = \xi' S(K)\xi$ of a linear combination $\xi'\hat{\delta}_n$, where $S(K)$ is given in (3) for the 2SLS and in (4) for the C2SLS, and ξ is a vector of linear combination coefficients. Let $\Omega_1(K) = \mathrm{E}(V_n' P_K \epsilon_n)\mathrm{E}(\epsilon_n' P_K V_n)$, and

$$\Omega_2(K) = \mathrm{E}(V_n' P_K V_n) = \begin{bmatrix} \mathrm{tr}(M_K' M_K)(\gamma_0'\Sigma_u\gamma_0 + 2\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2) & * \\ \mathrm{tr}(M_K)(\Sigma_u\gamma_0 + \sigma_{u\epsilon}') & K\Sigma_u \end{bmatrix}.$$

For the 2SLS, $S_\xi(K)$ can be estimated, up to an additive constant, by

$$\hat{S}_\xi(K) = \frac{1}{n}\xi'\hat{H}_n^{-1}\hat{\Omega}_1\hat{H}_n^{-1}\xi + \frac{1}{n}\hat{\sigma}_\epsilon^2\xi'\hat{H}_n^{-1}[Z_n'(I_n - P_K)Z_n + \hat{\Omega}_2]\hat{H}_n^{-1}\xi, \tag{5}$$

[10] Note $H_n^{-1}[V_n' P_K \epsilon_n - \mathrm{E}(V_n' P_K \epsilon_n)]/\sqrt{n}$ has zero mean.

with some preliminary estimators given below. Note that

$$\frac{1}{n}\mathrm{E}[Z_n'(I_n-P_K)Z_n] = \mathrm{E}[(F_n+V_n)'(I_n-P_K)(F_n+V_n)] = \frac{1}{n}F_n'(I_n-P_K)F_n+\frac{1}{n}\mathrm{E}(V_n'V_n)-\frac{1}{n}\mathrm{E}(V_n'P_KV_n).$$

Hence, $\frac{1}{n}[Z_n'(I_n-P_K)Z_n+\hat{\Omega}_2]$ is an estimator of $\frac{1}{n}F_n(I_n-P_K)F_n$ in $S(K)$, up to a constant $\frac{1}{n}\mathrm{E}(V_n'V_n)$ that does not depend on K. Similarly, for the C2SLS, $S_\xi(K)$ can be estimated, up to an additive constant, by

$$\hat{S}_\xi(K) = \frac{1}{n}\xi'\hat{H}_n^{-1}(\hat{\Pi}_1+\hat{\Pi}_2)\hat{H}_n^{-1}\xi + \frac{1}{n}\hat{\sigma}_\epsilon^2\xi'\hat{H}_n^{-1}[Z_n'(I_n-P_K)Z_n+\hat{\Omega}_2]\hat{H}_n^{-1}\xi. \tag{6}$$

For a simple illustration of the preliminary estimators, consider a predetermined IV matrix $Q_{\bar{K}}$ with $\bar{K}$ instruments. Let $P_{\bar{K}} = Q_{\bar{K}}(Q_{\bar{K}}'Q_{\bar{K}})^{-}Q_{\bar{K}}'$, $\bar{\delta}_n = (Z_n'P_{\bar{K}}Z_n)^{-1}Z_n'P_{\bar{K}}y_n$, $\hat{\epsilon}_n = y_n - Z_n\bar{\delta}_n$ and $\hat{U}_n = (I_n-P_{\bar{K}})Z_{2n}$. Then, $\hat{\Omega}_1, \hat{\Omega}_2, \hat{\Pi}_1$ and $\hat{\Pi}_2$ can be obtained, with the preliminary estimators $\hat{\sigma}_{u\epsilon} = \frac{1}{n}\hat{\epsilon}_n'\hat{U}_n$, $\hat{\sigma}_\epsilon^2 = \frac{1}{n}\hat{\epsilon}_n'\hat{\epsilon}_n$, and $\hat{\Sigma}_u = \frac{1}{n}\hat{U}_n'\hat{U}_n$. On the other hand, a preliminary estimator of H_n could be $\hat{H}_n = \frac{1}{n}Z_n'P_{\bar{K}}Z_n$.

The optimal choice of K is given by $\hat{K} = \arg\min_K \hat{S}_\xi(K)$. $\hat{K}$ is optimal in the sense that the approximate MSE is asymptotically as small at $\hat{K}$ as at its minimum, i.e.,

$$\frac{S_\xi(\hat{K})}{\min_K S_\xi(K)} \xrightarrow{p} 1. \tag{7}$$

To show (7), we assume the consistency of the initial estimators in Assumption 6. We also employ constraints on the set of possible K values in Assumption 7. Let $e_F(K) = \frac{1}{n}F_n'(I_n-P_K)F_n$ and $\Delta_K = \mathrm{tr}(e_F(K))$.

Assumption 6 (i) $\bar{\delta}_n \xrightarrow{p} \delta_0$, $\hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2$, $\hat{\Sigma}_u \xrightarrow{p} \Sigma_u$, $\hat{\sigma}_{u\epsilon} \xrightarrow{p} \sigma_{u\epsilon}$, $\hat{H}_n \xrightarrow{p} \bar{H}$, (ii) for 2SLS, $\inf_K |S_\xi(K)|/(K^2+n\Delta_K) > 0$, and for C2SLS $\inf_K |S_\xi(K)|/(K+n\Delta_K) > 0$.

Assumption 7 $\sum_K [nS_\xi(K)]^{-1} \to 0$.

The following proposition shows that $\hat{K}$ is the high-order asymptotically optimal choice for the number of instruments.

Proposition 7 *Under Assumptions 1-7, for* $\hat{K} = \arg\min_K \hat{S}_\xi(K)$*, (7) is satisfied for both 2SLS and C2SLS.*

5 Simulation Study

We conduct a small Monte-Carlo experiment to study how the instrument selection procedure performs in finite samples. Some parts of the experimental design follow Donald and Newey (2001). The model considered is

$$y_n = \lambda_0 W_n y_n + \gamma_0 Z_{2n} + \epsilon_n, \qquad Z_{2n} = X_n \beta_0 + U_n,$$

where $\epsilon_n = (\epsilon_{n1}, \cdots, \epsilon_{nn})'$, $U_n = (u_{n1}, \cdots, u_{nn})'$, and $X_n = (X'_{n1}, \cdots, X'_{nn})'$. For $\lambda_0 = 0.6$ and $\gamma_0 = 1$, random samples are generated with $(\epsilon_{ni}, u_{ni})' \sim$ i.i.d.$N(0, \Sigma)$,

$$\Sigma = \begin{pmatrix} 1 & \sigma_{u\epsilon} \\ \sigma_{u\epsilon} & 1 \end{pmatrix},$$

and $X_{ni} \sim N(0, I_{\bar{q}})$, where $\bar{q}$ is the maximal number of instruments for Z_{2n}. As shown in Hahn and Hausman (2002), the specification implies a theoretical first stage R-squared $R_f^2 = \beta_0'\beta_0/(\beta_0'\beta_0 + 1)$.

We consider two models with different specifications of β_0. In Model 1, the kth element of β_0 is given by

$$\beta_{0k} = c(\bar{q})(1 - \frac{k}{\bar{q}+1})^4, \text{ for } k = 1, \cdots, \bar{q},$$

where $c(\bar{q})$ is chosen such that $\beta_0'\beta_0 = R_f^2/(1 - R_f^2)$ for some R_f^2 specified in the experiment. In this model, the $\bar{q}$ instruments are arranged in decreasing order of importance. This specification represents the case where some instruments are more important than others. This case seems relevant to many applications as argued by Donald and Newey (2001). For comparison purpose, we also consider a specification where all instruments in X_n are equally important. The kth element of β_0 is given by

$$\beta_{0k} = \sqrt{\frac{R_f^2}{\bar{q}(1 - R_f^2)}}, \text{ for } k = 1, \cdots, \bar{q},$$

According to Donald and Newey (2001), this specification represents a "worst case" for the instrumental selection procedure, as there is no a priori information to prefer one instrument over another.

We experiment with different specifications for $n \in \{98, 490\}$, $\sigma_{u\epsilon} \in \{0.1, 0.5, 0.9\}$, and $R_f^2 \in \{0.02, 0.1\}$. Let W_A denote the weights matrix for the study of crimes across 49 districts in Columbus, Ohio, in Anselin (1988). For $n = 98$, we set $\bar{q} = 5$ and $W_n = I_2 \otimes W_A$, and conduct 5000 replications. For $n = 490$, we set $\bar{q} = 10$ and $W_n = I_{10} \otimes W_A$, and conduct 1000 replications.

Let X_n^q be a matrix consisting of the first q columns of X_n, and $Q_{p,q} = [X_n^q, W_n X_n^q, \cdots, W_n^p X_n^q]$ for $p = 1, 2, 3$ and $q = 1, \cdots, \bar{q}$. The estimators considered are: (i) 2SLS-min: the 2SLS with $Q_{1,1}$; (ii) 2SLS-max: the 2SLS with $Q_{3,\bar{q}}$; (iii) 2SLS-op: the 2SLS obtained by choosing the (p, q) to minimize $\hat{S}_\xi(K)$ in (5) with $\xi = (1, 1)'$; (iv) C2SLS-max: the C2SLS with $Q_{3,\bar{q}}$; and (v) C2SLS-op: the C2SLS obtained by choosing the (p, q) to minimize $\hat{S}_\xi(K)$ in (6) with $\xi = (1, 1)'$. The initial estimates used in the bias-correction procedure are obtained with $Q_{1,1}$. The approximate MSEs for both 2SLS and C2SLS are estimated using $Q_{3,\bar{q}}$.

For summary statistics of each estimator, we use robust measures of central tendency and dispersion, namely, the median bias (Med. Bias), the median of the absolute deviations (Med. AD), the difference between the 0.1 and 0.9 quantile (Dec. Rge) in the empirical distribution, and the coverage rate (Cov. Rate) of a nominal 95% confidence interval. Let $Z_n = [W_n y_n, Z_{2n}]$ and $\hat{Z}_n = P_K Z_n$. The variance of $\hat{\delta}_n$ is estimated by $\hat{\mathrm{V}}(\hat{\delta}_n) = \frac{1}{n}(y_n - Z_n\hat{\delta}_n)'(y_n - Z_n\hat{\delta}_n)(\hat{Z}_n' Z_n)^{-1}\hat{Z}_n'\hat{Z}_n(Z_n'\hat{Z}_n)^{-1}$.

[Tables 1 and 2 approximately here]

Tables 1 and 2 report the summary statistics for the estimators of Model 1. Relative to the 2SLS-min, instrument selection reduces median absolute deviation and dispersion (interdecile range), and improves coverage probability. When the reduced-form R^2 is low, the 2SLS-op also has smaller median bias than the 2SLS-min. Instrument selection reduces median bias of the 2SLS-max in most cases. With moderate sample size $n = 490$ and $\sigma_{u\epsilon} \in \{0.5, 0.9\}$, instrument selection also improves precision (in terms of median absolute deviation) of the 2SLS-max estimator of γ_0. Coverage probabilities of the 2SLS-op estimators are closer to their nominal levels than the 2SLS-max. The suggested bias-correction procedure substantially reduces the many-instrument bias. For the bias-corrected estimators, instrument selection leads to a reduction in median absolute deviation of the estimator of γ_0 and large improvements in coverage probability.

[Tables 3 and 4 approximately here]

Estimation results of Model 2 are reported in Tables 3 and 4. Choosing the number of instruments tends to raise precision and lower dispersion of the 2SLS-min. Instrument selection also reduces bias and improves coverage probability of the 2SLS-min estimator of λ_0. Instrument selection reduces the bias of the 2SLS-max in about half the cases. However, in terms of the precision and dispersion, the 2SLS-op are generally not as good as the 2SLS-max. This aligns with the finding in Donald and Newey (2001) that when instruments are equally important, instrument selection may be less useful for the 2SLS. Instrument selection improves the precision of the C2SLS estimator of γ_0. For the 2SLS and C2SLS, coverage probabilities are closer to their nominal levels when instruments are chosen.

[Tables 5 and 6 approximately here]

Tables 5 and 6 contain summary statistics for the estimated $\hat{p}$ and $\hat{q}$. For most cases in both models, only the first spatial lag ($p = 1$) is used by the proposed selection procedure. This is consistent with the fact that the approximation error of G_n by a linear combination of $(W_n, \cdots, W_n^p)$ diminishes very fast as p increases. On the other hand, the distribution of $\hat{q}$ is quite sensitive to the specification of β_0. Also, there is a tendency to use a larger q when the reduced-form R_f^2 is large.

[Tables 7 and 8 approximately here]

In the simulation study, we also consider the instrument selection procedure proposed in Donald and Newey (2001) by ignoring the spatial correlation. For a predetermined projector $P_{\bar{K}}$ with $\bar{K}$ instruments, let $\hat{V}_n = (I_n - P_{\bar{K}})Z_n$, $\hat{\sigma}_{v\epsilon} = \frac{1}{n}\hat{\epsilon}_n' \hat{V}_n$, and $\hat{\Sigma}_v = \frac{1}{n}\hat{V}_n'\hat{V}_n$. The approximate MSE of the 2SLS in Donald and Newey (2001) is given by

$$\hat{S}_{\xi\_dn}(K) = \frac{1}{n}\xi' \hat{H}_n^{-1}[K^2 \hat{\sigma}_{v\epsilon}' \hat{\sigma}_{v\epsilon} + \hat{\sigma}_\epsilon^2 (Z_n'(I_n - P_K)Z_n + K\hat{\Sigma}_v)]\hat{H}_n^{-1}\xi,$$

where $\hat{\sigma}_\epsilon^2$ and $\hat{H}_n$ can be estimated by the preliminary estimators described in the previous section. Denote the 2SLS estimator obtained by choosing the (p, q) to minimize $\hat{S}_{\xi\_dn}(K)$ with $\xi = (1,1)'$ by 2SLS-dn. The estimation results of Model 1 and the distribution of the estimated $\hat{p}$ and $\hat{q}$ are reported in Tables 7 and 8 respectively. For the estimation of λ_0, the 2SLS-op generally has smaller median bias but larger median absolute deviation than the 2SLS-dn. This is consistent with the observation that the 2SLS-dn tends to use more instruments. Coverage probabilities are closer to their nominal levels when the approximate MSE explicitly accounts for spatial correlation.

6 Concluding Remarks

This paper derives the approximate MSEs of the 2SLS and C2SLS estimators of the SAR model with endogenous regressors. The approximate MSE can be minimized to choose the instruments as in Donald and Newey (2001). We study the finite sample behavior of the instrument selection procedure in a limited Monte-Carlo experiment. We find that, for the case where some instruments are more important than others, the instrument selection often leads to smaller bias, better precision and more reliable inference.

In a recent paper, Liu and Lee (2010) has considered the estimation of a network model specified as a SAR model. They suggest that a measure for the interaction structure within a group can be used as an instrument for the endogenous social interaction effect (similar to the spatial lag term in the current model). As the number of groups increases in asymptotics, so does the number of the instruments based on the group interaction measures. The results of the current paper can be modified to estimate and/or choose instruments in that social interaction model with many instruments.

7 Summary of Notations

A_{ij} is the (i,j)th element of a matrix A.

For an $n \times n$ matrix A, $A^s = A + A'$, $vec_D(A) = (A_{11}, \cdots, A_{nn})'$, and A^- is a generalized inverse.

e_j is the jth unit (column) vector.

$Z_{2n} = \bar{Z}_{2n} + U_n$, where $\bar{Z}_{2n} = \mathrm{E}(Z_{2n}) = f(X_n)$.

$Z_n = (W_n y_n, Z_{2n}) = F_n + V_n$, where $F_n = [G_n \bar{Z}_{2n}\gamma_0, \bar{Z}_{2n}]$ and $V_n = [G_n(U_n\gamma_0 + \epsilon_n), U_n]$.

$\bar{U}_n = (\bar{u}'_{n1}, \cdots, \bar{u}'_{nn})'$, where $\bar{u}_{ni} = u_{ni}\gamma_0 + \epsilon_{ni}$. $\mathrm{E}(\epsilon^2_{ni}) = \sigma^2_\epsilon$, $\mathrm{E}(u_{ni}\epsilon_{ni}) = \sigma_{u\epsilon}$, and $\mathrm{E}(u'_{ni}u_{ni}) = \Sigma_u$.

$G_n = W_n(I_n - \lambda_0 W_n)^{-1}$, $G_n^{(p)} = (W_n, \cdots, W_n^{p+1})$.

$P_K = Q_K(Q'_K Q_K)^- Q'_K$, $M_K = P_K G_n$.

$e_F(K) = \frac{1}{n}F'_n(I_n - P_K)F_n$, $\Delta_K = \mathrm{tr}(e_F(K))$, $\rho_{K,n} = \mathrm{tr}(S(K))$.

$h_n = F'_n\epsilon_n/\sqrt{n}$, $H_n = \frac{1}{n}F'_nF_n$, $\bar{H} = \lim_{n\to\infty} H_n$.

$\Omega_1(K) = \mathrm{E}(V'_nP_K\epsilon_n)\mathrm{E}(\epsilon'_nP_KV_n)$, $\Omega_2(K) = \mathrm{E}[V'_n(I_n - P_K)V_n]$;

$$\Pi_1(K) = \begin{bmatrix} \mathrm{tr}(M'_KM_K)(\sigma_{u\epsilon}\gamma_0 + \sigma^2_\epsilon)^2 + \mathrm{tr}(M^2_K)\sigma^2_\epsilon(\gamma'_0\Sigma_u\gamma_0 + 2\sigma_{u\epsilon}\gamma_0 + \sigma^2_\epsilon) & * \\ \mathrm{tr}(M_K)[(\sigma_{u\epsilon}\gamma_0 + \sigma^2_\epsilon)\sigma'_{u\epsilon} + \sigma^2_\epsilon(\Sigma_u\gamma_0 + \sigma'_{u\epsilon})] & K(\sigma'_{u\epsilon}\sigma_{u\epsilon} + \sigma^2_\epsilon\Sigma_u) \end{bmatrix},$$

and

$$\Pi_2(K) = \begin{bmatrix} \Pi_{2,11} & * \\ [\frac{K}{n}\mathrm{tr}(G_n) - \mathrm{tr}(M_K)]\sigma_\epsilon^2(\Sigma_u\gamma_0 + \sigma_{u\epsilon}') & 0 \end{bmatrix},$$

where $\Pi_{2,11} = 2[\frac{1}{n}\mathrm{tr}(M_K)\mathrm{tr}(G_n) - \mathrm{tr}(M_K' M_K)]\sigma_\epsilon^2(\gamma_0'\Sigma_u\gamma_0 + 2\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2) + 2[\frac{1}{n}\mathrm{tr}(M_K)\mathrm{tr}(G_n) - \mathrm{tr}(P_K G_n^2)]\sigma_\epsilon^2(\sigma_{u\epsilon}\gamma_0 + \sigma_\epsilon^2)$.

For 2SLS,

$$\begin{aligned} S(K) &= \frac{1}{n}H_n^{-1}[\Omega_1(K) + \sigma_\epsilon^2 F_n'(I_n - P_K)F_n]H_n^{-1}, \\ \hat{S}_\xi(K) &= \frac{1}{n}\xi'\hat{H}_n^{-1}\hat{\Omega}_1\hat{H}_n^{-1}\xi + \frac{1}{n}\hat{\sigma}_\epsilon^2\xi'\hat{H}_n^{-1}[Z_n'(I_n - P_K)Z_n + \hat{\Omega}_2]\hat{H}_n^{-1}\xi. \end{aligned}$$

For C2SLS,

$$\begin{aligned} S(K) &= \frac{1}{n}H_n^{-1}[\Pi_1(K) + \Pi_2(K) + \sigma_\epsilon^2 F_n'(I_n - P_K)F_n]H_n^{-1}, \\ \hat{S}_\xi(K) &= \frac{1}{n}\xi'\hat{H}_n^{-1}(\hat{\Pi}_1 + \hat{\Pi}_2)\hat{H}_n^{-1}\xi + \frac{1}{n}\hat{\sigma}_\epsilon^2\xi'\hat{H}_n^{-1}[Z_n'(I_n - P_K)Z_n + \hat{\Omega}_2]\hat{H}_n^{-1}\xi. \end{aligned}$$

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Bekker, P. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* **62**: 657–681.

Bekker, P. and van der Ploeg, J. (2005). Instrumental variable estimation based on grouped data, *Statistica Neerlandica* **59**: 239–267.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics* **34**: 305–334.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments, *Econometrica* **69**: 1161–1191.

Hahn, J. and Hausman, J. (2002). A new specification test for the validity of instrumental variables, *Econometrica* **70**: 163–189.

Hansen, C., Hausman, J. and Newey, W. (2008). Estimation with many instrumental variables, *Journal of Business and Economic Statistics* **26**: 398–422.

Horn, R. and Johnson, C. (1985). *Matrix Analysis*, Cambridge Univsersity Press, Cambridge.

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

Kelejian, H. H. and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations, *Journal of Econometrics* **118**: 27–50.

Kelejian, H. H. and Prucha, I. R. (2007). Hac estimation in a spatial framework, *Journal of Econometrics* **140**: 131–154.

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

Morimune, K. (1983). Approximate distributions of k-class estimators when the degree of overidentification is large compared with the sample size, *Econometrica* **51**: 821–841.

Nagar, A. L. (1959). The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations, *Econometrica* **27**: 575–595.

Newey, W. (2007). Choosing among instrument sets. Manuscript, Department of Economics, MIT.

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

Table 1: 2SLS Estimation of Model 1, $R_f^2 = 0.02$

		$\lambda_0 = 0.6$	$\gamma_0 = 1.0$
	$n = 98$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.147 (0.506) [3.045] {0.994}	0.027 (0.643) [3.664] {0.999}
	2SLS-max	0.172 (0.175) [0.313] {0.760}	0.004 (0.160) [0.607] {0.960}
	2SLS-op	0.132 (0.269) [1.586] {0.964}	0.063 (0.345) [1.886] {0.989}
	C2SLS-max	0.018 (0.387) [4.240] {0.544}	0.111 (0.871) [5.743] {0.436}
	C2SLS-op	-0.055 (0.755) [12.54] {0.787}	0.169 (1.025) [13.42] {0.820}
0.5	2SLS-min	0.114 (0.421) [2.472] {0.994}	0.183 (0.632) [3.629] {0.980}
	2SLS-max	0.105 (0.112) [0.256] {0.826}	0.380 (0.381) [0.555] {0.531}
	2SLS-op	0.090 (0.187) [1.049] {0.959}	0.357 (0.462) [1.446] {0.795}
	C2SLS-max	0.004 (0.277) [3.458] {0.570}	0.247 (0.826) [5.628] {0.415}
	C2SLS-op	-0.036 (0.659) [13.21] {0.857}	0.270 (1.047) [17.17] {0.811}
0.9	2SLS-min	0.104 (0.339) [2.189] {0.993}	0.285 (0.568) [3.741] {0.885}
	2SLS-max	0.035 (0.044) [0.137] {0.901}	0.785 (0.785) [0.328] {0.004}
	2SLS-op	0.048 (0.107) [0.739] {0.958}	0.667 (0.717) [1.208] {0.446}
	C2SLS-max	-0.018 (0.222) [3.217] {0.555}	0.333 (0.741) [5.963] {0.371}
	C2SLS-op	-0.025 (0.484) [14.49] {0.873}	0.345 (0.868) [17.68] {0.727}
	$n = 490$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.118 (0.383) [2.180] {0.994}	-0.009 (0.377) [1.719] {0.998}
	2SLS-max	0.155 (0.155) [0.207] {0.548}	-0.000 (0.102) [0.389] {0.959}
	2SLS-op	0.119 (0.183) [0.830] {0.939}	0.029 (0.201) [0.931] {0.986}
	C2SLS-max	-0.010 (0.202) [2.708] {0.497}	0.036 (0.402) [2.614] {0.458}
	C2SLS-op	-0.024 (0.306) [4.270] {0.777}	0.089 (0.406) [2.888] {0.822}
0.5	2SLS-min	0.122 (0.337) [1.982] {0.994}	0.027 (0.345) [1.721] {0.971}
	2SLS-max	0.103 (0.103) [0.163] {0.659}	0.323 (0.323) [0.350] {0.335}
	2SLS-op	0.099 (0.150) [0.645] {0.946}	0.206 (0.305) [0.936] {0.766}
	C2SLS-max	-0.002 (0.178) [2.120] {0.489}	0.080 (0.422) [2.576] {0.422}
	C2SLS-op	-0.015 (0.287) [5.323] {0.809}	0.119 (0.399) [3.225] {0.833}
0.9	2SLS-min	0.129 (0.303) [1.939] {0.988}	0.058 (0.334) [1.767] {0.928}
	2SLS-max	0.046 (0.046) [0.091] {0.730}	0.679 (0.679) [0.217] {0.000}
	2SLS-op	0.112 (0.166) [0.664] {0.934}	0.277 (0.341) [0.950] {0.693}
	C2SLS-max	-0.014 (0.195) [3.193] {0.438}	0.107 (0.420) [3.636] {0.420}
	C2SLS-op	-0.017 (0.331) [6.884] {0.804}	0.142 (0.414) [4.668] {0.799}

Med. Bias (Med. AD) [Dec. Rge] {Cov. Rate}

Table 2: 2SLS Estimation of Model 1, $R_f^2 = 0.1$

		$\lambda_0 = 0.6$	$\gamma_0 = 1.0$
	$n = 98$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.094 (0.312) [1.787] {0.990}	-0.018 (0.283) [1.287] {0.991}
	2SLS-max	0.161 (0.163) [0.302] {0.766}	-0.005 (0.134) [0.515] {0.951}
	2SLS-op	0.107 (0.206) [1.034] {0.959}	0.014 (0.213) [0.933] {0.980}
	C2SLS-max	-0.009 (0.198) [2.118] {0.609}	0.020 (0.341) [1.849] {0.656}
	C2SLS-op	-0.028 (0.247) [2.630] {0.755}	0.052 (0.322) [1.832] {0.853}
0.5	2SLS-min	0.105 (0.293) [1.716] {0.989}	-0.005 (0.285) [1.351] {0.968}
	2SLS-max	0.114 (0.119) [0.249] {0.802}	0.255 (0.258) [0.468] {0.656}
	2SLS-op	0.107 (0.184) [0.913] {0.950}	0.130 (0.263) [0.953] {0.864}
	C2SLS-max	-0.003 (0.194) [2.260] {0.575}	0.032 (0.354) [2.215] {0.639}
	C2SLS-op	-0.004 (0.245) [3.906] {0.802}	0.074 (0.335) [2.475] {0.857}
0.9	2SLS-min	0.107 (0.276) [1.648] {0.986}	0.002 (0.271) [1.427] {0.936}
	2SLS-max	0.068 (0.071) [0.161] {0.807}	0.527 (0.527) [0.338] {0.066}
	2SLS-op	0.104 (0.199) [1.087] {0.948}	0.151 (0.287) [1.127] {0.809}
	C2SLS-max	-0.002 (0.197) [2.544] {0.547}	0.032 (0.353) [2.642] {0.622}
	C2SLS-op	0.005 (0.268) [6.076] {0.836}	0.077 (0.354) [3.843] {0.839}
	$n = 490$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.026 (0.173) [0.723] {0.979}	-0.014 (0.135) [0.564] {0.979}
	2SLS-max	0.130 (0.130) [0.196] {0.601}	-0.004 (0.075) [0.286] {0.951}
	2SLS-op	0.059 (0.109) [0.407] {0.935}	0.004 (0.099) [0.388] {0.951}
	C2SLS-max	-0.022 (0.080) [0.404] {0.738}	0.009 (0.129) [0.547] {0.762}
	C2SLS-op	-0.005 (0.083) [0.394] {0.830}	0.019 (0.111) [0.478] {0.848}
0.5	2SLS-min	0.032 (0.172) [0.865] {0.989}	-0.026 (0.141) [0.597] {0.972}
	2SLS-max	0.105 (0.105) [0.159] {0.611}	0.160 (0.160) [0.261] {0.633}
	2SLS-op	0.063 (0.113) [0.439] {0.946}	0.033 (0.110) [0.407] {0.925}
	C2SLS-max	-0.019 (0.083) [0.514] {0.676}	0.009 (0.137) [0.591] {0.752}
	C2SLS-op	0.005 (0.091) [0.459] {0.838}	0.019 (0.119) [0.456] {0.888}
0.9	2SLS-min	0.047 (0.165) [0.998] {0.990}	-0.017 (0.143) [0.631] {0.960}
	2SLS-max	0.083 (0.083) [0.110] {0.541}	0.327 (0.327) [0.194] {0.065}
	2SLS-op	0.059 (0.120) [0.529] {0.957}	0.037 (0.115) [0.438] {0.910}
	C2SLS-max	-0.016 (0.083) [0.787] {0.627}	0.010 (0.143) [0.731] {0.713}
	C2SLS-op	0.005 (0.103) [0.800] {0.829}	0.017 (0.122) [0.498] {0.893}

Med. Bias (Med. AD) [Dec. Rge] {Cov. Rate}

Table 3: 2SLS Estimation of Model 2, $R_f^2 = 0.02$

		$\lambda_0 = 0.6$	$\gamma_0 = 1.0$
	$n = 98$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.139 (0.570) [3.630] {0.995}	0.082 (0.900) [5.313] {1.000}
	2SLS-max	0.171 (0.174) [0.317] {0.759}	0.007 (0.161) [0.615] {0.957}
	2SLS-op	0.131 (0.271) [1.641] {0.959}	0.075 (0.367) [2.413] {0.989}
	C2SLS-max	0.024 (0.452) [5.417] {0.544}	0.155 (1.126) [7.771] {0.368}
	C2SLS-op	-0.055 (0.963) [15.90] {0.767}	0.238 (1.494) [21.11] {0.759}
0.5	2SLS-min	0.092 (0.437) [2.627] {0.997}	0.393 (0.892) [4.724] {0.987}
	2SLS-max	0.104 (0.111) [0.254] {0.830}	0.382 (0.383) [0.549] {0.531}
	2SLS-op	0.084 (0.178) [1.048] {0.958}	0.417 (0.507) [1.662] {0.778}
	C2SLS-max	-0.001 (0.293) [3.691] {0.589}	0.457 (1.063) [6.829] {0.324}
	C2SLS-op	-0.059 (0.689) [14.54] {0.850}	0.515 (1.395) [21.67] {0.754}
0.9	2SLS-min	0.041 (0.264) [1.682] {0.998}	0.658 (0.843) [3.599] {0.830}
	2SLS-max	0.034 (0.044) [0.136] {0.902}	0.784 (0.784) [0.331] {0.004}
	2SLS-op	0.031 (0.091) [0.574] {0.966}	0.781 (0.801) [1.057] {0.374}
	C2SLS-max	-0.009 (0.120) [2.035] {0.648}	0.654 (0.930) [4.895] {0.248}
	C2SLS-op	-0.027 (0.311) [10.11] {0.888}	0.699 (1.084) [15.67] {0.599}
	$n = 490$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.159 (0.517) [2.731] {0.997}	0.048 (0.723) [3.957] {0.998}
	2SLS-max	0.155 (0.155) [0.209] {0.542}	0.002 (0.104) [0.406] {0.961}
	2SLS-op	0.140 (0.205) [1.159] {0.939}	0.069 (0.254) [1.585] {0.987}
	C2SLS-max	0.025 (0.302) [3.655] {0.462}	0.121 (0.770) [4.856] {0.295}
	C2SLS-op	-0.009 (0.667) [9.910] {0.736}	0.186 (0.871) [10.54] {0.719}
0.5	2SLS-min	0.097 (0.391) [2.578] {0.999}	0.266 (0.693) [4.337] {0.984}
	2SLS-max	0.104 (0.104) [0.161] {0.667}	0.322 (0.322) [0.351] {0.341}
	2SLS-op	0.092 (0.145) [0.767] {0.948}	0.346 (0.429) [1.268] {0.720}
	C2SLS-max	0.003 (0.207) [3.091] {0.498}	0.307 (0.783) [5.109] {0.265}
	C2SLS-op	-0.038 (0.554) [10.55] {0.799}	0.379 (0.950) [12.34] {0.711}
0.9	2SLS-min	0.082 (0.307) [1.941] {0.997}	0.429 (0.639) [3.926] {0.866}
	2SLS-max	0.046 (0.046) [0.091] {0.723}	0.675 (0.675) [0.224] {0.000}
	2SLS-op	0.073 (0.143) [0.922] {0.968}	0.588 (0.640) [1.570] {0.553}
	C2SLS-max	-0.007 (0.132) [2.269] {0.492}	0.402 (0.651) [4.522] {0.216}
	C2SLS-op	-0.003 (0.326) [10.25] {0.836}	0.438 (0.796) [12.25] {0.633}

Med. Bias (Med. AD) [Dec. Rge] {Cov. Rate}

Table 4: 2SLS Estimation of Model 2, $R_f^2 = 0.1$

		$\lambda_0 = 0.6$	$\gamma_0 = 1.0$
	$n = 98$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.135 (0.481) [2.912] {0.994}	0.012 (0.576) [3.238] {0.999}
	2SLS-max	0.159 (0.162) [0.301] {0.769}	0.001 (0.135) [0.515] {0.951}
	2SLS-op	0.121 (0.212) [1.175] {0.946}	0.051 (0.256) [1.339] {0.977}
	C2SLS-max	0.032 (0.310) [3.464] {0.545}	0.061 (0.580) [3.910] {0.479}
	C2SLS-op	0.011 (0.394) [4.821] {0.667}	0.069 (0.556) [4.267] {0.690}
0.5	2SLS-min	0.116 (0.410) [2.455] {0.994}	0.138 (0.571) [3.154] {0.982}
	2SLS-max	0.113 (0.118) [0.246] {0.805}	0.259 (0.263) [0.466] {0.659}
	2SLS-op	0.100 (0.171) [0.889] {0.948}	0.252 (0.346) [1.153] {0.826}
	C2SLS-max	0.012 (0.240) [2.996] {0.554}	0.142 (0.573) [3.895] {0.459}
	C2SLS-op	0.005 (0.325) [4.816] {0.722}	0.188 (0.573) [4.892] {0.695}
0.9	2SLS-min	0.109 (0.342) [2.237] {0.993}	0.233 (0.527) [3.328] {0.903}
	2SLS-max	0.067 (0.070) [0.157] {0.810}	0.527 (0.527) [0.335] {0.065}
	2SLS-op	0.083 (0.177) [1.178] {0.963}	0.408 (0.494) [1.495] {0.668}
	C2SLS-max	0.008 (0.199) [2.943] {0.540}	0.193 (0.516) [4.346] {0.444}
	C2SLS-op	0.012 (0.291) [5.994] {0.763}	0.241 (0.554) [6.642] {0.670}
	$n = 490$		
$\sigma_{u\epsilon} = 0.1$	2SLS-min	0.125 (0.387) [2.370] {0.996}	-0.007 (0.397) [1.803] {0.997}
	2SLS-max	0.133 (0.133) [0.185] {0.574}	-0.004 (0.074) [0.281] {0.956}
	2SLS-op	0.110 (0.151) [0.598] {0.900}	0.019 (0.134) [0.680] {0.965}
	C2SLS-max	0.018 (0.163) [1.858] {0.496}	0.024 (0.246) [1.340] {0.512}
	C2SLS-op	0.020 (0.170) [1.632] {0.548}	0.030 (0.209) [1.014] {0.628}
0.5	2SLS-min	0.131 (0.341) [2.042] {0.995}	0.031 (0.377) [1.797] {0.972}
	2SLS-max	0.110 (0.110) [0.150] {0.592}	0.161 (0.162) [0.258] {0.626}
	2SLS-op	0.106 (0.151) [0.696] {0.930}	0.123 (0.220) [0.897] {0.874}
	C2SLS-max	0.006 (0.148) [1.923] {0.475}	0.047 (0.240) [1.541] {0.515}
	C2SLS-op	0.013 (0.162) [1.886] {0.565}	0.050 (0.201) [1.238] {0.656}
0.9	2SLS-min	0.138 (0.300) [1.989] {0.990}	0.067 (0.348) [1.806] {0.933}
	2SLS-max	0.082 (0.082) [0.108] {0.540}	0.326 (0.326) [0.194] {0.059}
	2SLS-op	0.130 (0.220) [1.287] {0.969}	0.178 (0.306) [1.270] {0.846}
	C2SLS-max	-0.006 (0.159) [2.230] {0.452}	0.063 (0.233) [1.901] {0.493}
	C2SLS-op	0.016 (0.160) [2.104] {0.597}	0.080 (0.206) [1.618] {0.679}

Med. Bias (Med. AD) [Dec. Rge] {Cov. Rate}

Table 5: The Distributions of $\hat{p}$ and $\hat{q}$ in Model 1

		2sls		c2sls	
		$\hat{p}$	$\hat{q}$	$\hat{p}$	$\hat{q}$
		$n = 98$			
$R_f^2 = 0.02$	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 2, 4]
	0.5	1 [1, 1, 2]	1 [1, 1, 3]	1 [1, 1, 2]	1 [1, 1, 2]
	0.9	1 [1, 1, 2]	1 [1, 1, 3]	1 [1, 1, 2]	1 [1, 1, 2]
0.1	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 1, 3]	1 [1, 3, 5]
	0.5	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 2, 3]
	0.9	1 [1, 1, 1]	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 1, 2]
		$n = 490$			
$R_f^2 = 0.02$	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 2]	1 [1, 2, 4]	1 [1, 1, 2]	1 [1, 2, 4]
	0.5	1 [1, 1, 2]	1 [1, 2, 4]	1 [1, 1, 2]	1 [1, 2, 3]
	0.9	1 [1, 1, 1]	1 [1, 2, 2]	1 [1, 1, 1]	1 [1, 2, 3]
0.1	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 1]	2 [2, 3, 4]	1 [1, 2, 3]	3 [3, 5, 9]
	0.5	1 [1, 1, 1]	2 [2, 2, 3]	1 [1, 1, 2]	3 [3, 4, 6]
	0.9	1 [1, 1, 1]	2 [1, 2, 2]	1 [1, 1, 1]	2 [2, 3, 4]

Mode [L.Q., Med., U.Q.]

Table 6: The Distributions of $\hat{p}$ and $\hat{q}$ in Model 2

		2sls		c2sls	
		$\hat{p}$	$\hat{q}$	$\hat{p}$	$\hat{q}$
		$n = 98$			
$R_f^2 = 0.02$	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 2]	1 [1, 1, 3]	1 [1, 1, 2]	1 [1, 2, 4]
	0.5	1 [1, 1, 2]	1 [1, 2, 4]	1 [1, 1, 2]	1 [1, 1, 3]
	0.9	1 [1, 1, 2]	1 [1, 2, 3]	1 [1, 1, 2]	1 [1, 1, 3]
0.1	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 2]	1 [1, 2, 4]	1 [1, 1, 3]	5 [2, 4, 5]
	0.5	1 [1, 1, 2]	1 [1, 2, 4]	1 [1, 1, 2]	5 [2, 3, 5]
	0.9	1 [1, 1, 1]	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 3, 4]
		$n = 490$			
$R_f^2 = 0.02$	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 2]	1 [1, 2, 6]	1 [1, 1, 2]	1 [1, 3, 7]
	0.5	1 [1, 1, 2]	1 [1, 2, 6]	1 [1, 1, 1]	1 [1, 2, 5]
	0.9	1 [1, 1, 1]	1 [1, 1, 2]	1 [1, 1, 1]	1 [1, 1, 4]
0.1	$\sigma_{u\epsilon} = 0.1$	1 [1, 1, 1]	1 [1, 4, 9]	1 [1, 1, 3]	10 [9,10,10]
	0.5	1 [1, 1, 1]	1 [1, 2, 4]	1 [1, 1, 1]	10 [8,10,10]
	0.9	1 [1, 1, 1]	1 [1, 1, 2]	1 [1, 1, 1]	10 [6, 9,10]

Mode [L.Q., Med., U.Q.]

Table 7: 2SLS Estimation of Model 1 with Selected IVs based on $\hat{S}_{\xi\_dn}(K)$

			$\lambda_0 = 0.6$	$\gamma_0 = 1.0$
		$n = 98$		
$R_f^2 = 0.02$	$\sigma_{u\epsilon} = 0.1$	2SLS-dn	0.150 (0.207) [0.715] {0.911}	0.033 (0.220) [1.096] {0.987}
	0.5	2SLS-dn	0.097 (0.156) [0.654] {0.940}	0.373 (0.434) [1.113] {0.723}
	0.9	2SLS-dn	0.050 (0.098) [0.611] {0.955}	0.689 (0.734) [1.156] {0.402}
		$n = 490$		
	$\sigma_{u\epsilon} = 0.1$	2SLS-dn	0.140 (0.157) [0.394] {0.813}	0.018 (0.130) [0.564] {0.979}
	0.5	2SLS-dn	0.099 (0.126) [0.387] {0.907}	0.254 (0.304) [0.724] {0.678}
	0.9	2SLS-dn	0.110 (0.156) [0.579] {0.928}	0.297 (0.356) [0.917] {0.651}
		$n = 98$		
$R_f^2 = 0.1$	$\sigma_{u\epsilon} = 0.1$	2SLS-dn	0.142 (0.170) [0.474] {0.886}	0.005 (0.157) [0.659] {0.974}
	0.5	2SLS-dn	0.118 (0.146) [0.483] {0.910}	0.190 (0.270) [0.795] {0.772}
	0.9	2SLS-dn	0.114 (0.166) [0.777] {0.934}	0.197 (0.308) [1.045] {0.732}
		$n = 490$		
	$\sigma_{u\epsilon} = 0.1$	2SLS-dn	0.117 (0.121) [0.241] {0.741}	0.002 (0.079) [0.309] {0.956}
	0.5	2SLS-dn	0.093 (0.109) [0.300] {0.889}	0.053 (0.106) [0.382] {0.867}
	0.9	2SLS-dn	0.078 (0.114) [0.391] {0.938}	0.047 (0.108) [0.386] {0.879}

Med. Bias (Med. AD) [Dec. Rge] {Cov. Rate}

Table 8: The Distributions of $\hat{p}$ and $\hat{q}$ Selected based on $\hat{S}_{\xi\_dn}(K)$

		$n = 98$		$n = 490$	
		$\hat{p}$	$\hat{q}$	$\hat{p}$	$\hat{q}$
$R_f^2 = 0.02$	$\sigma_{u\epsilon} = 0.1$	1 [1, 2, 3]	1 [1, 3, 4]	3 [1, 2, 3]	10 [2, 5, 8]
	0.5	1 [1, 2, 3]	1 [1, 2, 4]	1 [1, 1, 3]	2 [2, 3, 7]
	0.9	1 [1, 1, 2]	1 [1, 2, 3]	1 [1, 1, 1]	1 [1, 2, 2]
0.1	$\sigma_{u\epsilon} = 0.1$	3 [1, 2, 3]	5 [2, 3, 5]	3 [2, 3, 3]	10 [5, 7, 9]
	0.5	1 [1, 2, 3]	1 [1, 2, 4]	1 [1, 1, 2]	3 [3, 3, 5]
	0.9	1 [1, 1, 2]	1 [1, 1, 2]	1 [1, 1, 1]	2 [2, 2, 2]

Mode [L.Q., Med., U.Q.]