# Irregular N2SLS and LASSO estimation of the matrix exponential spatial specification model

Fei Jin[a,b] and Lung-fei Lee[c*]

[a]School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, China

[b]Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China

[c]Department of Economics, The Ohio State University, Columbus, OH 43210, USA

June 2, 2016

**Abstract**

In this paper, we consider estimation of the matrix exponential spatial specification model with the Durbin and endogenous regressors. We find that the nonlinear two-stage least squares (N2SLS) estimator is in general consistent and asymptotically normal. However, when the Durbin and endogenous regressors are irrelevant, the gradient vector of the N2SLS criterion function has a singular covariance matrix with probability approaching one (w.p.a.1.). Some components of the N2SLS estimator have slower rates of convergence and their asymptotic distributions are nonstandard. The distance difference and gradient test statistics are derived to test for the irrelevance of the Durbin and endogenous regressors. As an alternative estimation and model selection approach, we propose the adaptive group LASSO, which penalizes the coefficients of the Durbin and endogenous explanatory variables. We show that the estimator has the oracle properties, so the true model can be selected w.p.a.1. and the estimator always has the $\sqrt{n}$-rate of convergence and asymptotic normal distribution. We propose to select the tuning parameter in the adaptive group LASSO estimation by minimizing an information criterion.

*Keywords:* matrix exponential spatial specification, singular covariance matrix, nonlinear two-stage least squares, LASSO, oracle properties

*JEL classification:* C12, C13, C21, R15

## 1 Introduction

The spatial autoregressive (SAR) model is a popular model in spatial econometrics.[1] As an alternative model with spatial dependence, LeSage and Pace (2007) propose the matrix exponential spatial specification (MESS). The

*Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 4192. E-mail addresses: jin.fei@sufe.edu.cn (F. Jin), lee.1777@osu.edu (L.-F. Lee).

[1]See, e.g., Anselin (1988), Kelejian and Prucha (1998), and Lee (2004).

MESS model may provide estimates and inference similar to those from the SAR model. Han and Lee (2013) find that the SAR and MESS models may not be easily tested against each other unless the spatial interaction is rather strong. But the quasi-maximum likelihood (QML) estimator of the MESS model is much easier to compute than that of the SAR model, in particular for models with high order spatial weights, since it does not involve computing the determinant of the Jacobian transformation matrix. In addition, there are no constraints on the parameter that captures spatial dependence because the reduced form of the MESS model always exists. Debarsy et al. (2015) find that the QML estimator of the MESS model is robust to unknown heteroskedasticity, which is a nice property not shared by the SAR model. These features may lead to wider applications of the MESS model when spatial interactions are stable.

LeSage and Pace (2007) present the maximum likelihood (ML) and Bayesian estimators of the MESS model. Debarsy et al. (2015) consider large sample properties of the QML and generalized method of moments (GMM) estimators. However, those researchers have not included endogenous regressors in the model. In addition, Durbin regressors $W_nX_n$, i.e., spatial lags of exogenous variables, where $X_n$ is a matrix of exogenous variables and $W_n$ is an $n \times n$ spatial weights matrix, may be included in the model to capture local spillovers (externalities) in exogenous variables, while both the SAR and MESS processes may capture global spatial interactions (Anselin, 2003). In the social interaction literature, the Durbin regressors are referred to as contextual effects and the global spatial dependence is called endogenous effects, reflecting the contemporaneous and reciprocal influences of peers (Manski, 1993; Brock and Durlauf, 2001).

In this paper, we consider estimation of the MESS model with both endogenous and Durbin's regressors. Our model allows for unknown heteroskedasticity and unknown spatial correlation in the disturbances, so we estimate it with the nonlinear two-stage least squares (N2SLS). The N2SLS estimation can be seen as a GMM estimation exploring only linear moments. So it is of special interest as other estimation methods such as the QML and GMM with quadratic moments would not be appropriate due to unknown pattern of spatial correlation in disturbances. In the spatial econometric literature, while inclusion of the Durbin regressors in an SAR model may have enriched economic content in terms of capturing exogenous externality effects and relaxed restrictions imposed on direct and indirect spatial effects by the SAR model (Elhorst, 2010), their presence as extra exogenous regressors does not induce conceptual econometric issues in a 2SLS estimation procedure if columns of $W_nX_n$ are linearly independent with columns of $X_n$.[2] They can simply be treated as exogenous regressors in the estimation of an SAR model from a methodological point of view. Nonetheless, the presence of Durbin's regressors in the MESS model creates an issue for the N2SLS estimation.

We show that the parameters of the model are, in general, identifiable and the N2SLS estimator can be $\sqrt{n}$-consistent and asymptotically normal. However, when the coefficients of the Durbin and endogenous regressors are zero (but unknown), even though parameters of the model are still identifiable and the N2SLS estimator is consistent, elements of the gradient vector of the N2SLS criterion function at the true parameter values are linearly dependent

[2] If $W_n$ is row-normalized and $X_n$ contains an intercept term, a column of ones should be deleted from $[X_n, W_nX_n]$ to avoid multicollinearity.

with probability approaching one (w.p.a.1.). This implies that the covariance matrix of the gradient vector at the true parameters is singular w.p.a.1. In fact, such an irregular phenomenon appears in a reduced MESS model with Durbin regressors but without endogenous explanatory variables, where the Durbin regressors are really irrelevant. This corresponds to the singular information matrix phenomenon in the likelihood framework.[3] Some authors have studied the asymptotic distributions of ML estimators (MLE) for parametric models with singular information matrices. Cox and Hinkley (1974) provide two examples where the score statistic is zero, and show that the asymptotic distribution of the estimators can be found by a reparameterization. Lee (1993) derives the asymptotic distribution of the MLE for parameters in a stochastic frontier function model with a singular information matrix. Rotnitzky et al. (2000) investigate a more general setting where an identifiable parametric model has a singular information matrix of rank being one less than the number of parameters. The methods in both Lee (1993) and Rotnitzky et al. (2000) involve reparameterizations and high order Taylor expansions of the first order conditions for the MLE. Dovonon and Renault (2013) derive the convergence rate of the GMM estimator and the nonstandard asymptotic distribution of the $J$-test statistic for overidentification.

Following Rotnitzky et al. (2000), by a reparameterization, we derive the asymptotic distribution of the N2SLS estimator for our MESS model where the elements of the gradient vector of the N2SLS criterion function are linearly dependent w.p.a.1. The asymptotic distribution is non-standard, and only the parameter estimators for the exogenous and endogenous variables have the $\sqrt{n}$-rate of convergence, while the spatial dependence parameter estimator and those for the Durbin regressors have the $n^{1/4}$-rate of convergence.[4] The model we consider is one with spatially correlated data and the elements of the gradient vector of the criterion function for the N2SLS estimation can be linearly dependent w.p.a.1. For such a situation, reparameterization and high order Taylor expansions of the first order conditions can still be employed to derive asymptotic distributions of the N2SLS estimators, as for the case with i.i.d. data in Rotnitzky et al. (2000).

Since the Durbin and endogenous regressors may lead to nonstandard asymptotic distribution of the N2SLS estimator for the MESS model, it is of interest to test whether they are relevant or not. The classical tests in the GMM framework, such as the Wald test, the gradient test and the distance difference test, are derived when elements of a gradient vector are not linearly dependent. We show that, even when elements of the gradient vector are linearly dependent, we can still derive the distance difference and the gradient test statistics, but they have nonstandard asymptotic distributions. The asymptotic distribution of the distance difference test statistic is a mixture of two chi-squared distributions, with the number of degrees of freedom equal to $p$ and $p-1$ respectively and with mixing probabilities equal to $1/2$, where $p$ is the number of restrictions. The gradient test statistic, constructed using the Moore-Penrose pseudoinverse due to the singular covariance matrix of the gradient vector,

[3]For various cases of singular information matrices or zero score statistics in the likelihood framework, see, among others, Silvey (1959), Cox and Hinkley (1974), Kiefer (1982), Waldman (1982), Schmidt and Lin (1984), Lee and Chesher (1986), and Sargan (1983).

[4]The asymptotic distribution derived in this paper is non-standard due to the necessary high order expansion ended at an even order, which is the feature in Rotnitzky et al. (2000). However, in Lee (1993) for the stochastic frontier model (as well as a sample selection model), it has a high order expansion at an odd order from which asymptotic truncated-normal (normal) distribution can be derived. The extra complication for a high order expansion ended at an even order is the need to determine the sign of an estimator.

is asymptotically distributed as a chi-squared distribution with $p-1$ degrees of freedom. We also investigate the local power properties of our tests. For the Pitman drift (McManus, 1991) with the rate $n^{1/2}$, there is a direction of the parameter drift for which the tests have trivial power.

As an alternative estimation and model selection method, we propose to estimate the model based on the LASSO, which can perform model selection and parameter estimation simultaneously. We note that, in the case that the parameter vector $\zeta$ of the Durbin and endogenous regressors has the true value $\zeta_0 = 0$, if the information of $\zeta_0 = 0$ is used, then the irregular phenomenon disappears and the N2SLS estimator has the usual $\sqrt{n}$-rate of convergence and asymptotic normal distribution. Since the cause of the irregular phenomenon is $\zeta_0 = 0$, we can use the adaptive group LASSO (AGLASSO) that appends a penalty function of $\zeta$ to the N2SLS criterion function (Yuan and Lin, 2006; Wang and Leng, 2008). We show that the AGLASSO has the oracle properties, i.e., it can select the correct model w.p.a.1 and the resulted estimator satisfies the properties as if we knew the true model (Fan and Li, 2001; Zou, 2006). As a result, there is always no irregular phenomenon in the AGLASSO estimation, and the AGLASSO estimator has the $\sqrt{n}$-rate of convergence and asymptotic normal distribution.

The AGLASSO involves a tuning parameter. The oracle properties are satisfied when the tuning parameter has certain order in asymptotic analysis. But in finite samples, it is not clear what tuning parameter should be used in order that the AGLASSO can perform well. We select the tuning parameter by minimizing an information criterion for our AGLASSO. We show that the proposed data-driven procedure can identify the true model consistently. Due to the irregular phenomenon of the MESS model, the proposed information criterion differs from traditional ones.[5]

The rest of this paper is organized as follows. In Section 2, we introduce the MESS model with Durbin's regressors and endogenous explanatory variables, and show consistency and asymptotic distributions of the N2SLS estimators in the regular and irregular cases. In Section 3, we derive the distance difference and gradient tests and investigate their local power properties. In Section 4, we consider the AGLASSO estimation of the MESS model. In Section 5, we present some Monte Carlo results. We conclude in Section 6. All lemmas and proofs are collected in an online supplementary file.

## 2 N2SLS estimator

In this section, we consider the N2SLS estimation of the MESS model with the Durbin and endogenous explanatory variables. The model is as follows:

$$e^{\alpha W_n} Y_n = X_n^* \beta_1 + W_n l_n \beta_2 + W_n X_{n1} \beta_3 + Z_n \beta_4 + V_n, \quad V_n = T_n \epsilon_n, \tag{1}$$

where $n$ is the sample size, $Y_n$ is an $n \times 1$ vector of observations on the dependent variable, $l_n$ is a vector of ones, $X_{n1}$ is an $n \times (k_x - 1)$ matrix of exogenous variables that does not contain an intercept term, $Z_n$ is an $n \times k_z$ matrix of endogenous variables, $\epsilon_n$ is an $n \times 1$ vector of innovations with mean zero and covariance matrix

[5]For properties of various information criteria, see, among others, Wang et al. (2007), Wang and Leng (2007), Wang et al. (2009), Zhang et al. (2010), and Liao (2010).

being an identity matrix, $T_n$ is an $n \times n$ nonstochastic matrix whose elements are unknown to allow for unknown heteroskedasticity and spatial correlation in the disturbance vector $V_n$ as in Kelejian and Prucha (2007), and $W_n$ is an $n \times n$ spatial weights matrix with all diagonal elements being zero. The spatial weights matrix $W_n$ can be row-normalized or not row-normalized. When $W_n$ is not row-normalized, $X_n^* = X_n$ with $X_n = [X_{n1}, l_n]$; when $W_n$ is row-normalized, as $W_n l_n = l_n$, $X_n^* = X_{n1}$ and the intercept term is written as $W_n l_n$ for the convenience of later analyses. The Durbin regressors $W_n X_{n1}$ and $W_n l_n$ when $W_n$ is not row-normalized can be seen as neighbors' characteristics to capture exogenous externality. The $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_4$ are conformable parameter vectors. The matrix exponential $e^{\alpha W_n}$ with a scalar parameter $\alpha$, which captures spatial dependence, is defined as $\sum_{i=0}^{\infty} \frac{\alpha^i}{i!} W_n^i$. Since the inverse of $e^{\alpha W_n}$ always exists and equals $e^{-\alpha W_n}$ (Chiu et al., 1996), the reduced form of the model always exists and no constraints need to be imposed on the parameter space of $\alpha$. If the model is regarded as a game with complete information, the Nash equilibrium exists and is unique. For an SAR model, $e^{\alpha W_n} Y_n$ in (1) is replaced by $(I_n - \lambda W_n)$ for some scalar $\lambda$. Since $(I_n - \lambda W_n)^{-1} = \sum_{i=0}^{\infty} \lambda^i W_n^i$ if $\|\lambda W_n\| < 1$ for some matrix norm $\|\cdot\|$, the SAR model shows a geometrical decay of spatial dependence, while the MESS model shows an exponential decay. Furthermore, $|e^{\alpha W_n}| = e^{\alpha\, \mathrm{tr}(W_n)} = 1$ as the diagonal elements of $W_n$ are all zero, where $|\cdot|$ denotes the determinant of a square matrix. Thus, the likelihood function of the MESS model does not involve any determinant of the Jacobian transformation matrix and it has a computational advantage over the SAR model (LeSage and Pace, 2007).

Let $D_n = [X_n^*, W_n l_n, W_n X_{n1}, Z_n]$, $\beta = (\beta_1', \beta_2, \beta_3', \beta_4')'$, $\theta = (\alpha, \beta')'$, and $F_n$ be a full rank $n \times k_f$ instrumental variable (IV) matrix with $k_f$ not smaller than the total number of coefficients. As the variance of $F_n' V_n$ conditional on $F_n$ is $\Pi_n = F_n' T_n T_n' F_n$, the criterion function $Q_n(\theta)$ of the infeasible N2SLS estimation, as if $T_n$ were known, is

$$Q_n(\theta) = (e^{\alpha W_n} Y_n - D_n \beta)' F_n \Pi_n^{-1} F_n' (e^{\alpha W_n} Y_n - D_n \beta). \tag{2}$$

To focus on the N2SLS estimation, we first consider the large sample properties of the infeasible N2SLS estimator $\check{\theta}_n$ that minimizes $Q_n(\theta)$. A feasible version by the use of a nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator in Kelejian and Prucha (2007) will be investigated in the last part of this section, which shows that asymptotic results remain valid for the feasible version.

We first discuss some regularity conditions needed for the N2SLS estimation. Let "BRC" stand for "bounded in both row and column sum norms". A typical assumption on the spatial weights matrix $W_n$ in spatial econometrics is that it is BRC. This assumption, originated in Kelejian and Prucha (1998, 1999), restricts the degree of spatial dependence. Similarly, the unknown $T_n$ can be assumed to be BRC to restrict the degree of spatial dependence in the disturbances as in Kelejian and Prucha (2007) for the setting of disturbances with general unknown heteroskedasticity and spatial correlation. For the exogenous variables matrix $X_{n1}$, to allow for stochastic regressors with unknown spatial correlation, we may assume that $X_{n1} = B_{n1} \xi_{n1}$, where $B_{n1}$ is an $n \times n$ unknown BRC nonstochastic matrix and rows of $\xi_{n1}$ are i.i.d. As in Kelejian and Prucha (2004), (1) can represent an equation in a system of spatially correlated equations, then $Z_n = A_{n1} \mathbb{X}_n \gamma + A_{n2} u_n$, where $\mathbb{X}_n$ denotes all the exogenous variables in the system, $\gamma$ is a parameter matrix, $A_{n1}$ and $A_{n2}$ are $n \times n$ BRC nonstochastic matrices, and rows of $u_n$ are i.i.d. The matrix $A_{n1}$ depends on the spatial weights matrix $W_n$. Under regularity conditions, $A_{n1} = \sum_{i=1}^{\infty} \rho_i W_n^i$ for scalars $\rho_i$'s. Then the IV matrix $F_n$ in the N2SLS estimation can be formed by the independent columns of $[\mathbb{X}_n, W_n\mathbb{X}_n, \ldots, W_n^s\mathbb{X}_n]$ for some $s$. As $\mathbb{X}_n$ may also be the product of a BRC nonstochastic matrix and a stochastic matrix with i.i.d. rows, it is plausible to assume that $Z_n = B_{n2}\xi_{n2} + B_{n3}\xi_{n3}$ and $F_n = B_{n4}\xi_{n4}$, where $B_{n2}$, $B_{n3}$ and $B_{n4}$ are $n \times n$ unknown BRC nonstochastic matrices, and rows of stochastic matrices $\xi_{n2}$, $\xi_{n3}$ and $\xi_{n4}$ are i.i.d. One or more elements of $Z_n$ can also be generated by a nonlinear model.[6] With underlying exogenous variables similar to $X_{n1}$ satisfying $X_{n1} = B_{n1}\xi_{n1}$, $Z_n$ and $F_n$ can also have the forms discussed above.[7] Let $\epsilon_{ni}$ be the $i$th element of $\epsilon_n$, and $\xi_{jn,i}$'s be the $i$th rows of $\xi_{jn}$ for $j = 1, \ldots, 4$. We need the existence of the fourth moment of $\|(\xi_{n1,i}, \xi_{n2,i}, \xi_{n3,i})\|_\infty$ for the applicability of a law of large numbers, and the existence of a moment of $\|(\xi_{n4,i}, \epsilon_{ni})\|_\infty$ higher than the four order for the applicability of a central limit theorem and for proving that the HAC estimator for the covariance matrix $\frac{1}{n}\Pi_n$ is consistent. To formulate the N2SLS estimation, $\lim_{n\to\infty} \frac{1}{n} \mathrm{E}(\Pi_n)$ is assumed to be nonsingular. Although the reduced form of the model always exists for any value of $\alpha$, we assume that $\alpha$ is bounded as in Debarsy et al. (2015). As a consequence, $e^{\alpha W_n}$ is BRC uniformly in $\alpha$.[8] The basic assumptions are summarized below.

**Assumption 1.** *The nonstochastic matrices $\{W_n\}$ are BRC, and their diagonal elements are all zero.*

**Assumption 2.** *The unknown nonstochastic matrices $\{T_n\}$ are nonsingular, $\{T_n\}$ and $\{T_n^{-1}\}$ are BRC, and $\lim_{n\to\infty} \frac{1}{n} \mathrm{E}(\Pi_n)$ exists and is nonsingular.*

**Assumption 3.** *$X_n = B_{n1}\xi_{n1}$, $Z_n = B_{n2}\xi_{n2} + B_{n3}\xi_{n3}$, $F_n = B_{n4}\xi_{n4}$, where $B_{n1}$, $B_{n2}$, $B_{n3}$ and $B_{n4}$ are $n \times n$ unknown BRC nonstochastic matrices, $(\xi_{n1,i}, \xi_{n2,i}, \xi_{n3,i}, \xi_{n4,i}, \epsilon_{ni})$'s are i.i.d., $\xi_{n1,i}$'s and $\xi_{n4,i}$'s are independent of $\epsilon_{ni}$'s, $\mathrm{E}(\epsilon_n) = 0$, $\mathrm{E}(\epsilon_n\epsilon_n') = I_n$, $\mathrm{E}\,\|(\xi_{n1,i}, \xi_{n2,i}, \xi_{n3,i})\|_\infty^4 < \infty$, $\mathrm{E}\,\|(\xi_{n4,i}, \epsilon_{ni})\|_\infty^\tau < \infty$ for some $\tau > 4$, and $\lim_{n\to\infty} \frac{1}{n} \mathrm{E}(\Pi_n)$ exists and is nonsingular.*

**Assumption 4.** *There exists a constant $\eta > 0$ such that $|\alpha| \leq \eta$ and the true parameter $\alpha_0$ is in the interior of the parameter space $[-\eta, \eta]$.*

Let $\theta_0$ be the true parameter vector of $\theta$. The identification of $\theta_0$ requires a unique solution of $\lim_{n\to\infty} \frac{1}{n} \mathrm{E}[F_n'(e^{\alpha W_n} Y_n - D_n\beta)] = 0$ at $\theta_0$. Under regularity conditions,

$$\frac{1}{n} \mathrm{E}[F_n'(e^{\alpha W_n} Y_n - D_n\beta)] = \frac{1}{n}[\mathrm{E}(F_n' e^{(\alpha-\alpha_0)W_n} D_n)\beta_0, \mathrm{E}(F_n' D_n)] \begin{pmatrix} 1 \\ -\beta \end{pmatrix}.$$

When $\alpha = \alpha_0$, the preceding expression reduces to $\frac{1}{n} \mathrm{E}(F_n' D_n)(\beta_0 - \beta)$. An identification condition can be as in the following assumption.

**Assumption 5.** $\lim_{n\to\infty} \frac{1}{n}[\mathrm{E}(F_n' e^{(\alpha-\alpha_0)W_n} D_n)\beta_0, \mathrm{E}(F_n' D_n)]$ *has full column rank for any $\alpha \neq \alpha_0$.*

[6]In this case, we may have a many IV problem as in Liu and Lee (2013), which is beyond the scope of this paper.

[7]More discussions on the IV selection can be found in Kelejian and Prucha (2007).

[8]Such an assumption simplifies the argument on uniform convergence of the minimized sample average objective function over its parameter space.

Note that the condition of Assumption 5 implies, in particular, that $\lim_{n\to\infty} \frac{1}{n}\,\mathrm{E}(F_n' D_n)$ has full column rank. The condition would not hold when $\beta_0 = 0$; otherwise, it holds in general. Under the above regularity assumptions, the consistency of $\check{\theta}_n$ follows.

**Proposition 2.1.** *Under Assumptions 1–5, $\check{\theta}_n = \theta_0 + o_p(1)$.*

## 2.1 Asymptotic distribution: The regular case

We now consider the asymptotic distribution of the N2SLS estimator. Let the moment vector be $g_n(\theta) = F_n'(e^{\alpha W_n} Y_n - D_n\beta)$. Its Jacobian matrix is $G_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta'} = F_n'(W_n e^{\alpha W_n} Y_n, -D_n)$. Then,

$$
\begin{aligned}
\mathrm{E}[G_n(\theta_0)] &= [\mathrm{E}(F_n' W_n D_n)\beta_0, -\mathrm{E}(F_n' D_n)] \\
&= [\mathrm{E}(F_n' W_n X_n^*)\beta_{10} + \mathrm{E}(F_n' W_n^2 l_n)\beta_{20} + \mathrm{E}(F_n' W_n^2 X_{n1})\beta_{30} + \mathrm{E}(F_n' W_n Z_n)\beta_{40}, \\
&\quad - \mathrm{E}(F_n' X_n^*), -\mathrm{E}(F_n' W_n l_n), -\mathrm{E}(F_n' W_n X_{n1}), -\mathrm{E}(F_n' Z_n)].
\end{aligned} \tag{3}
$$

Let $\delta = (\beta_1', \beta_2)'$ and $\zeta = (\beta_3', \beta_4')'$ when $W_n$ is row-normalized; $\delta = \beta_1$ and $\zeta = (\beta_2, \beta_3', \beta_4')'$ when $W_n$ is not row-normalized. If $W_n$ is row-normalized and $\zeta_0 \neq 0$, the first column of $\mathrm{E}[G_n(\theta_0)]$ is generally not linearly dependent on $\mathrm{E}(F_n' D_n)$, since $\mathrm{E}(F_n' W_n^2 X_{n1})\beta_{30}$ or $\mathrm{E}(F_n' W_n Z_n)\beta_{40}$ appears in the first column. If $W_n$ is not row-normalized and $\zeta_0 \neq 0$, $\mathrm{E}(F_n' W_n^2 l_n)\beta_{20}$ might also appear in the first column of $\mathrm{E}[G_n(\theta_0)]$. Thus $\mathrm{E}[G_n(\theta_0)]$ generally has full rank. As a result, the asymptotic distribution of the N2SLS estimator can be derived as usual by applying the mean value theorem to the first order condition of the criterion function.

**Proposition 2.2.** *Under Assumptions 1–5, when $\zeta_0 \neq 0$, the N2SLS estimator $\check{\theta}_n$ has the asymptotic distribution*

$$
\sqrt{n}(\check{\theta}_n - \theta_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty} \big(\frac{1}{n}\bar{G}_n' \bar{\Pi}_n^{-1} \bar{G}_n\big)^{-1}\big),
$$

*where $\bar{\Pi}_n = \mathrm{E}(\Pi_n)$ and $\bar{G}_n = \mathrm{E}[G_n(\theta_0)]$, provided that $\operatorname{plim}_{n\to\infty} \frac{1}{n}\bar{G}_n$ has full rank. The best IV matrix $F_n$ is the matrix formed by the independent columns of $(T_n T_n')^{-1}[X_n^*, W_n X_n, W_n^2 X_n, \mathrm{E}(Z_n, W_n Z_n|\mathbb{X}_n)]$, where $\mathbb{X}_n$ denotes the matrix of all exogenous variables.*

Proposition 2.2 excludes the case that $\zeta_0 = 0$. This case turns out to be irregular, which needs special attention.

## 2.2 Asymptotic distribution: The irregular case

By (3), the Jacobian matrix of the moment vector at the true parameter vector is rank deficient w.p.a.1. when $\zeta_0 = 0$, i.e., when the Durbin regressors and endogenous explanatory variables are irrelevant, even $\delta_0 \neq 0$. In this subsection, we consider the N2SLS estimation of model (1) in this situation.

Although $\zeta_0 = 0$, the identification condition in Assumption 5 still holds when $\delta_0 \neq 0$. Thus the N2SLS estimator will be consistent. But in this case, the expected Jacobian matrix of the moment vector at the true parameter vector does not have full rank, so the usual way to derive the asymptotic distribution by the mean value theorem will not work. Instead, we analyze high order Taylor expansions of the first order condition of the N2SLS criterion function (2). Let $H_n = F_n \Pi_n^{-1} F_n'$. The first order derivatives of $Q_n(\theta)$ are:

$$\frac{\partial Q_n(\theta)}{\partial \alpha} = 2Y_n' e^{\alpha W_n'} W_n' H_n (e^{\alpha W_n} Y_n - D_n \beta), \tag{4}$$

$$\frac{\partial Q_n(\theta)}{\partial \beta} = -2D_n' H_n (e^{\alpha W_n} Y_n - D_n \beta). \tag{5}$$

Note that at $\theta_0 = (\alpha_0, \delta_0', 0)'$, $e^{\alpha_0 W_n} Y_n = X_n \delta_0 + V_n$, and

$$\frac{1}{\sqrt{n}} \frac{\partial Q_n(\theta_0)}{\partial \alpha} = \frac{2}{\sqrt{n}} (X_n \delta_0 + V_n)' W_n' H_n V_n = \frac{2}{\sqrt{n}} (W_n X_n \delta_0)' H_n V_n + O_p(\frac{1}{\sqrt{n}}) = O_p(1),$$

$$\frac{1}{\sqrt{n}} \frac{\partial Q_n(\theta_0)}{\partial \beta} = -\frac{2}{\sqrt{n}} D_n' H_n V_n = O_p(1).$$

Let $k_{x^*}$ be the number of columns in $X_n^*$, and $\delta = (\delta_1', \delta_2)'$ with $\delta_2$ being a scalar. Then,

$$\frac{1}{\sqrt{n}} \frac{\partial Q_n(\theta_0)}{\partial \alpha} + \frac{1}{\sqrt{n}} \frac{\partial Q_n(\theta_0)}{\partial \beta'} (0_{1 \times k_{x^*}}, \delta_{20}, \delta_{10}', 0_{1 \times k_z})' = o_p(1),$$

i.e., $\frac{1}{\sqrt{n}} \frac{\partial Q_n(\theta_0)}{\partial \alpha}$ and $\frac{1}{\sqrt{n}} \frac{\partial Q_n(\theta_0)}{\partial \beta}$ are linearly dependent w.p.a.1. As a result, $\frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \frac{\partial Q_n(\theta_0)}{\partial \theta'}$ is singular w.p.a.1. From (23)–(25) in Appendix A, $\frac{1}{n} \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} = \frac{2}{n} (-W_n X_n \delta_0, D_n)' H_n (-W_n X_n \delta_0, D_n) + o_p(1)$, which is also singular with large $n$ as $D_n$ contains $W_n X_n$ in its columns. We note that $\frac{1}{n} \mathrm{E}(\frac{\partial Q_n(\theta)}{\partial \theta} \frac{\partial Q_n(\theta)}{\partial \theta'})$ generally has full rank when $\theta \neq \theta_0$. Rothenberg (1971) shows that, in the likelihood theory of parametric models, if the information matrix has constant rank in an open neighborhood of the true parameter vector, then local identification of parameters is equivalent to nonsingularity of the information matrix at the true parameter vector. In the current case, the rank of $\frac{1}{n} \mathrm{E}(\frac{\partial Q_n(\theta)}{\partial \theta} \frac{\partial Q_n(\theta)}{\partial \theta'})$ evaluated at $\theta = \theta_0$ differs from that at $\theta \neq \theta_0$. So even though $\frac{1}{n} \mathrm{E}(\frac{\partial Q_n(\theta_0)}{\partial \theta} \frac{\partial Q_n(\theta_0)}{\partial \theta'})$ is singular w.p.a.1., the parameters may still be identifiable as previously argued.

Although the elements of the gradient vector are linearly dependent w.p.a.1., none of the elements is zero. In Rotnitzky et al. (2000), the asymptotic distribution of the MLE is first derived for a parametric model for which an element of the score is zero, where estimators corresponding to zero and nonzero scores have different convergence rates. If none of the elements of the score is zero but these elements are linearly dependent, the model is first reparameterized to be one for which one element of the score is zero. Following Rotnitzky et al. (2000), for our N2SLS estimation of the MESS model, consider the reparameterization $\omega = \binom{\alpha}{\beta + K'(\alpha - \alpha_0)} \equiv (\phi, \psi')'$, where $\phi$ is a scalar, and $K = \big[\mathrm{plim}_{n \to \infty} \frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \alpha} \frac{\partial Q_n(\theta_0)}{\partial \beta'}\big] \big[\mathrm{plim}_{n \to \infty} \frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \beta} \frac{\partial Q_n(\theta_0)}{\partial \beta'}\big]^{-1}$. At $\theta_0$, we have $\omega = \theta_0$. Denote $\omega_0 = \theta_0$. Then $Q_n(\theta) = Q_n(\phi, \psi - K'(\phi - \alpha_0)) \equiv Q_n^*(\omega)$, and $\frac{\partial Q_n^*(\omega_0)}{\partial \phi} = \frac{\partial Q_n(\theta_0)}{\partial \alpha} - [\mathrm{plim}_{n \to \infty} \frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \alpha} \frac{\partial Q_n(\theta_0)}{\partial \beta'}][\mathrm{plim}_{n \to \infty} \frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \beta} \frac{\partial Q_n(\theta_0)}{\partial \beta'}]^{-1} \frac{\partial Q_n(\theta_0)}{\partial \beta}$ is approximately the residual vector for the population regression of $\frac{\partial Q_n(\theta_0)}{\partial \alpha}$ on $\frac{\partial Q_n(\theta_0)}{\partial \beta}$. Because of the linear dependence of these two random vectors w.p.a.1., as a residual, $\frac{\partial Q_n^*(\omega_0)}{\partial \phi}$ must have a smaller order than $\frac{\partial Q_n(\theta_0)}{\partial \alpha}$. In our current case, $K = (0_{1 \times k_x^*}, -\delta_{20}, -\delta_{10}', 0_{1 \times k_z})$. So the reparameterization has $\omega = (\phi, \psi_1', \psi_2, \psi_3', \psi_4')' = (\alpha, \beta_1', \beta_2 - \delta_{20}(\alpha - \alpha_0), \beta_3' - \delta_{10}(\alpha - \alpha_0), \beta_4')'$ and

$$Q_n^*(\omega) = V_n'(\omega) H_n V_n(\omega), \tag{6}$$

where $V_n(\omega) = e^{\phi W_n} Y_n - X_n^* \psi_1 - W_n l_n [\psi_2 + \delta_{20}(\phi - \phi_0)] - W_n X_{n1} [\psi_3 + \delta_{10}(\phi - \phi_0)] - Z_n \psi_4$. The N2SLS estimator $\breve{\omega}_n$ minimizes $Q_n^*(\omega)$. Because of the one-to-one correspondence between $\theta$ and $\omega$, the consistency of the N2SLS estimator $\breve{\theta}_n$ implies the consistency of $\breve{\omega}_n$ to $\omega_0$.

To derive the asymptotic distribution of $\check{\omega}_n$, we need to investigate high order Taylor expansions of the first order condition $\frac{\partial Q_n^*(\check{\omega}_n)}{\partial \omega} = 0$. Here, we sketch the derivation of the asymptotic distribution, with details in the proof of Proposition 2.3. It turns out that we need a third order Taylor expansion of the first order conditions at the true parameter vector. In terms of $\check{\phi}_n$, $\check{\psi}_n$, and their true values, we find $\sqrt{n}(\check{\phi}_n - \phi_0)^2 = O_p(1)$ and $\sqrt{n}(\check{\psi}_n - \psi_0) = O_p(1)$, and, by further eliminating $\check{\psi}_n$ by substitution, the expansion yields:

$$\begin{aligned} 0 &= 2n^{-1/4} V_n' W_n' \mathbb{M}_D V_n - n^{-3/4} (W_n^2 X_n \delta_0)' \mathbb{P}_D W_n V_n \sqrt{n}(\check{\phi}_n - \phi_0)^2 \\ &\quad + n^{1/4}(\check{\phi}_n - \phi_0)\big[R_n + S_n \sqrt{n}(\check{\phi}_n - \phi_0)^2\big] + o_p(n^{-1/4}) \\ &= n^{1/4}(\check{\phi}_n - \phi_0)\big[R_n + S_n \sqrt{n}(\check{\phi}_n - \phi_0)^2\big] + O_p(n^{-1/4}), \end{aligned} \tag{7}$$

where $\mathbb{P}_D = H_n D_n (D_n' H_n D_n)^{-1} D_n' H_n$, $\mathbb{M}_D = H_n - \mathbb{P}_D$, $R_n = \frac{2}{\sqrt{n}} (W_n^2 X_n \delta_0)' \mathbb{M}_D V_n$, and $S_n = \frac{1}{n} (W_n^2 X_n \delta_0)' \mathbb{M}_D W_n^2 X_n \delta_0 = O(1)$. Note that $\mathbb{M}_D = H_n^{1/2} M_{H^{1/2}D} H_n^{1/2}$, where $H_n^{1/2}$ is a symmetric matrix such that $H_n = H_n^{1/2} H_n^{1/2}$, and $M_{H^{1/2}D} = I_n - H_n^{1/2} D_n (D_n' H_n D_n)^{-1} D_n' H_n^{1/2}$ is the orthogonal projector onto the null space of $D_n' H_n^{1/2}$. Then by the partitioned matrix formula, we have $R_n = O_p(1)$ and $S_n > 0$ w.a.p.1. under the following assumption:

**Assumption 6.** $\lim_{n\to\infty} \frac{1}{n}[\mathrm{E}(F_n' W_n^2 X_n)\delta_0, \mathrm{E}(F_n' D_n)]$ *has full column rank.*

Furthermore, when $R_n > 0$, as $S_n \sqrt{n}(\check{\phi}_n - \phi_0)^2 \geq 0$, we must have $\sqrt{n}(\check{\phi}_n - \phi_0)^2 = o_p(1)$; when $R_n < 0$, $R_n + S_n \sqrt{n}(\check{\phi}_n - \phi_0)^2 = o_p(1)$ and thus $\sqrt{n}(\check{\phi}_n - \phi_0)^2 = J_{1n} + o_p(1)$, where $J_{1n} = -S_n^{-1} R_n$. Note that $R_n = \frac{1}{\sqrt{n}} (W_n^2 X_n \delta_0)' [I_n - H_n D_n (D_n' H_n D_n)^{-1} D_n'] F_n \Pi_n^{-1} F_n' V_n$ is asymptotically normal by applying the central limit theorem in Qu and Lee (2012) to $\frac{1}{\sqrt{n}} F_n' V_n$.[9] The asymptotic distribution of $\sqrt{n}(\check{\phi}_n - \phi_0)^2$ is thus a mixture of a truncated normal and the point 0.

For $\check{\psi}_n$, when $R_n > 0$, the expansion of $\frac{\partial Q_n^*(\check{\omega}_n)}{\partial \psi} = 0$ at $\omega_0$ implies that $0 = \frac{1}{\sqrt{n}} \frac{\partial Q_n^*(\omega_0)}{\partial \psi} + \frac{1}{n} \frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'} \sqrt{n}(\check{\psi}_n - \psi_0) + o_p(1)$. Thus, $\sqrt{n}(\check{\psi}_n - \psi_0) = L_n + o_p(1)$, where $L_n = (\frac{1}{n} D_n' H_n D_n)^{-1} \frac{1}{\sqrt{n}} D_n' H_n V_n$ is the leading order term of $-(\frac{1}{n} \frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'})^{-1} \frac{1}{\sqrt{n}} \frac{\partial Q_n^*(\omega_0)}{\partial \psi}$. Note that $L_n \xrightarrow{d} L$, where $L$ is $N(0, \lim_{n\to\infty} [\frac{1}{n} \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} \mathrm{E}(F_n' D_n)]^{-1})$. When $R_n < 0$, we are essentially solving the following:

$$0 = \begin{pmatrix} \frac{1}{\sqrt{n}} \frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi^2} \\ \frac{1}{\sqrt{n}} \frac{\partial Q_n^*(\omega_0)}{\partial \psi} \end{pmatrix} + \begin{pmatrix} \frac{1}{6n} \frac{\partial^4 Q_n^*(\omega_0)}{\partial \phi^4} & \frac{1}{n} \frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi'} \\ \frac{1}{2n} \frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi} & \frac{1}{n} \frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'} \end{pmatrix} \begin{pmatrix} \sqrt{n}(\check{\phi}_n - \phi_0)^2 \\ \sqrt{n}(\check{\psi}_n - \psi_0) \end{pmatrix} + o_p(1). \tag{8}$$

Thus, $\binom{\sqrt{n}(\check{\phi}_n - \phi_0)^2}{\sqrt{n}(\check{\psi}_n - \psi_0)} = \binom{J_{1n}}{J_{2n}} + o_p(1)$, where

$$\begin{pmatrix} J_{1n} \\ J_{2n} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & I_{k_d} \end{pmatrix} \Big[\frac{1}{n} (-W_n^2 X_n \delta_0, D_n)' H_n (-W_n^2 X_n \delta_0, D_n)\Big]^{-1} \frac{1}{\sqrt{n}} (-W_n^2 X_n \delta_0, D_n)' H_n V_n \tag{9}$$

with $k_d = k_x^* + k_x + k_z + 1$ is the leading order term of $-\begin{pmatrix} \frac{1}{6n} \frac{\partial^4 Q_n^*(\omega_0)}{\partial \phi^4} & \frac{1}{n} \frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi'} \\ \frac{1}{2n} \frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi} & \frac{1}{n} \frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}} \frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi^2} \\ \frac{1}{\sqrt{n}} \frac{\partial Q_n^*(\omega_0)}{\partial \psi} \end{pmatrix}$, and $J_{1n}$ has the explicit form $J_{1n} = -S_n^{-1} R_n$, as in the last paragraph. By (9), $J_{2n} = L_n + (\frac{2}{n} D_n' H_n D_n)^{-1} \frac{1}{n} D_n' H_n W_n^2 X_n \delta_0 J_{1n}$.

[9] Note that $F_n$ is allowed to be stochastic in Assumption 3, so $\frac{1}{\sqrt{n}} F_n' V_n = \frac{1}{\sqrt{n}} \xi_{n4}' B_{n4}' T_n \epsilon_n$ is a special case of a general linear-quadratic form in Qu and Lee (2012).

Note that $J_n = (J_{1n}, J_{2n}')'$ can be further written as $J_n = \mathbb{J}_n + o_p(1)$, where

$$\mathbb{J}_n = \begin{pmatrix} 2 & 0 \\ 0 & I_{k_d} \end{pmatrix} \Big[\frac{1}{n} \mathrm{E}[(-W_n^2 X_n \delta_0, D_n)' F_n] \bar{\Pi}_n^{-1} \mathrm{E}[F_n'(-W_n^2 X_n \delta_0, D_n)]\Big]^{-1} \frac{1}{\sqrt{n}} \mathrm{E}[(-W_n^2 X_n \delta_0, D_n)' F_n] \bar{\Pi}_n^{-1} F_n' V_n.$$

Thus $J_n$ has the asymptotic distribution $J_n \xrightarrow{d} J$, where $J = (J_1, J_2')'$ is $N(0, \lim_{n\to\infty} \Delta_n)$ with $J_1$ being the first element of $J$ and

$$\Delta_n = \begin{pmatrix} 2 & 0 \\ 0 & I_{k_d} \end{pmatrix} \Big[\frac{1}{n} \mathrm{E}[(-W_n^2 X_n \delta_0, D_n)' F_n] \bar{\Pi}_n^{-1} \mathrm{E}[F_n'(-W_n^2 X_n \delta_0, D_n)]\Big]^{-1} \begin{pmatrix} 2 & 0 \\ 0 & I_{k_d} \end{pmatrix}$$

being the variance of $\mathbb{J}_n$. Thus, $L = J_2 - \lim_{n\to\infty}[\frac{2}{n} \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} \mathrm{E}(F_n' D_n)]^{-1} \frac{1}{n} \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} \mathrm{E}(F_n' W_n^2 X_n) \delta_0 J_1$, which is $N(0, \lim_{n\to\infty}[\frac{1}{n} \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} \mathrm{E}(F_n' D_n)]^{-1})$ as above.

From the above, only the asymptotic distribution of $\sqrt{n}(\check{\phi}_n - \phi_0)^2$ has been derived, but the sign of $n^{1/4}(\check{\phi}_n - \phi_0)$ has not. As we are interested in $(\check{\phi}_n - \phi_0)$, a further analysis for the sign of $n^{1/4}(\check{\phi}_n - \phi_0)$ is needed. For a fourth order Taylor expansion of $Q_n^*(\check{\omega}_n)$, the sign of $n^{1/4}(\check{\phi}_n - \phi_0)$ does not affect the leading order term of the fourth order polynomial. As $\check{\phi}_n$ is the N2SLS estimator, the sign of $n^{1/4}(\check{\phi}_n - \phi_0)$ should be chosen to minimize the remainder term of the fourth order Taylor expansion. Essentially, we derive the leading order term of the remainder by investigating a higher order—fifth order—Taylor expansion of $Q_n^*(\check{\omega}_n)$. When $R_n < 0$, $n^{1/4}(\check{\phi}_n - \phi_0)$ being positive is equivalent to some random variable being negative asymptotically. To describe this random variable, we define two random vectors that are uncorrelated with $\mathbb{J}_n$:

i) $U_{1n} = \frac{1}{\sqrt{n}} F_n' V_n - \Upsilon_{1n} \mathbb{J}_n$, where $\Upsilon_{1n} = \frac{1}{\sqrt{n}} \mathrm{E}(F_n' V_n \mathbb{J}_n') \Delta_n^{-1} = [-\frac{1}{2n} \mathrm{E}(F_n' W_n^2 X_n \delta_0), \frac{1}{n} \mathrm{E}(F_n' D_n)]$ and

ii) $U_{2n} = \frac{1}{\sqrt{n}} F_n' W_n V_n - \Upsilon_{2n} \mathbb{J}_n$, where $\Upsilon_{2n} = \frac{1}{\sqrt{n}} \mathrm{E}(F_n' W_n V_n \mathbb{J}_n') \Delta_n^{-1} = \mathrm{E}(F_n' W_n T_n T_n' F_n) \bar{\Pi}_n^{-1} \Upsilon_{1n}$.

The $U_n = (U_{1n}', U_{2n}')'$ is uncorrelated with $\mathbb{J}_n$ since it is the residual random vector for a population regression of $\frac{1}{\sqrt{n}}(V_n' F_n, V_n' W_n' F_n)'$ on $\mathbb{J}_n$. According to the fifth order Taylor expansion of $Q_n^*(\check{\omega}_n)$, we find that $\mathrm{P}(I(n^{1/4}(\check{\phi}_n - \phi_0) < 0) = I(K_n^* > 0) | R_n < 0) \to 1$ as $n \to \infty$, where $I(\cdot)$ is the set indicator and

$$\begin{aligned} K_n^* &= 2(U_{2n} + \Upsilon_{2n} \mathbb{J}_n)' \big(\frac{1}{n} \bar{\Pi}_n\big)^{-1} (U_{1n} + \Upsilon_{1n} \mathbb{J}_n) \\ &\quad + \mathbb{J}_n' \Big[ \begin{pmatrix} \frac{1}{3}[\mathrm{E}(F_n' W_n^3 X_n \delta_0)]' \bar{\Pi}_n^{-1} U_{1n} + [\mathrm{E}(F_n' W_n^2 X_n \delta_0)]' \bar{\Pi}_n^{-1} U_{2n} \\ -2 \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} U_{2n} \end{pmatrix} + \begin{pmatrix} [\mathrm{E}(F_n' W_n^2 X_n \delta_0)]' \bar{\Pi}_n^{-1} \Upsilon_{2n} \\ -2 \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} \Upsilon_{2n} \end{pmatrix} \mathbb{J}_n \Big] + o_p(1). \end{aligned} \tag{10}$$

Since $(U_n', J_n')' \xrightarrow{d} (U', J')'$, where $U = N(0, \lim_{n\to\infty} \mathrm{E}(U_n U_n'))$ is independent of $J$, $\check{\omega}_n$ has the following asymptotic distribution:

**Proposition 2.3.** *Under Assumptions 1–6, when* $\zeta_0 = 0$,

$$\begin{pmatrix} n^{1/4}(\check{\phi}_n - \phi_0) \\ \sqrt{n}(\check{\psi}_n - \psi_0) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (-1)^B J_1^{1/2} \\ J_2 \end{pmatrix} I(J_1 > 0) + \begin{pmatrix} 0 \\ L \end{pmatrix} I(J_1 < 0),$$

*where $B$ is a Bernoulli random variable with success probability equal to* $\mathrm{P}(K^* > 0 | J_1 > 0)$ *with* $K^* = 2 \lim_{n\to\infty} (U_2 + \Upsilon_{2n} J)' (\frac{1}{n} \bar{\Pi}_n)^{-1} (U_1 + \Upsilon_{1n} J) + \lim_{n\to\infty} J' \big[ \big( \begin{smallmatrix} \frac{1}{3}[\mathrm{E}(F_n' W_n^3 X_n \delta_0)]' \bar{\Pi}_n^{-1} U_1 + [\mathrm{E}(F_n' W_n^2 X_n \delta_0)]' \bar{\Pi}_n^{-1} U_2 \\ -2 \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} U_2 \end{smallmatrix} \big) + \big( \begin{smallmatrix} [\mathrm{E}(F_n' W_n^2 X_n \delta_0)]' \bar{\Pi}_n^{-1} \Upsilon_{2n} \\ -2 \mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} \Upsilon_{2n} \end{smallmatrix} \big) J \big]$.

Since $\alpha = \phi$, $\beta_1 = \psi_1$, $\beta_2 = \psi_2 + \delta_{20}(\phi - \phi_0)$, $\beta_3 = \psi_3 + \delta_{10}(\phi - \phi_0)$, and $\beta_4 = \psi_4$, the asymptotic distribution of $(\check{\alpha}_n, \check{\beta}_{1n}', \check{\beta}_{2n}, \check{\beta}_{3n}', \check{\beta}_{4n}')'$ follows by the continuous mapping theorem. Note that

$$n^{1/4}(\check{\beta}_{2n} - \beta_{20}) = \delta_{20} n^{1/4}(\check{\phi}_n - \phi_0) + n^{1/4}(\check{\psi}_{2n} - \psi_{20}) = \delta_{20} n^{1/4}(\check{\phi}_n - \phi_0) + o_p(1).$$

Similarly, $n^{1/4}(\check{\beta}_{3n} - \beta_{30}) = \delta_{10} n^{1/4}(\check{\phi}_n - \phi_0) + o_p(1)$. Hence, $\check{\alpha}_n$, $\check{\beta}_{1n}$ and $\check{\beta}_{2n}$ have rates of convergence that are slower than the usual $\sqrt{n}$-rate.

**Corollary 2.1.** *Under Assumptions 1–6, when $\zeta_0 = 0$,*

$$\begin{pmatrix} n^{1/4}(\check{\alpha}_n - \alpha_0) \\ n^{1/2}(\check{\beta}_{1n} - \beta_{10}) \\ n^{1/4}(\check{\beta}_{2n} - \beta_{20}) \\ n^{1/4}(\check{\beta}_{3n} - \beta_{30}) \\ n^{1/2}(\check{\beta}_{4n} - \beta_{40}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (-1)^B J_1^{1/2} \\ J_{2x^*} \\ (-1)^B \delta_{20} J_1^{1/2} \\ (-1)^B \delta_{10} J_1^{1/2} \\ J_{2z} \end{pmatrix} I(J_1 > 0) + \begin{pmatrix} 0 \\ L_{x^*} \\ 0_{k_x \times 1} \\ L_z \end{pmatrix} I(J_1 < 0), \tag{11}$$

*where $J_{2x^*}$ and $L_{x^*}$ are the subvectors consisting of the first $k_{x^*}$ elements of, respectively, $J_2$ and $L$, and $J_{2z}$ and $L_z$ are the subvectors consisting of the last $k_z$ elements of, respectively, $J_2$ and $L$.*

### 2.2.1 A special case: A MESS model with irrelevant Durbin's regressors but without endogenous regressors

The irregular phenomenon above appears when $\zeta_0 = 0$. It is of interest to note that if the endogenous variables are really relevant, i.e., $\beta_{40} \neq 0$, even though the Durbin regressors are irrelevant, the presence of relevant identifiable endogenous variables helps the identification of parameters including the unknown $\zeta_0$, which may be zero. If the Durbin regressors are irrelevant and not included in the model, even though $\beta_{40} = 0$, the Jacobian matrix of the moment vector at the true parameter vector in general has full rank w.p.a.1.. It is the Durbin regressors with unknown zero coefficients but not endogenous explanatory variables that lead to the irregular phenomenon. Thus, for the following MESS model without endogenous variables

$$e^{\alpha W_n} Y_n = X_n^* \beta_1 + W_n l_n \beta_2 + W_n X_{n1} \beta_3 + V_n, \quad V_n = T_n \epsilon_n, \tag{12}$$

we have a rank deficient expected Jacobian matrix of the moment vector for the N2SLS estimation at the true parameter vector with $\zeta_0 = 0$, where $\beta_4$ is excluded from $\zeta$, which is now $\zeta = \beta_3$ when $W_n$ is row-normalized and $\zeta = (\beta_2, \beta_3')'$ when $W_n$ is not row-normalized.

The MESS model (12) might be of interest in it own right. The N2SLS criterion function for (12) is $Q_n(\theta) = (e^{\alpha W_n} Y_n - X_n^* \beta_1 - W_n l_n \beta_2 - W_n X_{n1} \beta_3)' H_n (e^{\alpha W_n} Y_n - X_n^* \beta_1 - W_n l_n \beta_2 - W_n X_n \beta_3)$ with $\theta = (\alpha, \beta')' = (\alpha, \beta_1', \beta_2, \beta_3')' = (\alpha, \delta', \zeta')'$. Similar to model (1), when $\zeta_0 = 0$, $\frac{1}{n}\mathrm{E}(\frac{\partial Q_n(\theta_0)}{\partial \theta} \frac{\partial Q_n(\theta_0)}{\partial \theta'})$ and $\frac{1}{n}\mathrm{E}(\frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'})$ are singular w.p.a.1. Consider the reparameterization $\omega = \binom{\alpha}{\beta + K'(\alpha - \alpha_0)} \equiv (\phi, \psi')'$, where

$$K = -(0_{1 \times k^*}, -\delta_{20}, -\delta_{10}') = \big[\mathrm{plim}_{n\to\infty} \frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \alpha} \frac{\partial Q_n(\theta_0)}{\partial \beta'}\big]\big[\mathrm{plim}_{n\to\infty} \frac{1}{n} \frac{\partial Q_n(\theta_0)}{\partial \beta} \frac{\partial Q_n(\theta_0)}{\partial \beta'}\big]^{-1}.$$

Then $Q_n(\theta) = Q_n(\phi, \psi - K'(\phi - \alpha_0)) \equiv Q_n^*(\omega)$, and $\frac{\partial Q_n^*(\omega_0)}{\partial \phi}$ has a smaller order than $\frac{\partial Q_n(\theta_0)}{\partial \alpha}$. The reparameterization is $(\phi, \psi_1', \psi_2, \psi_3') = (\alpha, \beta_1', \beta_2 - \delta_{20}(\alpha - \alpha_0), \beta_3' - \delta_{10}'(\alpha - \alpha_0))$. Thus, we have $Q_n^*(\omega) = V_n'(\omega) H_n V_n(\omega)$, where $V_n(\omega) = e^{\phi W_n} Y_n - X_n^* \psi_1 - W_n l_n[\psi_2 + \delta_{20}(\phi - \phi_0)] - W_n X_{n1}[\psi_3 + \delta_{10}(\phi - \phi_0)]$. With some slight modifications of the assumptions to account for the exclusion of $Z_n$, the asymptotic distribution of $\check{\omega}_n$ that minimizes $Q_n^*(\omega)$ is in Proposition 2.3 by replacing all $D_n$ in relevant terms by $(X_n^*, W_n l_n, W_n X_{n1})$. Furthermore, the asymptotic distribution of the N2SLS estimator $(\check{\alpha}_n, \check{\beta}_{1n}', \check{\beta}_{2n}, \check{\beta}_{3n}')'$ that minimizes $Q_n(\theta)$ is in Corollary 2.1.

### 2.3 Feasible N2SLS estimator

The N2SLS estimator in the above analysis is infeasible as the criterion function $Q_n(\theta)$ involves the unknown matrix $T_n$. We consider a feasible N2SLS estimator using the spatial HAC estimator in Kelejian and Prucha (2007).

For the formulation of a feasible N2SLS estimator, a consistent estimator for the covariance matrix $\frac{1}{n}\Pi_n = (\varpi_{n,rs})$ is needed. This can be derived as follows. First we can derive a $\sqrt{n}$-consistent but may be inefficient estimator $\dot{\theta}_n$ from some feasible N2SLS estimation, e.g., the minimizer of $(e^{\alpha W_n} Y_n - D_n\beta)' F_n (F_n' F_n)^{-1} F_n' (e^{\alpha W_n} Y_n - D_n\beta)$, where $F_n$ consists of observable IV variables. Then the residual vector can be computed as $\hat{V}_n = e^{\dot{\alpha}_n W_n} Y_n - D_n \dot{\beta}_n$. Let $\hat{V}_n = (\hat{v}_{ni})$ and $F_n = (f_{n,ir})$. The spatial HAC estimator $\hat{\varpi}_{n,rs}$ of $\varpi_{n,rs}$ has the form

$$\hat{\varpi}_{n,rs} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} f_{n,ir} f_{n,js} \hat{v}_{ni} \hat{v}_{nj} \mathcal{K}(d_{n,ij}^*/d_n), \tag{13}$$

where $d_{n,ij}^* = d_{n,ji}^* \geq 0$ is a distance measured by the researcher for a meaningful distance measure $d_{n,ij} = d_{n,ji} \geq 0$ between spatial units $i$ and $j$,[10] $d_n > 0$ is a selected distance by the researcher such that $d_n \to \infty$ as $n \to \infty$, and $\mathcal{K}(\cdot)$ is a kernel function. The $d_{n,ij}^*$ can have some error for $d_{n,ij}$, but the errors should be uniformly bounded. Different from Kelejian and Prucha (2007), our IV matrix $F_n$ is allowed to be stochastic, so we maintain a set of assumptions similar to those in Kelejian and Prucha (2007) but account for randomness of $F_n$.

**Assumption 7.** *(i)* $\sqrt{n}(\dot{\theta}_n - \theta_0) = O_p(1)$; *(ii) for* $\ell_n = \max_{1 \leq i \leq n}(\ell_{i,n})$ *with* $\ell_{i,n} = \sum_{j=1}^n I(d_{n,ij}^* \leq d_n)$, $\mathrm{E}(\ell_n^2) = o(n^{2c_\ell})$, *where* $c_\ell \leq (\tau - 2)/[2(\tau - 1)]$ *for the* $\tau$ *in Assumption 3, and* $\sum_{j=1}^n |\sigma_{n,ij}| d_{n,ij}^{\rho_S} \leq c_S$ *for some* $\rho_S \geq 1$ *and* $0 < c_S < \infty$, *where* $\sigma_{n,ij}$ *is the* $(i,j)$*th element of* $T_n T_n'$; *(iii)* $d_{n,ij}^* = d_{n,ij} + \xi_{5n,ij} \geq 0$, *where* $\xi_{5n,ij} = \xi_{5n,ji}$ *denotes the measurement errors,* $\xi_{5n,ij} \leq c_d$ *with* $0 < c_d < \infty$, *and* $\xi_{5n,i}$*'s are independent of* $\epsilon_{ni}$*'s and* $\xi_{4n,i}$*'s; (iv) the kernel* $\mathcal{K}: \mathbb{R} \to [-1, 1]$, *with* $\mathcal{K}(0) = 1$, $\mathcal{K}(x) = \mathcal{K}(-x)$, $\mathcal{K}(x) = 0$ *for* $|x| > 1$, *satisfies* $|\mathcal{K}(x) - 1| \leq c_{\mathcal{K}} |x|^{\rho_{\mathcal{K}}}$ *for* $|x| \leq 1$, *some* $\rho_{\mathcal{K}} \geq 1$ *and* $0 < c_{\mathcal{K}} < \infty$.

The $\sqrt{n}$-consistency of $\dot{\theta}_n$ in Assumption 7$(i)$ can be established as for $\check{\theta}_n$ in Proposition 2.2. Assumption 7$(ii)$–$(iv)$ are the same as assumptions 4, 5 and 7 in Kelejian and Prucha (2007), except the independence of $\xi_{5n,i}$'s and $\xi_{4n,i}$'s. The $\ell_n$ in Assumption 7$(ii)$ plays the same role as the bandwidth parameter in time series literature and it limits the number of sample covariances entering into the HAC estimator. The second part of Assumption 7$(ii)$ restricts the degree of correlation that relates to distances between spatial units. Assumption 7$(iii)$ requires the measurement errors to be uniformly bounded and independent of the model disturbances and the disturbances for

[10] Multiple distances are also allowed in Kelejian and Prucha (2007), which we omit here for simplicity.

the instruments. Assumption 7($iv$) specifies a kernel as usual in time series literature and the properties are satisfied by many kernels. Detailed discussions can be found in Kelejian and Prucha (2007).

Under the maintained conditions, we may show that $\hat{\varpi}_{n,rs} - \varpi_{n,rs} = o_p(1)$. Let $\frac{1}{n}\hat{\Pi}_n = (\hat{\varpi}_{n,rs})$. The criterion function $\hat{Q}_n(\theta)$ of the feasible N2SLS estimator $\hat{\theta}_n$ is

$$\hat{Q}_n(\theta) = (e^{\alpha W_n} Y_n - D_n\beta)'\hat{H}_n(e^{\alpha W_n} Y_n - D_n\beta), \tag{14}$$

where $\hat{H}_n = F_n \hat{\Pi}_n^{-1} F_n'$. With the reparameterization in Section 2.2, let $\hat{Q}_n^*(\omega) = \hat{Q}_n(\phi, \psi - K'(\phi - \alpha_0))$ and $\hat{\omega}_n$ be the minimizer of $\hat{Q}_n^*(\omega)$.

**Proposition 2.4.** *Under Assumptions 1–7,* $\frac{1}{n}\hat{\Pi}_n - \frac{1}{n}\Pi_n = o_p(1)$, *and the results in Propositions 2.1 and 2.2 if* $\check{\theta}_n$ *is replaced by* $\hat{\theta}_n$ *and the result in Proposition 2.3 if* $\check{\omega}_n$ *is replaced by* $\hat{\omega}_n$ *still hold.*[11]

# 3 Testing for the irrelevance of the Durbin and endogenous regressors

In this section, we derive the distance difference and gradient tests that test for the irrelevance of the Durbin and endogenous variables, and also investigate their local power properties.[12]

## 3.1 Test statistics

With the restriction $\zeta = 0$ imposed, the restricted N2SLS estimator $\tilde{\Psi}_n$ minimizes the criterion function

$$\hat{Q}_n(\Psi, 0) = (e^{\alpha W_n} Y_n - X_n\delta)'\hat{H}_n(e^{\alpha W_n} Y_n - X_n\delta),$$

where $\Psi = (\alpha, \delta')'$. This restricted estimation does not have irregular features. As the estimation of this restricted model is regular, the asymptotic distribution of $\tilde{\Psi}_n$ follows from the equation:

$$\sqrt{n}(\tilde{\Psi}_n - \Psi_0) = \Big[\frac{1}{n}(-W_n X_n \delta_0, X_n)' H_n (-W_n X_n \delta_0, X_n)\Big]^{-1} \frac{1}{\sqrt{n}}(-W_n X_n \delta_0, X_n)' H_n V_n + o_p(1). \tag{15}$$

This relation for the restricted model will be useful for deriving asymptotic distributions of test statistics.

### 3.1.1 The distance difference test

For the unrestricted model, we take a fourth order Taylor expansion of $Q_n^*(\hat{\omega}_n)$ and collect terms that go to zero in probability into a remainder. As shown in Section 2.2, the estimator behaves differently for the two cases $R_n < 0$

[11]With unknown $T_n$, the N2SLS estimation with the best IV matrix in Proposition 2.2 is not feasible, since the best IV matrix cannot be consistently estimated.

[12]The Wald test is not considered here, because the usual Wald test is a quadratic form of the estimator and has an asymptotic chi-squared distribution, but from Corollary 2.1, a quadratic form of $\hat{\zeta}_n$ will not have an asymptotic chi-squared distribution due to the irregular feature under $H_0 : \zeta_0 = 0$.

and $R_n > 0$. When $R_n < 0$, from the proof of Proposition 2.3 and its analysis, we have:

$$
\begin{aligned}
\hat{Q}_n^*(\hat{\omega}_n) - \hat{Q}_n^*(\omega_0) &= \Big(\frac{1}{2\sqrt{n}}\frac{\partial^2 \hat{Q}_n^*(\omega_0)}{\partial \phi^2}, \frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n^*(\omega_0)}{\partial \psi'}\Big)\begin{pmatrix}\sqrt{n}(\hat{\phi}_n-\phi_0)^2\\ \sqrt{n}(\hat{\psi}_n-\psi_0)\end{pmatrix} \\
&\quad + \begin{pmatrix}\sqrt{n}(\hat{\phi}_n-\phi_0)^2\\ \sqrt{n}(\hat{\psi}_n-\psi_0)\end{pmatrix}' \begin{pmatrix}\frac{1}{24n}\frac{\partial^4 \hat{Q}_n^*(\omega_0)}{\partial \phi^4} & \frac{1}{4n}\frac{\partial^3 \hat{Q}_n^*(\omega_0)}{\partial \phi^2 \partial \psi'} \\ \frac{1}{4n}\frac{\partial^3 \hat{Q}_n^*(\omega_0)}{\partial \phi^2 \partial \psi} & \frac{1}{2n}\frac{\partial^2 \hat{Q}_n^*(\omega_0)}{\partial \psi \partial \psi'}\end{pmatrix} \begin{pmatrix}\sqrt{n}(\hat{\phi}_n-\phi_0)^2\\ \sqrt{n}(\hat{\psi}_n-\psi_0)\end{pmatrix} + o_p(1) \\
&= -V_n' \mathbb{P}_{(-W^2X\delta_0, D)} V_n + o_p(1),
\end{aligned}
$$

where the second equality follows by using (8) and orders of relevant derivatives in Appendix A, and $\mathbb{P}_{(-W^2X\delta_0,D)} = H_n(-W_n^2X_n\delta_0, D_n)[(-W_n^2X_n\delta_0, D_n)'H_n(-W_n^2X_n\delta_0, D_n)]^{-1}(-W_n^2X_n\delta_0, D_n)'H_n$. On the other hand, when $R_n > 0$, we have:

$$
\hat{Q}_n^*(\hat{\omega}_n) - \hat{Q}_n^*(\omega_0) = \frac{\partial \hat{Q}_n^*(\omega_0)}{\partial \psi'}(\hat{\psi}_n - \psi_0) + \frac{1}{2}(\hat{\psi}_n-\psi_0)'\frac{\partial^2 \hat{Q}_n^*(\omega_0)}{\partial \psi \partial \psi'}(\hat{\psi}_n-\psi_0) + o_p(1) = -V_n'\mathbb{P}_D V_n + o_p(1),
$$

because terms associated with $(\hat{\phi}_n - \phi_0)$ and its powers have small order $o_p(1)$ due to derivatives with respect to $\phi$ (in Appendix A) having small orders and $n^{1/4}(\hat{\phi}_n - \phi_0) = o_p(1)$ when $R_n > 0$ as shown in the proof of Proposition 2.3. With $\hat{Q}_n(\Psi, 0)$ of the constrained model and its constrained estimator $\tilde{\theta}_n = (\tilde{\Psi}_n', 0)'$, by a first order Taylor expansion of $\hat{Q}_n(\tilde{\theta}_n)$ at $\theta_0 = (\Psi_0', 0)'$, we have:

$$
\hat{Q}_n(\theta_0) - \hat{Q}_n(\tilde{\theta}_n) = \frac{1}{2}\sqrt{n}(\Psi_0 - \tilde{\Psi}_n)'\frac{1}{n}\frac{\partial^2 \hat{Q}_n(\ddot{\Psi}_n, 0)}{\partial \Psi \partial \Psi'}\sqrt{n}(\Psi_0 - \tilde{\Psi}_n) = V_n'\mathbb{P}_{(-WX\delta_0, X)}V_n + o_p(1),
$$

where $\mathbb{P}_{(-WX\delta_0,X)} = H_n(-W_nX_n\delta_0, X_n)[(-W_nX_n\delta_0, X_n)'H_n(-W_nX_n\delta_0, X_n)]^{-1}(-W_nX_n\delta_0, X_n)'H_n$, $\ddot{\Psi}_n$ lies between $\Psi_0$ and $\tilde{\Psi}_n$, and the second equality follows by (15) and small orders of second derivative terms in Appendix A.

Thus, as $\hat{Q}_n(\hat{\theta}_n) = \hat{Q}_n^*(\hat{\omega}_n)$ and $\hat{Q}_n(\theta_0) = \hat{Q}_n^*(\omega_0)$,

$$
\begin{aligned}
\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n) &= I(R_n < 0)V_n'(\mathbb{P}_{(-W^2X\delta_0,D)} - \mathbb{P}_{(-WX\delta_0,X)})V_n \\
&\quad + I(R_n > 0)V_n'(\mathbb{P}_D - \mathbb{P}_{(-WX\delta_0,X)})V_n + o_p(1).
\end{aligned} \tag{16}
$$

In the two cases $R_n < 0$ and $R_n > 0$, the test statistic is asymptotically distributed as chi-squared random variables, but the degree of freedom in the latter case is one less than that in the former case.

**Proposition 3.1.** *Under Assumptions 1–7, when $\zeta_0 = 0$, $\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n) \xrightarrow{d} T$, where $T$ is a mixture of a $\chi^2(k_{x^*} + k_z)$ and a $\chi^2(k_{x^*} + k_z - 1)$ random variable with mixing probabilities equal to $1/2$.*[13]

We may compute the $p$-value of the test or solve for the critical value via $\mathrm{P}(T > t) = \frac{1}{2}\mathrm{P}(\chi^2(k_{x^*} + k_z) > t) + \frac{1}{2}\mathrm{P}(\chi^2(k_{x^*} + k_z - 1) > t)$.

### 3.1.2 The gradient test

The gradient test is based on the asymptotic distribution of $\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta}$, where $\tilde{\theta}_n = (\tilde{\Psi}_n', 0)'$ is the restricted N2SLS estimator with $\zeta = 0$ imposed. Under the null hypothesis with $\theta_0 = (\Psi_0', 0)'$, by the mean value theorem and second

[13] $\chi^2(0)$ means the constant 0.

order derivatives in Appendix A,

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta} &= \frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\theta_0)}{\partial \zeta} + \frac{1}{n}\frac{\partial^2 \hat{Q}_n(\ddot{\theta}_n)}{\partial \zeta \partial \Psi'}\sqrt{n}(\tilde{\Psi}_n - \Psi_0) \\ &= -\frac{2}{\sqrt{n}}(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}V_n + o_p(1) \\ &\xrightarrow{d} N\big(0, \operatorname{plim}_{n\to\infty}\frac{4}{n}(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)\big),\end{aligned} \tag{17}$$

where $\mathbb{M}_{(-WX\delta_0,X)} = H_n - \mathbb{P}_{(-WX\delta_0,X)}$, $\ddot{\theta}_n$ lies between $\tilde{\theta}_n$ and $\theta_0$, $X_n^{**} = (l_n, X_{n1})$ when $W_n$ is not row-normalized, and $X_n^{**} = X_{n1}$ when $W_n$ is row-normalized. Because $\mathbb{M}_{(-WX\delta_0,X)}W_n X_n \delta_0 = 0$, the covariance matrix $(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)$ is singular. We may show that the rank of this covariance matrix is $k_{x^*} + k_z - 1$, i.e., one less than the number of its columns. Using the asymptotically normally distributed gradient vector $\frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta}$, an asymptotic chi-squared statistic can be constructed via the Moore-Penrose pseudoinverse of its asymptotic covariance matrix. Let $A^+$ be the Moore-Penrose pseudoinverse of a square matrix $A$, and $\tilde{\mathbb{M}}_{(-WX\delta_0,X)}$ be the matrix obtained by replacing $\delta_0$ in $\mathbb{M}_{(-WX\delta_0,X)}$ with $\tilde{\delta}_n$. Then we have the following proposition.

**Proposition 3.2.** *Under Assumptions 1–7, when $\zeta_0 = 0$,*

$$\frac{1}{4}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta'}[(W_n X_n^{**}, Z_n)'\tilde{\mathbb{M}}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)]^+\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta} \xrightarrow{d} \chi^2(k_{x^*} + k_z - 1).$$

## 3.2 Local power

We consider the local power of the test statistics under the alternative hypothesis that the true parameter $\zeta_0$ of the model with sample size $n$ is subject to the Pitman drift in Assumption 8 (the Durbin regressors and endogenous explanatory variables are relevant).

**Assumption 8.** $\zeta_{0n} = \frac{1}{\sqrt{n}}\kappa$, *where $\kappa$ is a $(k_{x^*} + k_z) \times 1$ nonzero vector.*

When $\zeta_0 = 0$, the N2SLS estimators $\hat{\beta}_{1n}$ and $\hat{\beta}_{4n}$ are $\sqrt{n}$-consistent, but $\hat{\alpha}_n$, $\hat{\beta}_{2n}$ and $\hat{\beta}_{3n}$ can only be $n^{1/4}$-consistent. The distance difference test integrates the information of all components of the N2SLS estimator, so it might be able to detect the small drift $\frac{1}{\sqrt{n}}\kappa$ from $\zeta_0 = 0$. Under the local alternative in Assumption 8, the restricted estimator $\tilde{\Psi}_n$ with the restriction $\zeta = 0$ imposed satisfies $\tilde{\Psi}_n = \Psi_0 + o_p(1)$. By the mean value theorem, under Assumption 8,

$$\begin{aligned}\sqrt{n}(\tilde{\Psi}_n - \Psi_0) &= -\big(\frac{1}{n}\frac{\partial^2 \hat{Q}_n(\ddot{\Psi}_n, 0)}{\partial \Psi \partial \Psi'}\big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\Psi_0, 0)}{\partial \Psi} \\ &= \big[\frac{1}{n}(-W_n X_n \delta_0, X_n)'H_n(-W_n X_n \delta_0, X_n)\big]^{-1}(-W_n X_n \delta_0, X_n)'H_n\big[\frac{1}{n}(W_n X_n^{**}, Z_n)\kappa + \frac{1}{\sqrt{n}}V_n\big] + o_p(1),\end{aligned} \tag{18}$$

where $\ddot{\Psi}_n$ is between $\Psi_0$ and $\tilde{\Psi}_n$. When $\kappa$ is not proportional to $(\delta_{20}, \delta_{10}', 0)'$, both $\tilde{\alpha}_n$ and $\tilde{\delta}_n$ can be asymptotically biased. Only $\tilde{\alpha}_n$ is asymptotically biased if $\kappa$ is proportional to $(\delta_{20}, \delta_{10}', 0)'$. Anyhow, $\tilde{\Psi}_n$ has the usual rate $O_p(n^{-1/2})$. Then, the gradient test might detect the small drift $\frac{1}{\sqrt{n}}\kappa$ from $\zeta_0 = 0$.

### 3.2.1 The distance difference test

Under the local alternative in Assumption 8, the N2SLS estimator $\hat{\theta}_n$ will still satisfy $\hat{\theta}_n = \theta_{0n} + o_p(1)$, where $\theta_{0n} = (\alpha_0, \delta_0', \zeta_{0n}')'$. We use the same reparameterization $\omega$ as in Section 2.2. Thus, corresponding to the drift $\kappa/\sqrt{n}$ in Assumption 8, $\omega_{0n} = \theta_{0n}$, i.e., $\phi_0 = \alpha_0$, $\psi_{0n} = (\psi_{10}', \psi_{20,n}', \psi_{30,n}', \psi_{40,n}')'$ with $\psi_{10} = \beta_{10}$ and $(\psi_{20,n}', \psi_{30,n}', \psi_{40,n}')' = \kappa/\sqrt{n}$ when the spatial weights matrix is not row-normalized, and $\psi_{0n} = (\psi_{10}', \psi_{20}', \psi_{30,n}', \psi_{40,n}')'$ with $\psi_{10} = \beta_{10}$, $\psi_{20} = \beta_{20}$ and $(\psi_{30,n}', \psi_{40,n}')' = \kappa/\sqrt{n}$ when the spatial weights matrix is row-normalized. Relevant derivatives of $\hat{Q}_n^*(\omega)$ at $\omega_{0n}$ have the same orders as before in Appendix A. By an analysis similar to that in Section 2.2, the estimator $\hat{\omega}_n$ has the following asymptotic distribution under the local alternative in Assumption 8.

**Proposition 3.3.** *Under Assumptions 1–8,*

$$\begin{pmatrix} n^{1/4}(\hat{\phi}_n - \phi_0) \\ \sqrt{n}(\hat{\psi}_n - \psi_{0n}) \end{pmatrix} = \begin{pmatrix} (-1)^B J_{1n}^{1/2} \\ J_{2n} \end{pmatrix} I(R_n < 0) + \begin{pmatrix} 0 \\ L_n \end{pmatrix} I(R_n > 0) + o_p(1) \xrightarrow{d} \begin{pmatrix} (-1)^B J_1^{1/2} \\ J_2 \end{pmatrix} I(J_1 > 0) + \begin{pmatrix} 0 \\ L \end{pmatrix} I(J_1 < 0),$$

*where $B$ is a Bernoulli random variable with success probability equal to* $\mathrm{P}(K^* > 0 | J_1 > 0)$, *and* $K^* = 2 \lim_{n\to\infty} \big(U_2 + \Upsilon_{2n} J + \frac{1}{n}[\mathrm{E}(F_n' W_n^2 X_n^{**}), \mathrm{E}(F_n' W_n Z_n)]\kappa\big)' (\frac{1}{n}\bar{\Pi}_n)^{-1}(U_1 + \Upsilon_{1n} J) + \lim_{n\to\infty} J' \Big[ \Big( {\textstyle \frac{1}{3}[\mathrm{E}(F_n' W_n^3 X_n \delta_0)]' \bar{\Pi}_n^{-1} U_1 + [\mathrm{E}(F_n' W_n^2 X_n \delta_0)]' \bar{\Pi}_n^{-1} U_2 \atop -2\,\mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} U_2} \Big) + \Big( {\textstyle [\mathrm{E}(F_n' W_n^2 X_n \delta_0)]' \bar{\Pi}_n^{-1} \Upsilon_{2n} \atop -2\,\mathrm{E}(D_n' F_n) \bar{\Pi}_n^{-1} \Upsilon_{2n}} \Big) J + \Big( {\textstyle \frac{1}{n}[\mathrm{E}(F_n' W_n^2 X_n \delta_0)]' \bar{\Pi}_n^{-1} [\mathrm{E}(F_n' W_n^2 X_n^{**}), \mathrm{E}(F_n' W_n Z_n)]\kappa \atop -\frac{2}{n}[\mathrm{E}(F_n' D_n)]' \bar{\Pi}_n^{-1} [\mathrm{E}(F_n' W_n^2 X_n^{**}), \mathrm{E}(F_n' W_n Z_n)]\kappa} \Big) \Big]$.

We use Proposition 3.3 to derive the asymptotic distribution of the distance difference test statistic under the local alternative in Assumption 8. The test statistic needs to be expanded differently for the two cases $R_n < 0$ and $R_n > 0$. Let $\chi^2(a, b)$ be a noncentral chi-squared distribution with $a$ degrees of freedom and noncentrality parameter $b$..

**Proposition 3.4.** *Under Assumptions 1–8, $\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n) \xrightarrow{d} [r^2 + \chi^2(k_{x*} + k_z - 1, c_1)] I(r > 0) + \chi^2(k_{x*} + k_z - 1, c_1) I(r < 0)$, where $c_1 = \mathrm{plim}_{n\to\infty} \frac{1}{n} \kappa' (W_n X_n^{**}, Z_n)' \mathbb{M}_{(-WX\delta_0, X)} (W_n X_n^{**}, Z_n) \kappa$ and $r$ is a standard normal random variable which is independent of $\chi^2(k_{x*} + k_z - 1, c_1)$, i.e., the asymptotic distribution of $\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n)$ is a mixture of two noncentral chi-squared distributions $\chi^2(k_{x*} + k_z, c_1)$ and $\chi^2(k_{x*} + k_z - 1, c_1)$, with both noncentrality parameters equal to $c_1$ and with mixing probabilities equal to* $1/2$.

When $\kappa$ is proportional to $(\delta_{20}, \delta_{10}', 0)'$, the noncentrality parameters are zero and the test has trivial power for the Pitman drift $\zeta_{0n} = \frac{1}{\sqrt{n}}\kappa$.

### 3.2.2 The gradient test

We now consider the local power of the gradient test by assuming that the DGP is subject to the Pitman drift in Assumption 8. Note that $\frac{1}{\sqrt{n}} \frac{\partial \hat{Q}_n(\Psi_0, 0)}{\partial \zeta} = -\frac{2}{\sqrt{n}}(W_n X_n^{**}, Z_n)' H_n (e^{\alpha_0 W_n} Y_n - X_n \delta_0) = -\frac{2}{n}(W_n X_n^{**}, Z_n)' H_n (W_n X_n^{**}, Z_n)\kappa - \frac{2}{\sqrt{n}}(W_n X_n^{**}, Z_n)' H_n V_n$ and $\frac{1}{n} \frac{\partial^2 \hat{Q}_n(\tilde{\Psi}_n, 0)}{\partial \zeta \partial \Psi'} = \frac{2}{n}(W_n X_n^{**}, Z_n)' H_n (-W_n e^{\tilde{\alpha}_n W_n} Y_n, X_n) = \frac{2}{n}(W_n X_n^{**}, Z_n)' H_n (-W_n X_n \delta_0, X_n) +$

$o_p(1)$. Then, by the mean value theorem and using (18),

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\tilde{\Psi}_n,0)}{\partial \zeta}\\
&= \frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\Psi_0,0)}{\partial \zeta} + \frac{1}{n}\frac{\partial^2 \hat{Q}_n(\ddot{\Psi}_n,0)}{\partial \zeta \partial \Psi'}\sqrt{n}(\tilde{\Psi}_n - \Psi_0)\\
&= -\frac{2}{n}(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)\kappa - \frac{2}{\sqrt{n}}(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}V_n + o_p(1)\\
&\xrightarrow{d} N\big(-\operatorname{plim}_{n\to\infty}\frac{2}{n}(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)\kappa, \operatorname{plim}_{n\to\infty}\frac{4}{n}(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)\big),
\end{aligned}$$

where $\ddot{\Psi}_n$ is between $\Psi_0$ and $\tilde{\Psi}_n$. Proposition 3.5 then follows.

**Proposition 3.5.** *Under Assumptions 1–8,*

$$\frac{1}{4}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta'}[(W_n X_n^{**}, Z_n)'\tilde{\mathbb{M}}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)]^{+}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta} \xrightarrow{d} \chi^2(k_{x^*} + k_z - 1, c_1).$$

Under the Pitman drift in Assumption 8, the gradient test statistic is asymptotically distributed as a noncentral chi-square random variable with the noncentrality parameter being the same as that for the distance difference test in Proposition 3.4. However, the distance difference test has one more degree of freedom with probability 0.5 and the same number of degrees of freedom with probability 0.5.

When $\kappa$ is proportional to $(\delta_{20}, \delta_{10}', 0)'$, the noncentrality parameter is zero and the test has trivial power. The test is not able to detect the small drift $n^{-1/2}(\delta_{20}, \delta_{10}', 0)'$ from $\zeta_0 = 0$. In this case, we should consider a larger Pitman drift in Assumption 9, which corresponds to the rate of convergence for $(\hat{\alpha}_n, \hat{\beta}_{2n}, \hat{\beta}_{3n}')'$.

**Assumption 9.** $\zeta_{0n} = n^{-1/4}(\delta_{20}, \delta_{10}', 0)'$.

Under Assumption 9, by the mean value theorem and using the derivatives in Appendix A,

$$\begin{aligned}
n^{1/4}(\tilde{\Psi}_n - \Psi_0) &= -\Big(\frac{1}{n}\frac{\partial^2 \hat{Q}_n(\ddot{\Psi}_n,0)}{\partial \Psi \partial \Psi'}\Big)^{-1} n^{-3/4}\frac{\partial \hat{Q}_n(\Psi_0,0)}{\partial \Psi}\\
&= \big[(-W_n X_n \delta_0, X_n)'H_n(-W_n X_n \delta_0, X_n)\big]^{-1}(-W_n X_n \delta_0, X_n)'H_n W_n X_n \delta_0 + O_p(n^{-1/4})\\
&= -\begin{pmatrix}1\\0\end{pmatrix} + O_p(n^{-1/4}),
\end{aligned}$$

where $\ddot{\Psi}_n$ is between $\Psi_0$ and $\tilde{\Psi}_n$. In this case, $\tilde{\Psi}_n - \Psi_0$ has the order $O_p(n^{-1/4})$ due to the drift, but $\tilde{\Psi}_n - \Psi_0 + n^{-1/4}\binom{1}{0}$ has the order $O_p(n^{-1/2})$. We may find the leading order term of $n^{1/2}(\tilde{\Psi}_n - \Psi_0 + n^{-1/4}\binom{1}{0})$ by expanding the first order condition $\frac{\partial \hat{Q}_n(\tilde{\Psi}_n,0)}{\partial \Psi} = 0$ at $(\alpha_0 - n^{1/4}, \delta_0')'$. Applying the result, then the asymptotic distribution of the gradient test statistic can be derived.

**Proposition 3.6.** *Under Assumptions 1–7 and 9,*

$$\frac{1}{4}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta'}[(W_n X_n^{**}, Z_n)'\tilde{\mathbb{M}}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)]^{+}\frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta} \xrightarrow{d} \chi^2(k_{x^*} + k_z - 1, c_2),$$

*where* $c_2 = \operatorname{plim}_{n\to\infty}\frac{1}{4n}(W_n^2 X_n \delta_0)'\mathbb{M}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)[(W_n X_n^{**}, Z_n)'\mathbb{M}_{(-WX\delta_0,X)}(W_n X_n^{**}, Z_n)]^{+}(W_n X_n^{**}, Z_n)'$ $\mathbb{M}_{(-WX\delta_0,X)}W_n^2 X_n \delta_0$.

Under the maintained regularity conditions, $c_2$ is generally non-zero. Thus, for the direction $(\delta_{20}, \delta_{10}', 0)'$, the gradient test can still have nontrivial power, but it is in terms of the larger drift $n^{-1/4}(\delta_{20}, \delta_{10}', 0)'$.

# 4 AGLASSO estimator

In this section, we consider estimation of the MESS model via the AGLASSO. The criterion function for the AGLASSO estimator is

$$\frac{1}{n}\hat{Q}_n(\theta) + \lambda_n \|\tilde{\zeta}_n\|^{-\mu}\|\zeta\|, \tag{19}$$

where $\lambda_n$ is a positive tuning parameter, $\tilde{\zeta}_n$ is an initial consistent estimator of $\zeta$, $\|\cdot\|$ denotes the Euclidean norm ($l_2$-norm), and $\mu$ is some positive number such as 1 or 2 as in the literature. The AGLASSO estimator $\hat{\theta}_n$ minimizes (19). Since the irregular phenomenon appears when $\zeta_0 = 0$, $\zeta$ is penalized in a group with an $l_2$-norm and there is no need to penalize other parameters. In addition, we are not interested in whether or not an individual component of $\zeta$ is zero, so a penalty term with an $l_1$-norm is not needed. The $\tilde{\zeta}_n$ can be the feasible N2SLS estimator in Section 2. Intuitively, $\tilde{\zeta}_n$ is small when $\zeta_0 = 0$, so the penalty term is large, and $\hat{\zeta}_n$ tends to be closer to zero. Otherwise, the effect of the penalty term is small.

**Assumption 10.** $\tilde{\zeta}_n = \zeta_0 + o_p(1)$.

## 4.1 Asymptotic properties

We study the asymptotic properties of the AGLASSO estimator in this subsection. the N2SLS estimator of $\beta$ has an explicit form for a given $\alpha$, so only the parameter space of $\alpha$ is assumed to be compact in Assumption 4. This is not the case for the AGLASSO estimator, so we make the following slightly stronger assumption.

**Assumption 4′.** *The true parameter vector $\theta_0$ is in the interior of the compact parameter space $\Theta$ for $\theta$.*

Assumption 11 is needed for the consistency of $\hat{\theta}_n$.

**Assumption 11.** $\lambda_n > 0$ *and* $\lambda_n = o(1)$.

**Proposition 4.1.** *Under Assumptions 1–3, 4′, 5, 7, 10 and 11,* $\hat{\theta}_n = \theta_0 + o_p(1)$.

We are interested in whether $\hat{\zeta}_n$ is equal to 0 w.p.a.1. in the case that $\zeta_0 = 0$, i.e., the sparsity property. This cannot be deduced from Proposition 4.1. To establish that property, Assumption 12 is needed. It requires the penalty term to have at least certain order when $\zeta_0 = 0$.

**Assumption 12.** *If* $\zeta_0 = 0$, $n^{1/2}\lambda_n\|\tilde{\zeta}_n\|^{-\mu} \to \infty$ *w.p.a.1.*

If $\tilde{\zeta}_n$ is the N2SLS estimator, then $\tilde{\zeta}_n = O_p(n^{-1/4})$. Thus, $n^{1/2+\mu/4}\lambda_n \to \infty$. If $\lambda_n = O(n^{-c})$ for some $c > 0$, then $c < \frac{1}{2} + \mu/4$.

**Proposition 4.2.** *Under Assumptions 1–3, 4′, 5, 7 and 10–12, if* $\zeta_0 = 0$, *then* $\mathrm{P}(\hat{\zeta}_n = 0) \to 1$ *as* $n \to \infty$.

For $\Psi = (\alpha, \delta')'$, we have the following oracle property.

**Proposition 4.3.** *Under Assumptions 1–3, 4′, 5, 7, and 10–12, if* $\zeta_0 = 0$, *then*

$$\sqrt{n}(\hat{\Psi}_n - \Psi_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty}\frac{1}{n}\big\{\mathrm{E}[(-W_nX_n\delta_0, X_n)'F_n]\bar{\Pi}_n^{-1}\,\mathrm{E}[F_n'(-W_nX_n\delta_0, X_n)]\big\}^{-1}\big).$$

Proposition 4.3 shows that, when $\zeta_0 = 0$, the AGLASSO estimator has the asymptotic distribution as if we knew the true parameter vector $\zeta_0 = 0$. We also derive the asymptotic distribution of $\hat{\theta}_n$ when $\zeta_0 \neq 0$. For that purpose, we first derive the rate of convergence of $\hat{\theta}_n$ when $\zeta_0 \neq 0$.

**Proposition 4.4.** *Under Assumptions 1–3, 4′, 5, 7, 10 and 11, if $\zeta_0 \neq 0$, $\hat{\theta}_n = \theta_0 + O_p(n^{-1/2} + \lambda_n)$.*

When $\zeta_0 \neq 0$, $\lambda_n$ may affect the convergence rate of $\hat{\theta}_n$ to $\theta_0$ and also the asymptotic distribution of $\hat{\theta}_n$. In order to eliminate the possible impact of the penalty term, the proper rate of $\lambda_n$ convergent to zero will be needed. For that purpose, we maintain Assumption 13.

**Assumption 13.** $\lambda_n = o(n^{-1/2})$.

**Proposition 4.5.** *Under Assumptions 1–3, 4′, 5, 7, 10, 11 and 13, if $\zeta_0 \neq 0$, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty}\big\{\frac{1}{n}\,\mathrm{E}[(-W_n D_n \beta_0, D_n)' F_n]\bar{\Pi}_n^{-1}\,\mathrm{E}[F_n'(-W_n D_n \beta_0, D_n)]\big\}^{-1}\big).$$

Proposition 2.2 shows that, if $\lambda_n$ is small enough, the AGLASSO estimator has the same asymptotic distribution as the N2SLS estimator when $\zeta_0 \neq 0$.

For the sparsity property of $\hat{\zeta}_n$ when $\zeta_0 = 0$, Assumption 12 requires $\lambda_n$ to be large enough. When $\zeta_0 \neq 0$, Assumption 13 requires $\lambda_n$ to be small enough in order to make the bias resulting from the penalty term small. Given $\mu$ and the N2SLS estimator $\tilde{\zeta}_n$, $\lambda_n = O(n^{-1/2-\mu/8})$ satisfies these assumptions.

## 4.2 Selection of the tuning parameter

In this section, we propose to select the tuning parameter $\lambda_n$ by minimizing an information criterion and we show that this data-driven procedure can identify the true model consistently.

To make explicit the dependence of the AGLASSO estimator on the tuning parameter, denote $\hat{\theta}_\lambda = \arg\min_{\theta\in\Theta}[\frac{1}{n}\hat{Q}_n(\theta) + \lambda\|\tilde{\zeta}_n\|^{-\mu}\|\zeta\|]$. Let $\Lambda = [0, \lambda_{\max}]$ be an interval from which the tuning parameter $\lambda$ is selected, where $\lambda_{\max}$ is a finite positive number. We propose to select the tuning parameter $\lambda$ that minimizes the following information criterion:

$$h_n(\lambda) = \frac{1}{n}\hat{Q}_n(\hat{\theta}_\lambda) - f(\hat{\zeta}_\lambda)\Gamma_n, \tag{20}$$

where $f(\hat{\zeta}_\lambda) = 1$ if $\hat{\zeta}_\lambda = 0$ and $f(\hat{\zeta}_\lambda) = 0$ otherwise, and $\{\Gamma_n\}$ is a positive sequence. That is, given $\Gamma_n$, the selected tuning parameter is $\hat{\lambda}_n = \arg\min_{\lambda\in\Lambda} h_n(\lambda)$. While $\frac{1}{n}\hat{Q}_n(\hat{\theta}_\lambda)$ measures the fit of the model, the term $f(\hat{\zeta}_\lambda)\Gamma_n$ gives extra bonus to setting $\zeta$ to zero. We take $\tilde{\zeta}_n$ to be the N2SLS estimator in Section 2. To guarantee model selection consistency, we make the following assumption.

**Assumption 14.** *$\Gamma_n > 0$, $\Gamma_n \to 0$ and $n^{1/2}\Gamma_n \to \infty$ as $n \to \infty$.*

To balance the requirements $\Gamma_n \to 0$ and $n^{1/2}\Gamma_n \to \infty$ in Assumption 14, we may take $\Gamma_n = O(n^{-1/4})$. Assumption 14 shows that the information criterion in (20) is different from the Akaike information criterion ($\Gamma_n = O(n^{-1})$), Bayesian information criterion ($\Gamma_n = O(n^{-1}\ln n)$) and Hannan-Quinn information criteria ($\Gamma_n =$

$O(n^{-1}\ln\ln n))$. This is because of the irregular convergence rate of the N2SLS estimator when $\zeta_0 = 0$. Let $\{\bar{\lambda}_n\}$ be an arbitrary sequence of tuning parameters which satisfy Assumptions 11–13, e.g., $\bar{\lambda}_n = n^{-1/2-\mu/8}$. Define $\Lambda_n = \{\lambda \in \Lambda : \hat{\zeta}_\lambda = 0 \text{ if } \zeta_0 \neq 0, \text{ and } \hat{\zeta}_\lambda \neq 0 \text{ if } \zeta_0 = 0\}$.

**Proposition 4.6.** *Under Assumptions 1–3, 4′, 5, 7, 10 and 14,* $\mathrm{P}(\inf_{\lambda\in\Lambda_n} h_n(\lambda) > h_n(\bar{\lambda}_n)) \to 1$ *as* $n \to \infty$.

Proposition 4.6 does not mean that the tuning parameter chosen by minimizing the information criterion in (20) must be $\bar{\lambda}_n$ because $h_n(\hat{\lambda}_n) \leq h_n(\bar{\lambda}_n)$. Instead, it means that any $\lambda$ that fails to identify the true model cannot be selected asymptotically as the optimal tuning parameter by the information criterion in (20), since it is less favorable than $\bar{\lambda}_n$. Consequently, the model selection consistency of our data-driven procedure is established.

### 4.3 Computation

In this section, we briefly discuss the computation of our group LASSO estimator. First, we note that, given $\alpha$ and $\zeta$, the AGLASSO estimator of $\delta$ is

$$\hat{\delta}_n(\alpha,\zeta) = (X_n'\hat{H}_nX_n)^{-1}X_n'\hat{H}_n[e^{\alpha W_n}y_n - (W_nX_n^{**}, Z_n)\zeta]. \tag{21}$$

Substituting $\hat{\delta}_n(\alpha,\zeta)$ into the AGLASSO criterion function yields the concentrated function:

$$L_n(\alpha,\zeta) = L_{n1}(\alpha,\zeta) + \lambda_n\|\tilde{\zeta}_n\|^{-\mu}\|\zeta\|,$$

where $L_{n1}(\alpha,\zeta) \equiv \frac{1}{n}\hat{Q}_n(\alpha,\hat{\delta}_n(\alpha,\zeta),\zeta) = \frac{1}{n}[\hat{\mathcal{M}}_ne^{\alpha W_n}y_n - \hat{\mathcal{M}}_n(W_nX_n^{**}, Z_n)\zeta]'[\hat{\mathcal{M}}_ne^{\alpha W_n}y_n - \hat{\mathcal{M}}_n(W_nX_n^{**}, Z_n)\zeta]$ with $\hat{\mathcal{M}}_n = \hat{\Pi}_n^{-1/2}F_n'[I_n - X_n(X_n'\hat{H}_nX_n)^{-1}X_n'\hat{H}_n]$. Note that $L_n(\alpha,\zeta)$ is an AGLASSO criterion function in the least squares framework for a given $\alpha$, then we can directly apply the algorithms for computing the usual group LASSO.[14] Let $\hat{\zeta}_n(\alpha)$ be the AGLASSO estimator of $\zeta$ for a given $\alpha$. Then $\hat{\alpha}_n$ can be obtained by minimizing $L_n(\alpha,\hat{\zeta}_n(\alpha))$.

## 5 Monte Carlo simulations

In this section, we conduct some Monte Carlo experiments to investigate the finite sample performance of the N2SLS estimator, the AGLASSO estimator and the test statistics for the MESS model.

The experimental design is as follows. The DGP is the following model:

$$e^{\alpha W_n}Y_n = X_{n1}\beta_1 + l_n\beta_2 + W_nX_{n1}\beta_3 + Z_n\beta_4 + V_n, \quad V_n = e^{-\varsigma W_n}\epsilon_n, \tag{22}$$

Two row-normalized spatial weights matrices are considered: one is based on the queen criterion and the other on the rook criterion. The spatial units are assumed to be located on a square grid at locations $\{(r,s) : r,s = 1,\dots,k\}$, thus the total number of units is $n = k^2$. The distance $d_{n,rs}$ between two spatial units $r$ and $s$ is given by the Euclidean distance. The $X_{n1}$ contains an exogenous variable drawn independently from $N(0,1)$. we consider

[14] In our Monte Carlo simulations, we use the Matlab package SLEP (Liu et al., 2009), which implements an efficient algorithm based on the accelerated gradient method in Liu and Ye (2010).

the case with one endogenous variable $Z_n = (z_{n1}, \ldots, z_{nn})'$, where $z_{ni} = \bar{z}_{ni} + u_{ni}$. The $\bar{z}_{ni}$ is an exogenous variable consisting of independent draws from $N(0, \sigma_0^2)$, and $(\epsilon_{ni}, u_{ni})$'s are independent draws from the bivariate normal distribution $N\left(0, \sigma_0^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$, where $\sigma_0^2$ is chosen such that $R^2 = 0.2$ or $0.8$, for $R^2 = \mathrm{var}(X_{n1}\beta_{10} + W_n X_{n1}\beta_{30})/[\mathrm{var}(X_{n1}\beta_{10} + W_n X_{n1}\beta_{30}) + \sigma_0^2]$. We set $\delta_0 = (\beta_{10}, \beta_{20})'$ to $(1,1)'$. The true parameter $\zeta_0 = (\beta_{30}, \beta_{40})'$ is $(0,0)'$, $(1,1)'$ or $(0,1)'$. The $\alpha_0$ is set to either $-0.2$ or $-1$, and $\varsigma_0$ is set to $-0.2$.[15] We use the IV matrix $[l_n, X_{n1}, W_n X_{n1}, W_n^2 X_{n1}, \bar{Z}_n, W_n \bar{Z}_n]$ with $\bar{Z}_n = (\bar{z}_{n1}, \ldots, \bar{z}_{nn})'$ in the estimation. Following Kelejian and Prucha (2007), in the spatial HAC estimation, we use the distance $d_n = [n^{1/4}]$, where $[r]$ denotes the nearest integer that is less than or equal to $r$, and the kernel function $\mathcal{K}(x) = (1-x)^2 I(0 \leq x \leq 1)$, which guarantees that the HAC estimator is positive semi-definite in finite samples. We focus on the case with no measure errors for the distances, so $d^*_{n,rs} = d_{n,rs}$. The nominal size of the tests is set to 0.05. For the investigation of the powers of the test statistics, the data are generated by MESS models with $\zeta$ values being $(1,0.5)'$, $(1,1)'$, $(1,1.5)'$, $(1,2)'$, $(1,2.5)'$, or $(1,3)'$. The tuning parameter $\lambda$ for the AGLASSO is selected by minimizing the information criterion (20) with $\Gamma_n = 0.06n^{-1/4}$.[16] We consider two sample sizes: $n = 144$ or 400. The number of Monte Carlo repetitions is 2,000.

To compare distributions of the N2SLS estimators in the regular and irregular cases with normal distributions, we first studentize estimators so that they have mean zero and unit variance and then plot in Figures 1–3 (solid line) their kernel density estimates, based on normal kernel functions with optimal bandwidths. The estimators are for the case with the rook matrix, $R^2 = 0.8$, $\alpha_0 = -0.2$, and $n = 400$.[17] The dashed lines represent the standard normal probability density function (PDF). Figure 1 shows the irregular case with $\zeta_0 = 0$. While the density estimates for $\beta_1$ and $\beta_4$ are close to the standard normal, those for $\alpha$, $\beta_2$ and $\beta_3$ show obvious deviations from the normal distribution. In particular, the density estimate for $\beta_3$ has an obvious bimodal behavior. For the regular case with $\zeta_0 = (1,1)'$, Figure 2 shows that all density estimates are close to the standard normal. For Figure 3, while the Durbin regressor is irrelevant, the endogenous explanatory variable is relevant ($\zeta_0 = (0,1)'$). As mentioned earlier, this is a regular case. We observe that all density estimates are close to the standard normal PDF.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[15] The MESS model and the SAR model have similar interpretations in some sense. The chosen values $-0.2$ and $-1$ in a MESS process correspond to low and high degrees of spatial dependence in the SAR model (LeSage and Pace, 2007; Debarsy et al., 2015).

[16] In theory, the information criterion (20) can achieve model selection consistency as long as $\Gamma_n$ satisfies the order requirement in Assumption 14. But the finite sample performance depends on the choice of $\Gamma_n$. From the proof of Proposition 4.6, when $\zeta_0 = 0$, $\Gamma_n$ should be larger than the difference of the N2SLS criterion function values divided by $n$ at the N2SLS estimate and at the restricted N2SLS estimate. For the queen matrix, $n = 144$, $R^2 = 0.2$ and $\alpha_0 = -1$, we compute the difference 1000 times, and set $\Gamma_n = cn^{-1/2}$ to be the sample mean plus 2 times the standard errors, which yields $c = 0.06$. We then set $\Gamma_n = 0.06n^{-1/2}$ in all cases and for all sample sizes.

[17] For other cases, the figures are similar, so they are omitted.

Table 1 presents the probabilities that the AGLASSO estimator selects the right model, i.e., the proportions of Monte Carlo repetitions where the AGLASSO estimate $\hat{\zeta}_n = 0$ when $\zeta_0 = 0$, or $\hat{\zeta}_n \neq 0$ when $\zeta_0 \neq 0$. For the irregular cases with $\zeta_0 = 0$, when $n = 144$ and $R^2 = 0.2$, about 80%–90% of $\hat{\zeta}_n$ estimates are zero; when the sample size increases to 400 but $R^2$ is still 0.2, the probabilities of correct model selection increase to higher than 96%; when $R^2$ increases to 0.8, more $\hat{\zeta}_n$ estimates are zero. The cases with the rook matrix have higher probabilities of correct model selection. The impact of $\alpha_0$ on the probabilities of correct model selection is ambiguous. When $\zeta_0 = (1,1)'$ or $\zeta_0 = (0,1)'$, $\hat{\zeta}_n$ estimates are 100% nonzero. Hence, in finite samples, the AGLASSO estimator may not select the right model with a positive probability, but the probability of making mistakes decreases as the sample size increases, which is consistent with the asymptotic theory.

[Table 1 about here.]

To investigate the relevance of the asymptotic distributions of the N2SLS, the AGLASSO and also the restricted N2LS estimators with the restriction $\zeta = 0$ imposed (N2SLS-r), we report the ratios of the SE when $n = 144$ to that when $n = 400$ in Table 2. Asymptotically, the theoretical ratio for estimators with the $\sqrt{n}$-rate of convergence is 1.67, but that for those with the $n^{1/4}$-rate is 1.29. When $\zeta_0 = 0$, the N2SLS estimators of $\alpha$, $\beta_2$ and $\beta_3$ are only $n^{1/4}$-consistent, but those of $\beta_1$ and $\beta_4$ are $n^{1/2}$-consistent, the AGLASSO estimators of $\alpha$, $\beta_2$ and $\beta_3$ are $\sqrt{n}$-consistent, and the N2SLS-r estimators of $\alpha$, $\beta_2$ and $\beta_3$ are expected to converge to some limits with the $\sqrt{n}$-rate. In this case, Table 2 shows that, for the N2SLS, the ratios of $\alpha$, $\beta_2$ and $\beta_3$ fluctuate around 1.29 and are significantly smaller than 1.67, but those of $\beta_1$ and $\beta_4$ are around 1.67; for the N2SLS-r, the ratios of $\alpha$, $\beta_1$ and $\beta_2$ are close to 1.67; for the AGLASSO, the ratios for $\beta_1$ are slightly larger than 1.67, but those for $\alpha$ and $\beta_2$ are significantly larger than 1.67. The observed large ratios for the AGLASSO might be due to the fact that, in finite samples, the correct model selection probabilities are higher in cases with larger sample sizes. When $\zeta_0 = (1,1)'$ or $\zeta_0 = (0,1)'$, the reported ratios are around 1.67 in most cases, because the N2SLS and AGLASSO estimators are $\sqrt{n}$-consistent, and the N2SLS-r estimators converge to their limits with the $\sqrt{n}$-rate. Overall, the ratios in the tables are consistent with our asymptotic theory.

[Table 2 about here.]

Tables 3–5 report the biases, standard errors (SEs) and coverage probabilities (CP) of 95% confidence intervals of the N2SLS, N2SLS-r and AGLASSO estimates when $n = 144$.[18] Table 3 shows the results in the irregular case with $\zeta_0 = 0$. We first focus on the N2SLS. Biases and SEs for $\alpha$, $\beta_2$ and $\beta_3$ are relatively larger than those for $\beta_1$ and $\beta_4$.[19] Specifically, while the biases for $\beta_1$ and $\beta_4$ are all smaller or equal to 0.058 and 0.005 respectively, and the SEs for $\beta_1$ and $\beta_4$ are all smaller or equal to 0.204 and 0.088 respectively, the biases and SEs for $\alpha$, $\beta_2$ and

[18]Results for $n = 400$ are reported in the supplementary file. They have patterns similar to those in Tables 3–5, but as expected we observe smaller biases and SEs and generally more accurate CPs in corresponding cases. Because the asymptotic distributions of the N2SLS estimators in the irregular case are very complicated, their confidence intervals are simulated, which are obtained by sampling from the asymptotic distribution in Corollary 2.1 1000 times and taking the 2.5% and 97.5% quantiles.

[19]Recall that the N2SLS estimators of $\alpha$, $\beta_2$ and $\beta_3$ only have the $n^{1/4}$-rate of convergence.

$\beta_3$ are usually several times larger than those of $\beta_1$ and $\beta_4$. The cases with $R^2 = 0.8$ have smaller biases and SEs than those with $R^2 = 0.2$ in most cases. Comparing cases with the queen matrix and the rook matrix, those with the rook matrix have smaller SEs, but they only have smaller biases in some cases. The impact of $\alpha_0$ on the bias is ambiguous. In general, cases with $\alpha_0 = -1$ have larger SEs than those with $\alpha_0 = -0.2$. The CPs are all close to 95% except those for $\beta_2$, which are only around 80% for cases with the queen matrix and $R^2 = 0.2$. Since the N2SLS-r estimator has imposed the right restriction, it has smaller bias and smaller SE than those of the N2SLS estimator in all cases. For the AGLASSO, due to a positive probability of making mistakes in model selection as seen from Table 1, its bias and SE are between those of the N2SLS-r and N2SLS, but they are generally significantly smaller than those of the N2SLS. Since the AGLASSO selects the right model with higher probabilities for cases with $R^2 = 0.8$ than for those with $R^2 = 0.2$, its bias and SE are closer to those of the N2SLS-r in those cases. The CPs for the N2SLS-r are all close to 95%, and the CPs of the AGLASSO are closer to 95% than those for the N2SLS in most cases.

[Table 3 about here.]

Table 4 presents results on biases, SEs and CPs in the regular case with $\zeta_0 = (1, 1)'$ and $n = 144$. The bias of the N2SLS estimator is smaller than or equal to 0.051 in all cases. Compared with the irregular case with $\zeta_0 = 0$ in Table 3, the biases of the N2SLS estimators are significantly smaller except those for $\beta_4$, the SEs of the N2SLS estimators for $\alpha$, $\beta_2$ and $\beta_3$ are significantly smaller, while those for $\beta_1$ and $\beta_4$ have similar magnitudes. It is still observed that the cases with a larger $R^2$ generally have smaller SEs, and the cases with the rook matrix generally have smaller SEs than those with the queen matrix. Since the N2SLS-r estimator has imposed the wrong restriction $\zeta = 0$, it has relatively large bias in all cases. As the AGLASSO estimates $\zeta$ as nonzero with probability one, it has the same bias and SE as the N2SLS estimator. The CPs for the N2SLS and AGLASSO are around 95% in all cases, but those of the N2SLS-r can be very low in some cases due to its large biases. Biases, SEs and CPs in the regular case with $\zeta_0 = (0, 1)'$ and $n = 144$ are reported in Table 5. Patterns similar to those for Table 4 are observed.

[Table 4 about here.]

[Table 5 about here.]

The empirical size and power properties of the distance difference test and gradient test are summarized in Table 6. The two tests have empirical sizes close to zero, but they have power larger than the nominal size in almost all cases. Thus the tests are conservative in finite samples.[20] For $n = 144$, the power generally increases as $\beta_4$ in the DGP increases. There are no clear patterns for the impacts of $R^2$ and the spatial weights matrix on the power. Cases with different $\alpha_0$ values have similar power. Except for several cases with the rook matrix and

[20]In the supplementary file, we also report the empirical size and power properties for the homoskedastic case without using the HAC estimator. The empirical sizes are close to the nominal 5% and the powers have similar patterns. Thus the size distortion may be largely due to the use of the HAC estimator.

$R^2 = 0.8$, the gradient test has larger power than that of the distance difference test. The power increases as the sample size increases from 144 to 400, thus the power are all close or equal to 1 when $n = 400$.

[Table 6 about here.]

# 6 Conclusion

In this paper, we consider estimation of the MESS model with both the Durbin and endogenous explanatory variables. As the disturbances of the MESS model are allowed to have heteroskedastic variances and spatial dependence of unknown form, the N2SLS estimation is employed and is a robust estimation method for such a general model. Optimal N2SLS estimation is feasible with a HAC estimated weighting matrix. For the N2SLS estimation, the model parameters are generally identifiable and the N2SLS estimator is consistent. If the true parameter vector for the Durbin and endogenous explanatory variables is nonzero, the N2SLS estimator has the usual $\sqrt{n}$-rate of convergence and is asymptotically normal; otherwise, only some components of the N2SLS estimator have the $\sqrt{n}$ convergence rate, while the remaining components have the $n^{1/4}$-rate, and the asymptotic distribution is nonstandard. Since the irrelevance of the Durbin and endogenous regressors causes the irregular phenomenon, in addition to estimation, it may be of interest to consider the tests of their irrelevance. We investigate the distance difference and gradient tests. These two tests can generally detect Pitman drifts with the rate $n^{-1/2}$. However, there is a direction with the rate $n^{-1/2}$ for which the tests have trivial power.

As an alternative estimation method based on the N2SLS estimation, we propose to estimate the MESS model and perform a model selection via the AGLASSO. We show that the proposed estimator has the oracle property under regularity conditions. As a result, the N2SLS estimator with penalty has the usual $\sqrt{n}$-rate of convergence and asymptotic normal distribution. By contrast, the N2SLS estimator has slower than the $\sqrt{n}$-rate of convergence and non-normal asymptotic distribution in an irregular case. The irregular case occurs when a component of the true parameter vector takes a certain value, but if the the component is restricted to be the true value in the N2SLS estimation, the irregular phenomenon disappears. Since the LASSO can perform simultaneous model selection and estimation and the proposed AGLASSO estimator has the oracle property, there is no irregular phenomenon in the AGLASSO estimator. The AGLASSO provides an alternative estimation strategy so there is no need to find the nonstandard asymptotic distribution of the N2SLS estimator and also a pre-test procedure may not be needed. We propose to select the tuning parameter in the AGLASSO estimation by minimizing an information criterion.

In Monte Carlo experiments, the N2SLS estimators of the parameters with only the $n^{1/4}$-rate of convergence in the irregular case have large biases and SEs, but the N2SLS estimators of all parameters in the regular case performs well. The AGLASSO estimators perform as well as the restricted N2SLS in the irregular case and as the unrestricted N2SLS in the regular case for a moderate sample size. The distance difference test and gradient test are undersized and are powerful for the sample sizes considered. Thus, for estimation, the N2SLS estimates should not be used directly and we suggest the AGLASSO method. If one is willing to implement a pre-test with the distance difference or gradient test, then the a further estimation of the restricted model is needed if the the null hypothesis of no Durbin and endogenous regressors is not rejected.

In the case that the endogenous explanatory variables in the MESS models are equal to a non-linear function of some exogenous variables plus disturbance terms, there is a many IV issue in the N2SLS estimation, which is an interesting research question to explore in the future.[21] Instead of unknown spatial dependence in disturbances, if a spatial dependence process is specified for the disturbances, we might also consider the GMM estimation that explores, in addition to the linear moments, quadratic moments that capture spatial dependence (Lee, 2007), even unknown heteroskedastic variances remain unspecified. The additional quadratic moments can make the expected Jacobian matrix of the moment vector have full rank. Thus a proper GMM estimator with additional higher order moments may have the usual $\sqrt{n}$-rate of convergence and asymptotic normal distribution. It is of interest to find the best moment conditions. Furthermore, in the event of homoskedastic or heteroskedastic disturbances with parametric variances, it is of interest on how to implement the best GMM estimation in practice.


## Acknowledgments

We are grateful to the editors Zongwu Cai, Yongmiao Hong and Cheng Hsiao and three anonymous referees for their insightful comments that lead to substantial improvement of this paper. We also would like to thank the seminar participants of the School of Economics, Huazhong University of Science and Technology, and the "International Symposium on Recent Developments in Econometric Theory with Applications" in Honor of Professor Takeshi Amemiya at the Xiamen University for helpful comments. First author gratefully acknowledges the financial support from the National Natural Science Foundation of China (No. 71501119), Shanghai Pujiang Program and Shanghai Chenguang Program.


# Appendix A Derivatives

The first order derivatives of $Q_n(\theta)$ are given in (4) and (5). The second order derivatives are:

$$\frac{\partial^2 Q_n(\theta)}{\partial \alpha^2} = 2Y_n' e^{\alpha W_n'} W_n'^2 H_n (e^{\alpha W_n} Y_n - D_n\beta) + 2Y_n' e^{\alpha W_n'} W_n' H_n W_n e^{\alpha W_n} Y_n, \tag{23}$$

$$\frac{\partial^2 Q_n(\theta)}{\partial \alpha \partial \beta} = -2D_n' H_n W_n e^{\alpha W_n} Y_n, \tag{24}$$

$$\frac{\partial^2 Q_n(\theta)}{\partial \beta \partial \beta'} = 2D_n' H_n D_n. \tag{25}$$

The derivatives of $Q_n^*(\omega)$ are:

$$\frac{\partial Q_n^*(\omega)}{\partial \phi} = 2(W_n e^{\phi W_n} Y_n - W_n X_n \delta_0)' H_n V_n(\omega),$$

$$\frac{\partial Q_n^*(\omega)}{\partial \psi} = -2D_n' H_n V_n(\omega),$$

[21]For the SAR model, the many IV issue has been studied in Liu and Lee (2013) and Jin and Lee (2013).

$$\frac{\partial^2 Q_n^*(\omega)}{\partial \phi^2} = 2(W_n^2 e^{\phi W_n} Y_n)' H_n V_n(\omega) + 2(W_n e^{\phi W_n} Y_n - W_n X_n \delta_0)' H_n (W_n e^{\phi W_n} Y_n - W_n X_n \delta_0),$$
$$\frac{\partial^2 Q_n^*(\omega)}{\partial \phi \partial \psi} = -2D_n' H_n (W_n e^{\phi W_n} Y_n - W_n X_n \delta_0),$$
$$\frac{\partial^2 Q_n^*(\omega)}{\partial \psi \partial \psi'} = 2D_n' H_n D_n,$$
$$\frac{\partial^3 Q_n^*(\omega)}{\partial \phi^3} = 2(W_n^3 e^{\phi W_n} Y_n)' H_n V_n(\omega) + 6(W_n^2 e^{\phi W_n} Y_n)' H_n (W_n e^{\phi W_n} Y_n - W_n X_n \delta_0),$$
$$\frac{\partial^3 Q_n^*(\omega)}{\partial \phi^2 \partial \psi} = -2D_n' H_n W_n^2 e^{\phi W_n} Y_n,$$
$$\frac{\partial^4 Q_n^*(\omega)}{\partial \phi^4} = 2(W_n^4 e^{\phi W_n} Y_n)' H_n V_n(\omega) + 8(W_n^3 e^{\phi W_n} Y_n)' H_n (W_n e^{\phi W_n} Y_n - W_n X_n \delta_0) + 6(W_n^2 e^{\phi W_n} Y_n)' H_n W_n^2 e^{\phi W_n} Y_n,$$
$$\frac{\partial^4 Q_n^*(\omega)}{\partial \phi^3 \partial \psi} = -2D_n' H_n W_n^3 e^{\phi W_n} Y_n,$$
$$\frac{\partial^5 Q_n^*(\omega)}{\partial \phi^5} = 2(W_n^5 e^{\phi W_n} Y_n)' H_n V_n(\omega) + 10(W_n^4 e^{\phi W_n} Y_n)' H_n (W_n e^{\phi W_n} Y_n - W_n X_n \delta_0) + 20(W_n^3 e^{\phi W_n} Y_n)' H_n W_n^2 e^{\phi W_n} Y_n,$$
$$\frac{\partial^5 Q_n^*(\omega)}{\partial \phi^4 \partial \psi} = -2D_n' H_n W_n^4 e^{\phi W_n} Y_n.$$

Other unlisted derivatives with order equal to or smaller than five are equal to zero. Specifically, as $\frac{\partial^2 Q_n^*(\omega)}{\partial \psi \partial \psi'}$ does not depend on $\omega$, any additional derivatives for this derivative are zero. Hence, by Lemma 1 in the supplementary file, with $\zeta_0 = 0$, we have:

$$\frac{\partial Q_n^*(\omega_0)}{\partial \phi} = 2V_n' W_n' H_n V_n = O_p(1),$$
$$\frac{\partial Q_n^*(\omega_0)}{\partial \psi} = -2D_n' H_n V_n = O_p(\sqrt{n}),$$
$$\frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi^2} = 2(X_n\delta_0 + V_n)' W_n'^2 H_n V_n + 2V_n' W_n' H_n W_n V_n = 2(W_n^2 X_n \delta_0)' H_n V_n + O_p(1) = O_p(\sqrt{n}),$$
$$\frac{\partial^2 Q_n^*(\omega_0)}{\partial \phi \partial \psi} = -2D_n' H_n W_n V_n = O_p(\sqrt{n}),$$
$$\frac{\partial^2 Q_n^*(\omega_0)}{\partial \psi \partial \psi'} = 2D_n' H_n D_n = O_p(n),$$
$$\begin{aligned}\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^3} &= 2(X_n\delta_0 + V_n)' W_n'^3 H_n V_n + 6(X_n\delta_0 + V_n)' W_n'^2 H_n W_n V_n = 2(W_n^3 X_n \delta_0)' H_n V_n + 6(W_n^2 X_n\delta_0)' H_n W_n V_n + O_p(1)\\ &= O_p(\sqrt{n}),\end{aligned}$$
$$\frac{\partial^3 Q_n^*(\omega_0)}{\partial \phi^2 \partial \psi} = -2D_n' H_n W_n^2 (X_n\delta_0 + V_n) = -2D_n' H_n W_n^2 X_n \delta_0 + O_p(n^{1/2}) = O_p(n),$$
$$\begin{aligned}\frac{\partial^4 Q_n^*(\omega_0)}{\partial \phi^4} &= 2(X_n\delta_0 + V_n)' W_n'^4 H_n V_n + 8(X_n\delta_0 + V_n)' W_n'^3 H_n W_n V_n + 6(X_n\delta_0 + V_n)' W_n'^2 H_n W_n^2 (X_n\delta_0 + V_n)\\ &= 6(W_n^2 X_n\delta_0)' H_n W_n^2 X_n \delta_0 + O_p(n^{1/2})\\ &= O_p(n),\end{aligned}$$
$$\frac{\partial^4 Q_n^*(\omega_0)}{\partial \phi^3 \partial \psi} = -2D_n' H_n W_n^3 (X_n\delta_0 + V_n) = -2D_n' H_n W_n^3 X_n \delta_0 + O_p(n^{1/2}) = O_p(n),$$

$\dfrac{\partial^5 Q_n^*(\omega)}{\partial \phi^5}$ and $\dfrac{\partial^5 Q_n^*(\omega)}{\partial \phi^4 \partial \psi}$ are of order $O_p(n)$ uniformly in $\omega$.

Orders of derivatives

| | | | | |
|---|---|---|---|---|
| $\frac{\partial Q_n^*(\omega_0)}{\partial\phi} = O_p(1)$ | $\frac{\partial Q_n^*(\omega_0)}{\partial\psi} = O_p(\sqrt{n})$ | $\frac{\partial^2 Q_n^*(\omega_0)}{\partial\phi^2} = O_p(\sqrt{n})$ | $\frac{\partial^2 Q_n^*(\omega_0)}{\partial\phi\partial\psi} = O_p(\sqrt{n})$ | $\frac{\partial^2 Q_n^*(\omega_0)}{\partial\psi\partial\psi'} = O_p(n)$ |
| $\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^3} = O_p(\sqrt{n})$ | $\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi^2\partial\psi} = O_p(n)$ | $\frac{\partial^3 Q_n^*(\omega_0)}{\partial\phi\partial\psi\partial\psi'} = 0$ | $\frac{\partial^4 Q_n^*(\omega_0)}{\partial\phi^4} = O_p(n)$ | $\frac{\partial^4 Q_n^*(\omega_0)}{\partial\phi^3\partial\psi} = O_p(n)$ |
| $\frac{\partial^5 Q_n^*(\omega)}{\partial\phi^5} = O_p(n)$ | $\frac{\partial^5 Q_n^*(\omega)}{\partial\phi^4\partial\psi} = O_p(n)$ | | | |

# References


Anselin, L., 1988. Spatial Econometrics: Methods and Models. Boston: Kluwer Academic Publishers.

Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review 26, 153–166.

Brock, W.A., Durlauf, S.N., 2001. Interactions-based models, in: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, North-Holand, Amsterdam. pp. 3297–3380.

Chiu, T.Y.M., Leonard, T., Tsui, K.W., 1996. The matrix-logarithmic covariance model. Journal of the American Statistical Association 91, 198–210.

Cox, D.R., Hinkley, D.V., 1974. Theoretical Statistics. Chapman and Hall, London.

Debarsy, N., Jin, F., Lee, L.F., 2015. Large sample properties of the matrix exponential spatial specification with an application to FDI. Journal of Econometrics 188, 1–21.

Dovonon, P., Renault, E., 2013. Testing for common conditionally heteroskedastic factors. Econometrica 81, 2561–2586.

Elhorst, J.P., 2010. Applied spatial econometrics: Raising the bar. Spatial Economic Analysis 5, 9–28.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.

Han, X., Lee, L.F., 2013. Model selection using J-test for the spatial autoregressive model vs. the matrix exponential spatial model. Regional Science and Urban Economics 43, 250–271.

Jin, F., Lee, L.F., 2013. Generalized spatial two stage least squares estimation of spatial autoregressive models with autoregressive disturbances in the presence of endogenous regressors and many instruments. Econometrics 1, 71–114.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99–121.

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

Kelejian, H.H., Prucha, I.R., 2007. HAC estimation in a spatial framework. Journal of Econometrics 140, 131–154.

Kiefer, N.M., 1982. A remark on the parameterization of a model for heterogeneity. Working paper no 278. Department of Economics, Cornell University, Ithaca, NY.

Lee, L.F., 1993. Asymptotic distribution of the maximum likelihood estimator for a stochastic frontier function model with a singular information matrix. Econometric Theory 9, 413–430.

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

Lee, L.F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489–514.

Lee, L.F., Chesher, A., 1986. Specification testing when score test statistics are identically zero. Journal of Econometrics 31, 121–149.

LeSage, J.P., Pace, R.K., 2007. A matrix exponential spatial specification. Journal of Econometrics 140, 190–214.

Liao, Z., 2010. Adaptive GMM shrinkage estimation with consistent moment selection. Working paper.

Liu, J., Ji, S., Ye, J., 2009. SLEP: Sparse Learning with Efficient Projections. Arizona State University. `http://www.public.asu.edu/~jye02/Software/SLEP`.

Liu, J., Ye, J., 2010. Efficient L1/Lq norm regularization. http://arxiv.org/abs/1009.4766v1.

Liu, X., Lee, L.F., 2013. Two stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments. Econometric Reviews 32, 734–753.

Manski, C.F., 1993. Identification of endogenous social effects: The reflection problem. Review of Economic Studies 60, 531–542.

McManus, D.A., 1991. Who invented local power analysis? Econometric Theory 7, 265–268.

Qu, X., Lee, L.F., 2012. LM tests for spatial correlation in spatial models with limited dependent variables. Regional Science and Urban Economics 42, 430–445.

Rothenberg, T.J., 1971. Identification in parametric models. Eonometrica 39, 577–591.

Rotnitzky, A., Cox, D.R., Bottai, M., Robins, J., 2000. Likelihood-based inference with singular information matrix. Bernoulli 6, 243–284.

Sargan, J.D., 1983. Identification and lack of identification. Econometrica 51, 1605–1633.

Schmidt, P., Lin, T.F., 1984. Simple tests of alternative specifications in stochastic frontier models. Journal of Econometrics 24, 349–361.

Silvey, S.D., 1959. The Lagrangean multiplier test. Annals of Mathematical Statistics 30, 389–407.

Waldman, D.M., 1982. A stationary point for the stochastic frontier likelihood. Journal of Econometrics 18, 275–279.

Wang, H., Leng, C., 2007. Unified LASSO estimation by least squares approximation. Journal of the American Statistical Association 102, 1039–1048.

Wang, H., Leng, C., 2008. A note on adaptive group lasso. Computational Statistics and Data Analysis 52, 5277–5286.

Wang, H., Li, B., Leng, C., 2009. Shrinkage tuning parameter selection with a diverging number of parameters. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 71, 671–683.

Wang, H., Li, R., Tsai, C.L., 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika 94, 553–568.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with group variables. Journal of the Royal Statistical Society, Series B 68, 49–67.

Zhang, Y., Li, R., Tsai, C.L., 2010. Regularization parameter selections via generalized information criterion. Journal of the American Statistical Association 105, 312–323.

Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–1429.

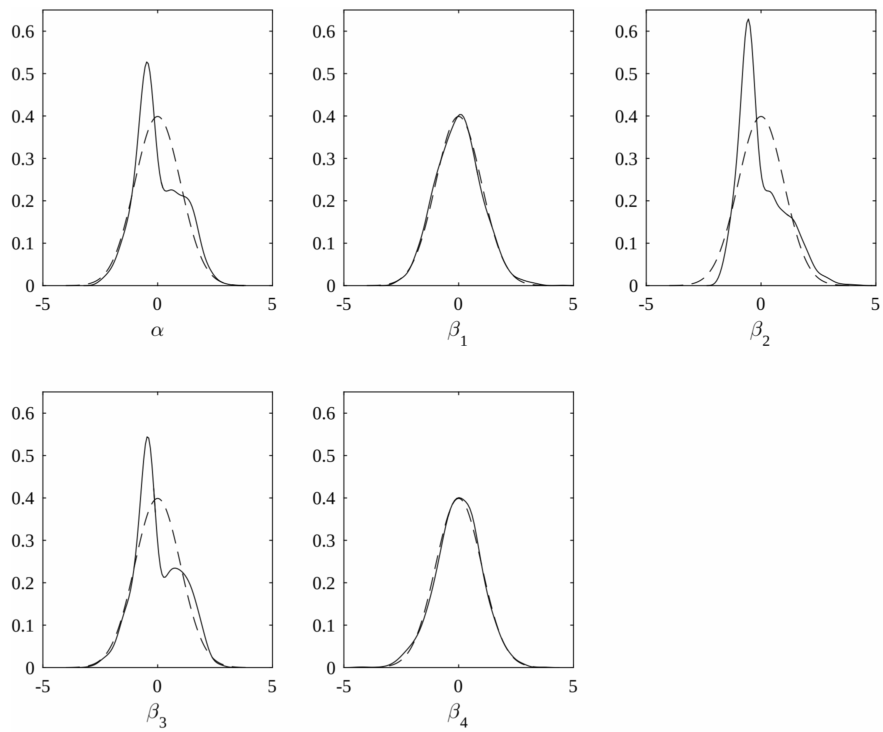


Figure 1: Kernel density estimates of the N2SLS estimators with $\zeta_0 = 0$ [Solid line: kernel density estimate; dashed line: standard normal PDF]

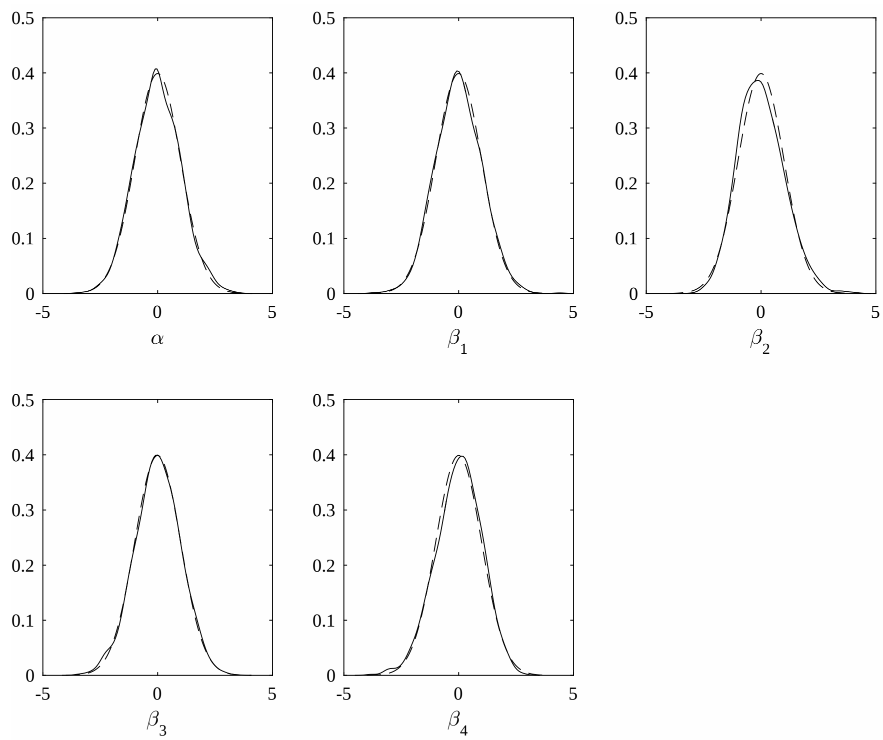


Figure 2: Kernel density estimates of the N2SLS estimators with $\zeta_0 = (1, 1)'$ [Solid line: kernel density estimate; dashed line: standard normal PDF]

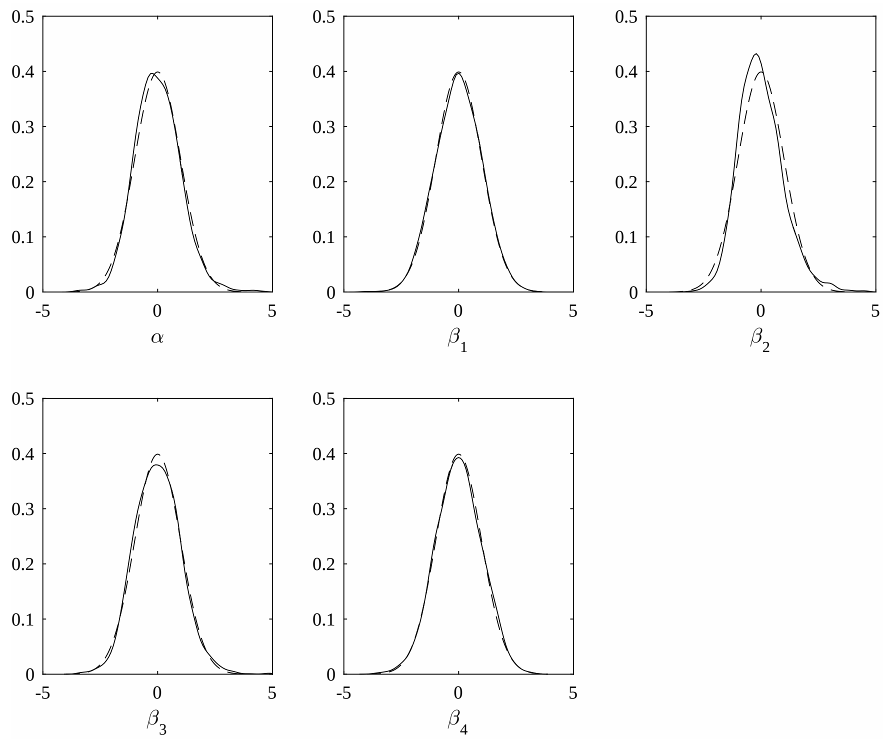


Figure 3: Kernel density estimates of the N2SLS estimators with $\zeta_0 = (0, 1)'$ [Solid line: kernel density estimate; dashed line: standard normal PDF]

Table 1: Probabilities that the AGLASSO estimator selects the right model

| | $n = 144$ | | | $n = 400$ | | |
|---|---|---|---|---|---|---|
| | $\zeta_0 = 0$ | $\zeta_0 = (1,1)'$ | $\zeta_0 = (0,1)'$ | $\zeta_0 = 0$ | $\zeta_0 = (1,1)'$ | $\zeta_0 = (0,1)'$ |
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 0.812 | 1.000 | 1.000 | 0.967 | 1.000 | 1.000 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | 0.847 | 1.000 | 1.000 | 0.973 | 1.000 | 1.000 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 0.884 | 1.000 | 1.000 | 0.987 | 1.000 | 1.000 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | 0.866 | 1.000 | 1.000 | 0.980 | 1.000 | 1.000 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 0.861 | 1.000 | 1.000 | 0.976 | 1.000 | 1.000 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | 0.886 | 1.000 | 1.000 | 0.983 | 1.000 | 1.000 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 0.900 | 1.000 | 1.000 | 0.989 | 1.000 | 1.000 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | 0.904 | 1.000 | 1.000 | 0.991 | 1.000 | 1.000 |

The numbers denote the proportions of Monte Carlo repetitions where the AGLASSO estimate $\hat{\zeta}_n = 0$ when $\zeta_0 = 0$, or $\hat{\zeta}_n \neq 0$ when $\zeta_0 \neq 0$. $\beta_{10} = 1$ and $\beta_{20} = 1$.

Table 2: Ratios of the SE when $n = 144$ to that when $n = 400$

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| $\zeta_0 = 0$ | | | | | |
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.221[1.547]1.596 | 1.748[1.713]1.805 | 0.993[1.560]1.624 | 1.248[—]— | 1.792[—]— |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.109[1.515]1.572 | 1.708[1.659]1.708 | 1.000[1.527]1.436 | 1.161[—]— | 1.701[—]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.126[1.641]1.657 | 1.572[1.643]1.677 | 0.969[1.775]1.709 | 1.199[—]— | 1.688[—]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.149[1.601]1.746 | 1.640[1.709]1.756 | 1.097[1.793]2.030 | 1.241[—]— | 1.667[—]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.261[1.633]2.226 | 1.655[1.670]1.739 | 1.236[1.634]2.374 | 1.243[—]— | 1.682[—]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.342[1.662]2.318 | 1.756[1.680]1.779 | 1.596[1.703]2.399 | 1.351[—]— | 1.753[—]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.352[1.585]2.507 | 1.702[1.706]1.764 | 1.422[1.607]2.673 | 1.374[—]— | 1.701[—]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.382[1.648]2.967 | 1.712[1.701]1.761 | 1.510[1.644]3.272 | 1.361[—]— | 1.765[—]— |
| $\zeta_0 = (1,1)'$ | | | | | |
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.764[1.675]1.764 | 1.782[1.698]1.782 | 2.391[2.550]2.391 | 1.679[—]1.679 | 1.675[—]1.675 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.704[1.640]1.704 | 1.747[1.741]1.747 | 1.830[1.863]1.830 | 1.676[—]1.676 | 1.700[—]1.700 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.691[1.826]1.691 | 1.683[1.752]1.683 | 1.789[2.176]1.789 | 1.626[—]1.626 | 1.778[—]1.778 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.928[1.815]1.928 | 1.722[1.731]1.722 | 1.944[2.220]1.944 | 1.720[—]1.720 | 1.718[—]1.718 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.744[1.722]1.744 | 1.696[1.737]1.696 | 1.798[1.782]1.798 | 1.666[—]1.666 | 1.777[—]1.777 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.750[1.647]1.750 | 1.689[1.661]1.689 | 1.729[1.678]1.729 | 1.674[—]1.674 | 1.692[—]1.692 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.722[1.760]1.748 | 1.719[1.640]1.719 | 1.714[1.739]1.720 | 1.740[—]1.765 | 1.725[—]1.727 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.774[1.712]1.774 | 1.740[1.646]1.740 | 1.817[1.736]1.817 | 1.742[—]1.742 | 1.712[—]1.712 |
| $\zeta_0 = (0,1)'$ | | | | | |
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.777[1.799]1.777 | 1.728[1.694]1.728 | 2.162[2.494]2.165 | 1.643[—]1.643 | 1.744[—]1.744 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.672[1.640]1.671 | 1.732[1.707]1.732 | 1.929[2.641]1.929 | 1.710[—]1.710 | 1.728[—]1.727 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | 1.759[1.741]1.759 | 1.638[1.609]1.638 | 1.817[1.970]1.817 | 1.671[—]1.671 | 1.665[—]1.665 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | 1.764[1.687]1.764 | 1.659[1.647]1.659 | 1.756[1.767]1.756 | 1.674[—]1.674 | 1.736[—]1.736 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.813[1.673]1.813 | 1.794[1.699]1.794 | 2.159[1.748]2.158 | 1.851[—]1.851 | 1.762[—]1.762 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.696[1.689]1.696 | 1.709[1.683]1.709 | 2.406[1.818]2.405 | 1.695[—]1.695 | 1.697[—]1.697 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | 1.812[1.666]1.812 | 1.716[1.659]1.716 | 1.855[1.711]1.855 | 1.808[—]1.808 | 1.763[—]1.763 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | 1.919[1.693]1.919 | 1.764[1.690]1.764 | 2.084[1.689]2.084 | 1.851[—]1.851 | 1.786[—]1.786 |

The numbers show the ratios of the SE when $n = 144$ to that when $n = 400$ in each case. The three numbers in each cell correspond to: N2SLS[N2SLS-r]AGLASSO. The ratios for the N2SLS-r estimates of $\beta_3$ and $\beta_4$ are not reported, because those estimates are restricted to zero. $\beta_{10} = 1$ and $\beta_{20} = 1$. The ratios for the AGLASSO estimates of $\beta_3$ and $\beta_4$ when $\zeta_0 = 0$ are not reported either, because Table 1 shows that those estimates are zero with very high probabilities.

Table 3: Biases, SEs and CPs when $\zeta_0 = 0$ and $n = 144$

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.592[0.861]0.969 | 0.058[0.204]0.990 | -0.227[0.779]0.764 | -0.419[0.858]0.993 | 0.005[0.088]0.982 |
| N2SLS-r | -0.172[0.446]0.985 | -0.004[0.179]0.981 | -0.068[0.532]0.917 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.272[0.609]0.973 | 0.004[0.198]0.980 | -0.097[0.656]0.875 | -0.157[0.516]— | 0.005[0.057]— |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.493[0.882]0.972 | 0.053[0.204]0.988 | -0.078[1.240]0.799 | -0.333[0.926]0.985 | 0.003[0.085]0.983 |
| N2SLS-r | -0.145[0.445]0.983 | -0.008[0.174]0.981 | -0.041[0.527]0.925 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.255[0.589]0.974 | -0.002[0.188]0.980 | -0.093[0.573]0.883 | -0.137[0.464]— | 0.003[0.052]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.147[0.613]0.985 | 0.046[0.195]0.987 | 0.050[0.867]0.888 | -0.151[0.674]0.995 | 0.003[0.085]0.988 |
| N2SLS-r | -0.039[0.311]0.995 | -0.009[0.178]0.978 | 0.010[0.409]0.965 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.083[0.418]0.982 | -0.005[0.188]0.979 | 0.000[0.517]0.932 | -0.068[0.331]— | 0.004[0.049]— |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.063[0.681]0.977 | 0.053[0.202]0.990 | 0.232[1.346]0.899 | -0.078[0.721]0.997 | 0.004[0.085]0.985 |
| N2SLS-r | -0.032[0.308]0.996 | -0.010[0.176]0.986 | 0.024[0.424]0.968 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.090[0.410]0.985 | -0.001[0.187]0.984 | 0.003[0.541]0.941 | -0.072[0.359]— | 0.004[0.050]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.173[0.510]0.968 | 0.020[0.049]0.987 | -0.036[0.565]0.880 | -0.123[0.518]0.967 | 0.004[0.087]0.983 |
| N2SLS-r | -0.010[0.124]0.968 | 0.001[0.044]0.979 | -0.003[0.135]0.965 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.082[0.310]0.916 | 0.006[0.047]0.979 | -0.039[0.285]0.895 | -0.059[0.268]— | 0.005[0.054]— |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.005[0.592]0.927 | 0.021[0.054]0.982 | 0.217[0.914]0.928 | 0.051[0.607]0.930 | 0.001[0.087]0.988 |
| N2SLS-r | -0.005[0.127]0.970 | -0.002[0.044]0.981 | 0.002[0.141]0.961 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.060[0.272]0.942 | 0.003[0.047]0.978 | -0.026[0.278]0.915 | -0.039[0.231]— | 0.004[0.048]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.069[0.372]0.958 | 0.017[0.050]0.990 | 0.151[0.469]0.968 | 0.063[0.366]0.983 | 0.001[0.085]0.984 |
| N2SLS-r | -0.002[0.081]0.982 | -0.001[0.044]0.978 | 0.001[0.097]0.976 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.018[0.176]0.951 | 0.003[0.047]0.979 | -0.002[0.209]0.946 | -0.012[0.160]— | 0.005[0.049]— |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.121[0.376]0.938 | 0.017[0.050]0.989 | 0.213[0.503]0.978 | 0.112[0.356]0.979 | 0.002[0.087]0.980 |
| N2SLS-r | -0.003[0.083]0.988 | -0.001[0.044]0.976 | -0.001[0.098]0.977 | 0.000[0.000]— | 0.000[0.000]— |
| AGLASSO | -0.011[0.176]0.961 | 0.003[0.046]0.975 | 0.005[0.221]0.948 | -0.003[0.156]— | 0.004[0.050]— |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{10} = 1$ and $\beta_{20} = 1$.

Table 4: Biases, SEs and CPs when $\zeta_0 = (1,1)'$ and $n = 144$

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.004[0.282]0.989 | -0.001[0.202]0.987 | 0.051[0.535]0.969 | 0.012[0.652]0.979 | -0.007[0.089]0.988 |
| N2SLS-r | -0.120[0.476]0.999 | -0.031[0.390]0.982 | 0.012[0.961]0.967 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.004[0.282]0.989 | -0.001[0.202]0.987 | 0.051[0.535]0.969 | 0.012[0.652]0.978 | -0.007[0.089]0.988 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.011[0.264]0.993 | -0.003[0.197]0.981 | 0.023[0.401]0.973 | -0.028[0.628]0.980 | -0.003[0.090]0.984 |
| N2SLS-r | -0.131[0.453]1.000 | -0.047[0.392]0.974 | -0.031[0.743]0.966 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.011[0.264]0.993 | -0.003[0.197]0.981 | 0.023[0.401]0.973 | -0.028[0.628]0.980 | -0.003[0.090]0.984 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.004[0.174]0.994 | -0.002[0.202]0.986 | 0.011[0.328]0.979 | 0.004[0.396]0.990 | -0.009[0.091]0.990 |
| N2SLS-r | -0.090[0.367]1.000 | -0.035[0.415]0.975 | -0.024[0.679]0.977 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.004[0.174]0.994 | -0.002[0.202]0.986 | 0.011[0.328]0.979 | 0.004[0.396]0.990 | -0.009[0.091]0.990 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.002[0.191]0.994 | 0.011[0.201]0.989 | 0.013[0.336]0.975 | 0.015[0.416]0.991 | -0.008[0.092]0.982 |
| N2SLS-r | -0.086[0.362]1.000 | -0.034[0.410]0.972 | -0.022[0.678]0.974 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.002[0.191]0.994 | 0.011[0.201]0.989 | 0.013[0.336]0.975 | 0.015[0.416]0.991 | -0.008[0.092]0.982 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.017[0.206]0.983 | 0.002[0.054]0.983 | 0.005[0.224]0.979 | -0.023[0.293]0.984 | -0.002[0.091]0.979 |
| N2SLS-r | -0.624[0.209]0.459 | -0.049[0.101]0.965 | -0.451[0.132]0.342 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.017[0.206]0.983 | 0.002[0.054]0.983 | 0.005[0.224]0.979 | -0.023[0.293]0.984 | -0.002[0.091]0.979 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.012[0.208]0.985 | 0.002[0.054]0.982 | 0.008[0.220]0.974 | -0.009[0.294]0.983 | -0.004[0.089]0.983 |
| N2SLS-r | -0.635[0.206]0.444 | -0.053[0.102]0.958 | -0.459[0.128]0.320 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.012[0.208]0.985 | 0.002[0.054]0.982 | 0.008[0.220]0.974 | -0.009[0.294]0.983 | -0.004[0.089]0.983 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.004[0.124]0.988 | 0.001[0.058]0.987 | 0.005[0.141]0.985 | -0.005[0.145]0.984 | -0.006[0.087]0.987 |
| N2SLS-r | -0.559[0.224]0.574 | -0.099[0.114]0.905 | -0.412[0.149]0.445 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.004[0.126]0.988 | 0.001[0.058]0.987 | 0.005[0.141]0.985 | -0.005[0.147]0.983 | -0.006[0.087]0.987 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.006[0.126]0.985 | -0.002[0.058]0.988 | 0.002[0.143]0.983 | -0.007[0.144]0.987 | -0.003[0.086]0.984 |
| N2SLS-r | -0.552[0.211]0.574 | -0.097[0.113]0.918 | -0.412[0.140]0.430 | -1.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.006[0.126]0.985 | -0.002[0.058]0.988 | 0.002[0.143]0.983 | -0.007[0.144]0.987 | -0.003[0.086]0.984 |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{10} = 1$ and $\beta_{20} = 1$.

Table 5: Biases, SEs and CPs when $\zeta_0 = (0,1)'$ and $n = 144$

| | $\alpha$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ |
|---|---|---|---|---|---|
| queen, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.004[0.279]0.990 | 0.005[0.179]0.988 | 0.051[0.488]0.969 | 0.018[0.585]0.984 | -0.007[0.090]0.984 |
| N2SLS-r | 0.028[0.509]0.999 | -0.012[0.361]0.979 | 0.211[1.018]0.982 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.004[0.279]0.990 | 0.005[0.179]0.988 | 0.051[0.488]0.969 | 0.018[0.585]0.983 | -0.007[0.090]0.983 |
| queen, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | 0.005[0.269]0.991 | 0.007[0.177]0.985 | 0.037[0.421]0.977 | 0.013[0.592]0.988 | -0.006[0.090]0.983 |
| N2SLS-r | 0.028[0.493]1.000 | -0.004[0.359]0.979 | 0.175[1.122]0.978 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.005[0.269]0.991 | 0.007[0.177]0.985 | 0.037[0.421]0.977 | 0.013[0.592]0.988 | -0.006[0.090]0.982 |
| rook, $R^2 = 0.2$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | -0.002[0.181]0.995 | 0.005[0.176]0.986 | 0.018[0.301]0.979 | -0.001[0.376]0.989 | -0.007[0.088]0.989 |
| N2SLS-r | 0.016[0.349]1.000 | -0.002[0.353]0.980 | 0.101[0.650]0.983 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.002[0.181]0.995 | 0.005[0.176]0.986 | 0.018[0.301]0.979 | -0.001[0.376]0.989 | -0.007[0.088]0.989 |
| rook, $R^2 = 0.2$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.001[0.188]0.995 | 0.001[0.175]0.986 | 0.015[0.296]0.975 | 0.005[0.377]0.986 | -0.006[0.090]0.981 |
| N2SLS-r | 0.021[0.356]1.000 | -0.011[0.359]0.978 | 0.084[0.611]0.987 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.001[0.188]0.995 | 0.001[0.175]0.986 | 0.015[0.296]0.975 | 0.005[0.377]0.986 | -0.006[0.090]0.981 |
| queen, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.001[0.273]0.978 | 0.006[0.046]0.988 | 0.043[0.360]0.959 | 0.010[0.310]0.974 | -0.006[0.092]0.985 |
| N2SLS-r | 0.003[0.232]0.981 | -0.000[0.091]0.977 | 0.028[0.273]0.970 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.001[0.273]0.978 | 0.006[0.046]0.988 | 0.043[0.360]0.959 | 0.010[0.310]0.974 | -0.006[0.092]0.985 |
| queen, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.012[0.274]0.976 | 0.006[0.045]0.985 | 0.034[0.433]0.958 | -0.002[0.304]0.973 | -0.005[0.090]0.983 |
| N2SLS-r | 0.002[0.232]0.985 | 0.002[0.092]0.969 | 0.034[0.280]0.974 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.012[0.274]0.977 | 0.006[0.045]0.985 | 0.034[0.433]0.958 | -0.002[0.304]0.972 | -0.005[0.090]0.983 |
| rook, $R^2 = 0.8$, $\alpha_0 = -0.2$ | | | | | |
| N2SLS | 0.001[0.188]0.977 | 0.005[0.045]0.986 | 0.017[0.206]0.969 | 0.000[0.204]0.973 | -0.010[0.092]0.986 |
| N2SLS-r | 0.010[0.162]0.983 | -0.001[0.090]0.977 | 0.023[0.195]0.979 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | 0.001[0.188]0.977 | 0.005[0.045]0.986 | 0.017[0.206]0.969 | 0.000[0.204]0.973 | -0.010[0.092]0.986 |
| rook, $R^2 = 0.8$, $\alpha_0 = -1$ | | | | | |
| N2SLS | -0.002[0.195]0.980 | 0.005[0.046]0.986 | 0.017[0.225]0.972 | -0.002[0.206]0.976 | -0.009[0.093]0.987 |
| N2SLS-r | 0.008[0.163]0.981 | 0.002[0.090]0.976 | 0.019[0.193]0.979 | 0.000[0.000]— | -1.000[0.000]— |
| AGLASSO | -0.002[0.195]0.980 | 0.005[0.046]0.986 | 0.017[0.225]0.972 | -0.002[0.206]0.976 | -0.009[0.093]0.987 |

"N2SLS" denotes the unrestricted N2SLS estimator and "N2SLS-r" denotes the restricted N2SLS estimator with the restriction $\zeta = 0$ imposed. The three numbers in each cell are: bias[SE]CP. $\beta_{10} = 1$ and $\beta_{20} = 1$.

Table 6: Size and power of the distance difference and gradient tests

| | distance difference test | | | | | | | gradient test | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | size | power | | | | | | size | power | | | | | |
| | | (1) | (2) | (3) | (4) | (5) | (6) | | (1) | (2) | (3) | (4) | (5) | (6) |
| $W_n$, $R^2$, $\alpha_0$ | n=144 | | | | | | | | | | | | | |
| queen, 0.2, $-0.2$ | 0.000 | 0.225 | 0.640 | 0.796 | 0.844 | 0.861 | 0.875 | 0.001 | 0.521 | 0.929 | 0.974 | 0.992 | 0.993 | 0.993 |
| queen, 0.2, $-1$ | 0.000 | 0.205 | 0.653 | 0.781 | 0.839 | 0.869 | 0.860 | 0.001 | 0.494 | 0.936 | 0.976 | 0.992 | 0.994 | 0.994 |
| rook, 0.2, $-0.2$ | 0.001 | 0.276 | 0.663 | 0.792 | 0.840 | 0.860 | 0.868 | 0.000 | 0.408 | 0.897 | 0.969 | 0.986 | 0.989 | 0.995 |
| rook, 0.2, $-1$ | 0.000 | 0.259 | 0.677 | 0.802 | 0.848 | 0.869 | 0.869 | 0.000 | 0.410 | 0.900 | 0.978 | 0.989 | 0.991 | 0.991 |
| queen, 0.8, $-0.2$ | 0.000 | 0.211 | 0.643 | 0.773 | 0.849 | 0.861 | 0.877 | 0.000 | 0.302 | 0.754 | 0.919 | 0.968 | 0.984 | 0.987 |
| queen, 0.8, $-1$ | 0.000 | 0.226 | 0.642 | 0.798 | 0.843 | 0.864 | 0.863 | 0.001 | 0.303 | 0.760 | 0.909 | 0.972 | 0.982 | 0.991 |
| rook, 0.8, $-0.2$ | 0.000 | 0.313 | 0.703 | 0.822 | 0.868 | 0.873 | 0.879 | 0.000 | 0.044 | 0.362 | 0.684 | 0.867 | 0.941 | 0.964 |
| rook, 0.8, $-1$ | 0.000 | 0.330 | 0.705 | 0.834 | 0.867 | 0.863 | 0.891 | 0.000 | 0.052 | 0.359 | 0.690 | 0.864 | 0.933 | 0.963 |
| | n=400 | | | | | | | | | | | | | |
| queen, 0.2, $-0.2$ | 0.001 | 0.991 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.001 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 0.999 |
| queen, 0.2, $-1$ | 0.004 | 0.990 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.002 | 0.997 | 1.000 | 1.000 | 1.000 | 0.999 | 1.000 |
| rook, 0.2, $-0.2$ | 0.001 | 0.996 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.002 | 0.983 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| rook, 0.2, $-1$ | 0.001 | 0.993 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.002 | 0.986 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| queen, 0.8, $-0.2$ | 0.001 | 0.988 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.003 | 0.951 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| queen, 0.8, $-1$ | 0.001 | 0.992 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.001 | 0.932 | 0.999 | 1.000 | 1.000 | 1.000 | 1.000 |
| rook, 0.8, $-0.2$ | 0.000 | 0.996 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.000 | 0.427 | 0.950 | 0.998 | 1.000 | 1.000 | 1.000 |
| rook, 0.8, $-1$ | 0.001 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.003 | 0.412 | 0.948 | 0.997 | 1.000 | 1.000 | 1.000 |

For the power, (1), (2), (3), (4), (5) and (6) in the table mean that in the DGP $\zeta_0 = (1, 0.5)'$, $\zeta_0 = (1, 1)'$, $\zeta_0 = (1, 1.5)'$, $\zeta_0 = (1, 2)'$, $\zeta_0 = (1, 2.5)'$ and $\zeta_0 = (1, 3)'$, respectively. $\beta_{10} = 1$ and $\beta_{20} = 1$.