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 30, 2018

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, which have irregular asymptotic distributions, 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 for the adaptive group LASSO by minimizing an information criterion.

Keywords: matrix exponential spatial specification, unknown heteroskedasticity, nonlinear two-stage least squares, singular covariance matrix, irregular estimates and test statistics, 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).

*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).

The MESS model may provide estimates and inference on spatial effects 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. The MESS is mainly motivated by computational consideration (LeSage and Pace, 2007). 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. The flexible parameter space is in particular useful for models with high order spatial weights so that the QML can be easily implemented.[2] Debarsy et al. (2015) find that the QML estimator of the MESS model may also be robust to unknown heteroskedasticity while disturbances have either no spatial correlation or a similar MESS process. Those are nice properties not shared by the SAR model.

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 $W_n X_n$, 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 generates 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 endogenous and Durbin's regressors, in which unknown heteroskedasticity is allowed. In a limited information setting, where endogenous regressors are present in an equation but without the specification of explicit structural equations for them, a popular estimation method is the nonlinear two-stage least squares (N2SLS). The N2SLS estimation can be seen as a GMM estimation exploring linear moments. In the spatial econometric literature, while inclusion of the Durbin regressors in an SAR model may capture exogenous externality effects and relax 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 if columns of $W_n X_n$ are linearly independent with columns of X_n.[3] 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

[2] By contrast, for a high order SAR model, it is very difficult to impose tractable parameter space on stability for a QML estimation (Elhorst et al., 2012).

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

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 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. Such an irregular phenomenon appears also in a 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.[4] Some authors have studied 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 a 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 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 estimates for the exogenous and endogenous variables have the $\sqrt{n}$-rate of convergence, while the spatial dependence parameter estimate and those for the Durbin regressors have the $n^{1/4}$-rate of convergence.[5] The model we consider is one with spatially correlated data and elements of the gradient vector of the N2SLS criterion function 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

[4] 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).

[5] 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.

probabilities equal to 1/2, where p is the number of restrictions. The gradient test statistic, constructed using the Moore-Penrose pseudoinverse of the singular covariance matrix of the gradient vector, is asymptotically distributed as a chi-squared distribution with $p-1$ degrees of freedom. We also investigate local power properties of our test statistics. For the Pitman drift (McManus, 1991) with the order $O(n^{-1/2})$, 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 simultaneously parameter estimation and model selection. Since the irregular phenomenon is due to $\zeta_0 = 0$ but subject to uncertainty, we consider using the adaptive group LASSO (AGLASSO) that appends a penalty function of ζ 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.[6]

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.[7]

2 N2SLS estimator

The MESS model with the Durbin and endogenous explanatory variables is as follows:

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

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, where k_x is the number of exogenous variables including an intercept in this equation, and Z_n is an $n \times k_z$ matrix of endogenous variables. For disturbances, $V_n = (v_{ni})$ is an $n \times 1$ vector of innovations with mean zero and unknown heteroskedastic variances.

[6] 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).

[7] It is available at `http://econ.shufe.edu.cn/kindeditor-4.1.10/attached/file/20170109/20170109135722_32428.pdf`.

The spatial weights matrix W_n is an $n \times n$ matrix with all diagonal elements being zero. The W_n may be row-normalized or not. 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 implicitly written as $W_n l_n$ for the convenience of later analyses. The Durbin regressors $W_n X_{n1}$ can be seen as neighbors' characteristics to capture exogenous externality.[8] The $W_n l_n$, when W_n is not row-normalized with binary elements, is known as out-degrees, which measures the overall numbers of links for each individual.[9] The β_1, β_2, β_3 and β_4 are conformable parameter vectors. The matrix exponential $e^{\alpha W_n}$ with a scalar parameter α, 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 α. If the model is regarded as a game with complete information, the Nash equilibrium exists and is unique. Let the utility function of spatial unit i be $U_{ni}(Y_{ni}) = c_{ni} Y_{ni} - \frac{1}{2} Y_{ni}^2$, where c_{ni} is the ith row of $e^{-\alpha W_n}(X_n \beta + V_n)$ that gives benefit, and the quadratic term is a cost for i's action Y_{ni}. Then the maximization of $U_{ni}(Y_{ni})$ yields $Y_n = (Y_{n1}, \ldots, Y_{nn})' = e^{-\alpha W_n}(X_n \beta + V_n)$. For an SAR model, $e^{\alpha W_n} Y_n$ in (1) is replaced by $(I_n - \lambda W_n)$ for some scalar λ. 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 pattern of spatial dependence across spatial units, while the MESS model shows an exponential decay. If the economic theory implies an exponential decay, then the MESS model can provide a better description of real data. Pfaffermayr (2012) and Koch (2013) derive the MESS model for spatial income convergence from theory. Furthermore, the determinant $|e^{\alpha W_n}| = e^{\alpha \operatorname{tr}(W_n)} = 1$ as the diagonal elements of W_n are all zero. 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 $F_n = (f_{n,ij})$ be a full rank $n \times k_f$ instrumental variable (IV) matrix with k_f not smaller than the total number of coefficients. To allow for conditional heteroskedasticity, we assume that v_{ni}'s are independent conditional on F_n but can have different conditional variances. As the variance of $F_n' V_n$ conditional on F_n is $\Pi_n = F_n' \Sigma_n F_n$, where $\Sigma_n = \mathrm{E}(V_n V_n' | F_n)$ is a diagonal matrix of conditional variances, the criterion function $Q_n(\theta)$ of the infeasible N2SLS estimation, as if Σ_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 large sample properties of the infeasible N2SLS estimator $\check{\theta}_n$ that minimizes $Q_n(\theta)$ as if Π_n were known. A feasible version by the use of a White-type (White, 1980) consistent estimator for Π_n will be investigated in the last part of this section, which shows that asymptotic results remain valid

[8]Even when X_{n1} is a dummy variable, $W_n X_{n1}$ can be meaningful. For example, suppose that X_{n1} represents an individual's gender, which takes the value 1 if the individual is female, and takes 0 otherwise. Then if elements of W_n are 0 and 1, where 1 represents a friend link, each element of $W_n X_{n1}$ would represent the number of female friends; if W_n is row-normalized from the above matrix of binary elements, then the variable $W_n X_{n1}$ represents the share of female friends among male and female friends. In this case, $W_n X_{n1}$ captures the effect of female friends on outcome Y_n.

[9]In the case that W_n is not row-normalized and its elements are binary, if each spatial unit has the same number of neighbors, $W_n l_n$ is proportional to l_n, so it is absorbed in l_n and should not be included in the model. This is similar to the case of a row-normalized W_n. We thank an anonymous referee for pointing this out.

for the feasible version. The function $Q_n(\theta)$ is a generalized version of the N2SLS criterion function in Amemiya (1985). It is one as in the estimation of implicit nonlinear simultaneous equations (Amemiya, 1985, p. 255). Since the moment conditions are based on orthogonality of IVs with disturbances, $Q_n(\theta)$ can also be regarded as a GMM criterion function (Hansen, 1982).[10]

We first discuss some regularity conditions needed for the N2SLS estimation. Abbreviate "bounded in row and column sum norms" as "BRC". A typical assumption on the sequence of spatial weights matrices $\{W_n\}$ in spatial econometrics is that they are BRC. This assumption, originated in Kelejian and Prucha (1998, 1999), restricts the degree of spatial dependence. As in Kelejian and Prucha (2004), (1) can represent an equation in a system of spatially correlated equations, if $Z_n = A_{n1}\mathbb{X}_n\gamma + A_{n2}u_n$, where $\mathbb{X}_n$ denotes all the exogenous variables in the system, γ is a parameter matrix, A_{n1} and A_{n2} are $n \times n$ BRC nonstochastic matrices, and u_n is a matrix of error terms. The matrix A_{n1} may depend on the spatial weights matrix W_n such that, under regularity conditions, $A_{n1} = \sum_{i=0}^{\infty} \rho_i W_n^i$ for some scalars ρ_i's. Under this circumstance, the IV matrix F_n in the N2SLS estimation can be formed by the independent columns of $[\mathbb{X}_n, W_n\mathbb{X}_n, \dots, W_n^s\mathbb{X}_n]$ for some s. We may have similar IVs if one or more elements of Z_n are generated by a nonlinear model.[11] For our analyses, we allow X_n and F_n to be stochastic. It is only assumed that X_n, F_n and Z_n have bounded second moments in the sense that $\frac{1}{n}\|\operatorname{E}(X_{n1}'X_{n1})\| = O(1)$, $\frac{1}{n}\|\operatorname{E}(F_n'F_n)\| = O(1)$ and $\frac{1}{n}\|\operatorname{E}(Z_n'Z_n)\| = O(1)$ for the Euclidean matrix norm $\|\cdot\|$. The disturbances v_{ni}'s are independent conditional on F_n but can have different conditional variances. We assume that $f_{n,ij}v_{nk}$'s have uniformly bounded moments higher than the second order so that a central limit theorem can be applied in the analyses. Some regularity conditions are also needed so that laws of large numbers can be applied to terms involving X_n, Z_n or F_n. For those, we shall make high level assumptions in the main text and discuss low level ones via spatial near-epoch dependence in the supplementary file. The low level conditions allow spatial dependence in X_n, Z_n and F_n, which can be the case from the above discussion. For the N2SLS estimation, $\lim_{n\to\infty} \frac{1}{n}\operatorname{E}(\Pi_n)$ is assumed to be nonsingular. Although the reduced form of the model always exists for any value of α, we assume that α is bounded as in Debarsy et al. (2015). As a consequence, $e^{\alpha W_n}$ is BRC uniformly in α.[12] These basic assumptions are summarized below.

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

Assumption 2. *v_{ni}'s in $V_n = (v_{n1}, \dots, v_{nn})'$ are independent with mean zero conditional on F_n but can have different variances conditional on F_n. Furthermore, $\sup_n \sup_{1\le i,j,k\le n} \operatorname{E}|f_{n,ij}v_{nk}|^{\tau} < \infty$ for some $2 < \tau < \infty$, $\frac{1}{n}\|\operatorname{E}(F_n'F_n)\| = O(1)$, $\frac{1}{n}\|\operatorname{E}(X_{n1}'X_{n1})\| = O(1)$ and $\frac{1}{n}\|\operatorname{E}(Z_n'Z_n)\| = O(1)$ for the Euclidean matrix norm $\|\cdot\|$.*

Assumption 3. *$\frac{1}{n}\Pi_n - \frac{1}{n}\operatorname{E}(\Pi_n) = o_p(1)$, and $\lim_{n\to\infty} \frac{1}{n}\operatorname{E}(\Pi_n)$ exists and is nonsingular.*

[10] Since W_n has a zero diagonal, $|e^{\alpha W_n}| = e^{\alpha \operatorname{tr}(W_n)} = 1$ for any α. Then $e^{\alpha W_n}$ would not tend to a matrix whose elements are all close to zero when α goes to minus infinity. Thus the criterion function $Q_n(\theta)$ does not have a numerical problem similar to that for the Box-Cox model (Davidson and MacKinnon, 1993, pp. 243–244). We thank an anonymous referee for directing us to this problem.

[11] In this case, we may have a many IV problem as in Liu and Lee (2013), which is beyond the scope of this paper. More discussions on the IV selection can be found in Kelejian and Prucha (2007).

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

Assumption 4. *There exists a constant $\eta > 0$ such that $|\alpha| \leq \eta$ and the true parameter α_0 of α is in the interior of the parameter space $[-\eta, \eta]$.*

Using the reduced form of Y_n, we may write the moment condition $\frac{1}{n}F_n'(e^{\alpha W_n}Y_n - D_n\beta) = \frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}D_n\beta_0 + \frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}V_n - \frac{1}{n}F_n'D_n\beta$. For the identification of the true parameter vector θ_0 of θ, a law of large number is needed for $\frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}D_n$.

Assumption 5. $\frac{1}{n}F_n'e^{(\alpha-\alpha_0)W_n}D_n - \frac{1}{n}\,\mathrm{E}(F_n'e^{(\alpha-\alpha_0)W_n}D_n) = o_p(1)$ *for any* $\alpha \in [-\eta, \eta]$.

The identification of θ_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 θ_0. Under Assumption 2,

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

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

Assumption 6. $\lim_{n\to\infty}\frac{1}{n}[\mathrm{E}(F_n'e^{(\alpha-\alpha_0)W_n}D_n)\beta_0, \mathrm{E}(F_n'D_n)]$ *has full column rank for any* $\alpha \neq \alpha_0$.

Note that the condition of Assumption 6 implies, in particular, that $\lim_{n\to\infty}\frac{1}{n}\,\mathrm{E}(F_n'D_n)$ has full column rank. The condition holds in general except the case $\beta_0 = 0$. So implicitly we assume that $\beta_0 \neq 0$. Under the above regularity assumptions, the consistency of $\check{\theta}_n$ follows.

Proposition 2.1. *Under Assumptions 1–6,* $\check{\theta}_n = \theta_0 + o_p(1)$.

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 $G_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta'} = F_n'(W_n e^{\alpha W_n}Y_n, -D_n)$. For the convergence of this matrix, we make the following assumption.

Assumption 7. $\frac{1}{n}F_n'W_nD_n - \frac{1}{n}\,\mathrm{E}(F_n'W_nD_n) = o_p(1)$.

Let $\bar{G}_n = \mathrm{E}[G_n(\theta_0)]$. Then,

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

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. So ζ represents effects of contextual variables and endogenous regressors in both cases. If W_n is row-normalized and $\zeta_0 \neq 0$, the first column of $\bar{G}_n$ is generally not linearly dependent on $\mathrm{E}(F_n'D_n)$, since $\mathrm{E}(F_n'W_n^2X_{n1})\beta_{30}$ or $\mathrm{E}(F_n'W_nZ_n)\beta_{40}$ appears in the first column. If W_n is not row-normalized and $\zeta_0 \neq 0$, $\mathrm{E}(F_n'W_n^2l_n)\beta_{20}$ might also appear in the first column of $\bar{G}_n$. Thus $\bar{G}_n$ generally has full rank as long as some contextual variables or endogenous regressors have relevant effects. 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. We assume a condition on Z_n for the convergence of the Jacobian matrix $G_n(\theta)$.

Proposition 2.2. *Under Assumptions 1–7, when $\zeta_0 \neq 0$, i.e., some contextual variables or endogenous regressors have relevant effects, 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 $\bar{\Pi}_n = \mathrm{E}(\Pi_n)$, provided that $\mathrm{plim}_{n\to\infty}\frac{1}{n}\bar{G}_n$ has full rank. The best IV matrix F_n is the matrix formed by the independent columns of $\Sigma_n^{-1}[X_n^*, W_nX_n, W_n^2X_n, \mathrm{E}(Z_n, W_nZ_n|\mathbb{X}_n)]$, 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), when $\zeta_0 = 0$, i.e., when the Durbin regressors and endogenous explanatory variables are irrelevant, the Jacobian matrix of the moment vector at the true parameter vector is rank deficient w.p.a.1. In this subsection, we consider the N2SLS estimation of model (1) in this situation.

Although $\zeta_0 = 0$, the identification condition in Assumption 6 still holds when $\delta_0 \neq 0$, and the N2SLS estimator can be consistent. However, 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

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

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

Let k_{x^*} be the number of columns in X_n^*, and $\delta = (\delta_1', \delta_2)'$, where δ_2 is the last element of δ and δ_1 contains the remaining elements.[13] Then,

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

[13] Recall that δ, defined in Section 2.1, has different expressions for the cases with a row-normalized W_n and a non-row-normalized W_n.

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 (22)–(24) 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 might 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), for analytical convenience, 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 ϕ 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 θ_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 $\check{\omega}_n$ minimizes $Q_n^*(\omega)$. Because of the one-to-one correspondence between θ and ω, the consistency of the N2SLS estimator $\check{\theta}_n$ implies the consistency of $\check{\omega}_n$ to ω_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. For our model, we need a third order Taylor expansion of the first order conditions at the true parameter vector as that gives the leading terms. 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 8. $\lim_{n\to\infty} \frac{1}{n}[\mathrm{E}(F_n' W_n^2 X_n)\delta_0, \mathrm{E}(F_n' D_n)]$ *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 Lemma 2 in the supplementary file to $\frac{1}{\sqrt{n}} F_n' V_n$. 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 ω_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), $J_{2n} = L_n + (\frac{2}{n} D_n' H_n D_n)^{-1} \frac{1}{n} D_n' H_n W_n^2 X_n \delta_0 J_{1n}$. 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 preceding analysis, 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 $\Upsilon_{2n} = \frac{1}{\sqrt{n}}\,\mathrm{E}(F_n' W_n V_n \mathbb{J}_n')\Delta_n^{-1} = \mathrm{E}(F_n' W_n \Sigma_n F_n)\bar{\Pi}_n^{-1}\Upsilon_{1n}$.

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

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

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 and 8, 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 $K^* = 2\,\mathrm{plim}_{n\to\infty}(U_2 + \Upsilon_{2n}J)'(\frac{1}{n}\Pi_n)^{-1}(U_1 + \Upsilon_{1n}J) + \mathrm{plim}_{n\to\infty} J'\big[\big(\begin{smallmatrix} \frac{1}{3}(F_n' W_n^3 X_n \delta_0)'\Pi_n^{-1} U_1 + (F_n' W_n^2 X_n \delta_0)'\Pi_n^{-1} U_2 \\ -2(D_n' F_n)\Pi_n^{-1} U_2 \end{smallmatrix}\big) + \big(\begin{smallmatrix} (F_n' W_n^2 X_n \delta_0)'\Pi_n^{-1}\Upsilon_{2n} \\ -2(D_n' F_n)\Pi_n^{-1}\Upsilon_{2n} \end{smallmatrix}\big)J\big]$.

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 and 8, 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

Consider the following MESS model with Durbin regressors but without endogenous regressors:

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

Let $\zeta = \beta_3$ when W_n is row-normalized and $\zeta = (\beta_2, \beta_3')'$ when W_n is not row-normalized. Then model (12) has an irregular phenomenon when $\zeta_0 = 0$, similar to model (1). In either model (1) or (12), if the Durbin regressors are irrelevant and not included in the model, the Jacobian matrix of the moment vector at the true parameter vector in general has full rank w.p.a.1. On the other hand, in model (1), it is of interest to note that if the endogenous variables are really relevant, i.e., $\beta_{40} \neq 0$, even when the Durbin regressors are irrelevant, there is no irregular phenomenon, so the presence of relevant identifiable endogenous variables helps the identification of parameters including those for the Durbin regressors, which might be zero. Thus, it is the Durbin regressors with unknown zero coefficients but not endogenous explanatory variables that lead to the irregular phenomenon.

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 $\breve{\omega}_n$ 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 $(\breve{\alpha}_n, \breve{\beta}_{1n}', \breve{\beta}_{2n}, \breve{\beta}_{3n}')'$ 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 $Q_n(\theta)$ in (2) involves the unknown covariance matrix Σ_n. We consider a feasible N2SLS estimator using the White-type estimator for $\Pi_n = F_n' \Sigma_n F_n$.

For the formulation of a feasible N2SLS estimator, a consistent estimator for the covariance matrix $\frac{1}{n}\Pi_n$ can be derived as follows. First we can derive a 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 $\hat{V}_n = e^{\dot{\alpha}_n W_n} Y_n - D_n \dot{\beta}_n = (\hat{v}_{ni})$. The White-type estimator for Π_n is $\hat{\Pi}_n = F_n' \hat{\Sigma}_n F_n$, where $\hat{\Sigma}_n = \mathrm{diag}(\hat{v}_{n1}^2, \ldots, \hat{v}_{nn}^2)$. The initial estimator $\dot{\theta}_n$ can be only $n^{1/4}$-

consistent as in last subsection. Thus we assume that $n^{1/4}(\dot{\theta}_n - \theta_0) = O_p(1)$.[14] We also assume that all elements of X_n, Z_n and F_n have uniformly bounded fourth moments in order to show the consistency of $\frac{1}{n}\hat{\Pi}_n$ for $\frac{1}{n}\Pi_n$.

Assumption 9. *$n^{1/4}(\dot{\theta}_n - \theta_0) = O_p(1)$, $\sup_n \sup_{1\leq i\leq n, 1\leq j\leq k_f} \mathrm{E}(f_{n,ij}^4) < \infty$, $\sup_n \sup_{1\leq i\leq n, 1\leq j\leq k_d} \mathrm{E}(z_{n,ij}^4) < \infty$ and $\sup_n \sup_{1\leq i\leq n} \mathrm{E}(v_{ni}^4) < \infty$.*

With $\hat{\Pi}_n$, the feasible N2SLS estimator $\hat{\theta}_n$ is the minimizer of

$$\hat{Q}_n(\theta) = (e^{\alpha W_n} Y_n - D_n\beta)' \hat{H}_n (e^{\alpha W_n} Y_n - D_n\beta), \tag{13}$$

where $\hat{H}_n = F_n \hat{\Pi}_n^{-1} F_n'$. For 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, 2 and 9, $\frac{1}{n}\hat{\Pi}_n - \frac{1}{n}\Pi_n = o_p(1)$. With the additional Assumption 9, the results in Propositions 2.1 and 2.2 with $\check{\theta}_n$ replaced by $\hat{\theta}_n$ and the result in Proposition 2.3 with $\check{\omega}_n$ replaced by $\hat{\omega}_n$ still hold.*[15]

3 Testing for the irrelevance of the Durbin and endogenous regressors

In this section, we derive the distance difference and gradient tests for the irrelevance of the Durbin and endogenous variables, and also investigate their local power properties.[16]

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:

$$\sqrt{n}(\tilde{\Psi}_n - \Psi_0) = \big[\frac{1}{n}(-W_n X_n \delta_0, X_n)' H_n (-W_n X_n \delta_0, X_n)\big]^{-1} \frac{1}{\sqrt{n}}(-W_n X_n \delta_0, X_n)' H_n V_n + o_p(1). \tag{14}$$

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

[14]This includes also the case that an initial estimator happens to be $\sqrt{n}$-consistent, i.e., $n^{1/2}(\dot{\theta}_n - \theta_0) = O_p(1)$. In the proof of Proposition 2.4, we need only the property that $n^{1/4}(\dot{\theta}_n - \theta_0) = O_p(1)$. On the other hand, the consistency property that $\dot{\theta}_n = \theta_0 + o_p(1)$ alone would not be strong enough.

[15]With unknown Σ_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.

[16]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 $H_0 : \zeta_0 = 0$.

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

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

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 ϕ (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 corresponding 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 Ψ_0 and $\tilde{\Psi}_n$, and the second equality follows by (14) 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 $\hat{Q}_n(\theta_0) = \hat{Q}_n^*(\omega_0)$,

$$\begin{aligned}\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n) &= I(R_n < 0)V_n'(\mathbb{P}_{(-W^2X\delta_0, D)} - \mathbb{P}_{(-WX\delta_0, X)})V_n \\ &\quad + I(R_n > 0)V_n'(\mathbb{P}_D - \mathbb{P}_{(-WX\delta_0, X)})V_n + o_p(1).\end{aligned} \tag{15}$$

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 1–9, 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* $1/2$.[17]

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

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

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 order derivatives in Appendix A,

$$\begin{aligned}
\frac{1}{\sqrt{n}} \frac{\partial \hat{Q}_n(\tilde{\theta}_n)}{\partial \zeta} &= \frac{1}{\sqrt{n}} \frac{\partial \hat{Q}_n(\theta_0)}{\partial \zeta} + \frac{1}{n} \frac{\partial^2 \hat{Q}_n(\ddot{\theta}_n)}{\partial \zeta \partial \Psi'} \sqrt{n}(\tilde{\Psi}_n - \Psi_0) \\
&= -\frac{2}{\sqrt{n}} (W_n X_n^{**}, Z_n)' \mathbb{M}_{(-WX\delta_0, X)} V_n + o_p(1) \\
&\xrightarrow{d} N\big(0, \operatorname{plim}_{n\to\infty} \frac{4}{n} (W_n X_n^{**}, Z_n)' \mathbb{M}_{(-WX\delta_0, X)} (W_n X_n^{**}, Z_n)\big),
\end{aligned} \tag{16}$$

where $\mathbb{M}_{(-WX\delta_0, X)} = H_n - \mathbb{P}_{(-WX\delta_0, X)}$, $\ddot{\theta}_n$ lies between $\tilde{\theta}_n$ and θ_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 $\tilde{\mathbb{M}}_{(-WX\tilde{\delta}, X)}$ be the matrix obtained by replacing δ_0 in $\mathbb{M}_{(-WX\delta_0, X)}$ with $\tilde{\delta}_n$ and replacing H_n with $\hat{H}_n$. Then we have the following proposition.

Proposition 3.2. *Under Assumptions 1–7 and 9, when $\zeta_0 = 0$,*

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

3.2 Local power

We consider the local power of the test statistics under the alternative hypothesis that the true parameter ζ_0 of the model with sample size n is subject to Pitman's drift (the Durbin regressors and endogenous explanatory variables are relevant).

Assumption 10. $\zeta_{0n} = \frac{1}{\sqrt{n}} \kappa$, *where κ 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 10, 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 10,

$$\begin{aligned}
\sqrt{n}(\tilde{\Psi}_n - \Psi_0) &= -\big(\frac{1}{n} \frac{\partial^2 \hat{Q}_n(\ddot{\Psi}_n, 0)}{\partial \Psi \partial \Psi'}\big)^{-1} \frac{1}{\sqrt{n}} \frac{\partial \hat{Q}_n(\Psi_0, 0)}{\partial \Psi} \\
&= \big[\frac{1}{n} (-W_n X_n \delta_0, X_n)' H_n (-W_n X_n \delta_0, X_n)\big]^{-1} (-W_n X_n \delta_0, X_n)' H_n \big[\frac{1}{n} (W_n X_n^{**}, Z_n) \kappa + \frac{1}{\sqrt{n}} V_n\big] + o_p(1),
\end{aligned} \tag{17}$$

where $\ddot{\Psi}_n$ is between Ψ_0 and $\tilde{\Psi}_n$. When κ 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 κ is proportional to $(\delta_{20}, \delta'_{10}, 0)'$. Anyhow, $\tilde{\Psi}_n$ has at most the usual order $O_p(n^{-1/2})$. 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 10, the N2SLS estimator $\hat{\theta}_n$ will still satisfy $\hat{\theta}_n = \theta_{0n} + o_p(1)$, where $\theta_{0n} = (\alpha_0, \delta'_0, \zeta'_{0n})'$. By using the reparameterization ω in Section 2.2, corresponding to the drift $\kappa/\sqrt{n}$ in Assumption 10, $\omega_{0n} = \theta_{0n}$, 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 ω_{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 10.

Proposition 3.3. *Under Assumptions 1–10,*

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

where B is a Bernoulli random variable with success probability equal to $\mathrm{P}(K^* > 0 | J_1 > 0)$, and $K^* = 2\,\mathrm{plim}_{n\to\infty}\big(U_2 + \Upsilon_{2n} J + \frac{1}{n} F'_n W_n (W_n X^{**}_n, Z_n)\kappa\big)'(\frac{1}{n}\Pi_n)^{-1}(U_1 + \Upsilon_{1n} J) + \mathrm{plim}_{n\to\infty} J'\Big[\Big(\begin{smallmatrix} \frac{1}{3}(F'_n W^3_n X_n \delta_0)'\Pi_n^{-1} U_1 + (F'_n W^2_n X_n \delta_0)'\Pi_n^{-1} U_2 \\ -2(D'_n F_n)\Pi_n^{-1} U_2 \end{smallmatrix}\Big) + \Big(\begin{smallmatrix} (F'_n W^2_n X_n \delta_0)'\Pi_n^{-1}\Upsilon_{2n} \\ -2(D'_n F_n)\Pi_n^{-1}\Upsilon_{2n} \end{smallmatrix}\Big) J + \Big(\begin{smallmatrix} \frac{1}{n}(F'_n W^2_n X_n \delta_0)'\Pi_n^{-1} F'_n W_n (W_n X^{**}_n, Z_n)\kappa \\ -\frac{2}{n}(F'_n D_n)'\Pi_n^{-1} F'_n W_n (W_n X^{**}_n, Z_n)\kappa \end{smallmatrix}\Big)\Big]$.

We use Proposition 3.3 to derive the asymptotic distribution of the distance difference test statistic under the local alternative in Assumption 10. 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 1–10, $\hat{Q}_n(\tilde{\theta}_n) - \hat{Q}_n(\hat{\theta}_n) \xrightarrow{d} [r^2 + \chi^2(k_{x*} + k_z - 1, c_1)]I(r > 0) + \chi^2(k_{x*} + k_z - 1, c_1)I(r < 0)$, 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 κ 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 10. Note that

$$\begin{aligned}\frac{1}{\sqrt{n}}\frac{\partial \hat{Q}_n(\Psi_0,0)}{\partial \zeta} &= -\frac{2}{\sqrt{n}}(W_nX_n^{**},Z_n)'\hat{H}_n(e^{\alpha_0 W_n}Y_n - X_n\delta_0)\\ &= -\frac{2}{n}(W_nX_n^{**},Z_n)'H_n(W_nX_n^{**},Z_n)\kappa - \frac{2}{\sqrt{n}}(W_nX_n^{**},Z_n)'H_nV_n + o_p(1)\end{aligned}$$

and

$$\frac{1}{n}\frac{\partial^2 \hat{Q}_n(\tilde{\Psi}_n,0)}{\partial\zeta\partial\Psi'} = \frac{2}{n}(W_nX_n^{**},Z_n)'\hat{H}_n(-W_ne^{\tilde{\alpha}_nW_n}Y_n,X_n) = \frac{2}{n}(W_nX_n^{**},Z_n)'H_n(-W_nX_n\delta_0,X_n)+o_p(1).$$

Then, by the mean value theorem and using (17),

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

where $\ddot{\Psi}_n$ is between Ψ_0 and $\tilde{\Psi}_n$. Proposition 3.5 then follows.

Proposition 3.5. *Under Assumptions 1–7 and 9–10,*

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

where c_1 is defined in Proposition 3.4.

Under the Pitman drift in Assumption 10, the gradient test statistic is asymptotically distributed as a noncentral chi-square random variable with its 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 κ 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 11, which corresponds to the rate of convergence for $(\hat{\alpha}_n,\hat{\beta}_{2n},\hat{\beta}_{3n}')'$.[18]

Assumption 11. $\zeta_{0n} = n^{-1/4}(\delta_{20},\delta_{10}',0)'$.

Under Assumption 11, by the mean value theorem and using the derivatives in Appendix A,

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

[18] We have not carried out a power analysis for the distance difference test under Assumption 11, because the asymptotic distribution of the unrestricted N2SLS estimator cannot be derived in a way similar to that in Section 2.2.

where $\ddot{\Psi}_n$ is between Ψ_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, we can derive the asymptotic distribution of the gradient test statistic.

Proposition 3.6. *Under Assumptions 1–7, 9 and 11,*

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

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 Assumption 8, 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

$$\frac{1}{n}\hat{Q}_n(\theta) + \lambda_n \|\tilde{\zeta}_n\|^{-\mu}\|\zeta\|, \tag{18}$$

where λ_n is a positive tuning parameter, $\tilde{\zeta}_n$ is an initial consistent estimator of ζ, $\|\cdot\|$ denotes the Euclidean norm (l_2-norm), and μ is some positive number such as 1 or 2 as in the literature. The AGLASSO estimator $\hat{\theta}_n$ minimizes (18). Since the irregular phenomenon appears when the whole vector $\zeta_0 = 0$, ζ is penalized in a group with an l_2-norm and there is no need to penalize other parameters for our issue under concern. We are not interested in whether or not an individual component of ζ is zero, so a penalty term with an l_1-norm is not needed.[19] 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. In general, $\tilde{\zeta}_n$ can be any consistent estimate.

Assumption 12. $\tilde{\zeta}_n = \zeta_0 + o_p(1)$.

4.1 Asymptotic properties

We study the asymptotic properties of the AGLASSO estimator in this subsection. the N2SLS estimator of β has an explicit form for a given α, so only the parameter space of α 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.

[19] Since our motivation is to avoid the irregular phenomenon that appears when the whole vector ζ_0 is zero, it is natural to use the AGLASSO. It is also possible to use only the adaptive LASSO by penalizing the parameters individually. But the adaptive LASSO makes selection based on the strength of individual variables and can result in selecting more variables than necessary (Yuan and Lin, 2006).

Assumption 4′. *The true parameter vector θ_0 is in the interior of the compact parameter space Θ for θ.*

Assumption 13 is needed for the consistency of $\hat{\theta}_n$.

Assumption 13. $\lambda_n > 0$ *and* $\lambda_n = o(1)$.

Proposition 4.1. *Under Assumptions 1–3, 4′, 6, 9, 12 and 13,* $\hat{\theta}_n = \theta_0 + o_p(1)$.

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 14 is needed. It requires the penalty term to have at least certain order when $\zeta_0 = 0$.

Assumption 14. *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 $\tilde{\zeta}_n = O_p(n^{-1/4})$. In order that $n^{1/2+\mu/4}\lambda_n \to \infty$, then if $\lambda_n = O(n^{-c})$ for some $c > 0$, it requires $c < 1/2 + \mu/4$, , i.e., λ_n may not be allowed to go to zero too fast with a rate equal to or faster than $n^{1/2+\mu/4}$.

Proposition 4.2. *Under Assumptions 1–3, 4′, 6, 9 and 12–14, 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′, 6, 9, and 12–14, if* $\zeta_0 = 0$, *then*

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

Proposition 4.3 shows that, when $\zeta_0 = 0$, the AGLASSO estimator for Ψ 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′, 6, 9, 12 and 13, if* $\zeta_0 \neq 0$, $\hat{\theta}_n = \theta_0 + O_p(n^{-1/2} + \lambda_n)$.

When $\zeta_0 \neq 0$, λ_n may affect the convergence rate of $\hat{\theta}_n$ to θ_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 λ_n convergent to zero will be needed. For that purpose, we maintain Assumption 15.

Assumption 15. $\lambda_n = o(n^{-1/2})$.

Proposition 4.5. *Under Assumptions 1–3, 4′, 6, 9, 12, 13 and 15, if* $\zeta_0 \neq 0$, *then*

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

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

For the sparsity property of $\hat{\zeta}_n$ when $\zeta_0 = 0$, Assumption 14 requires λ_n to be large enough. When $\zeta_0 \neq 0$, Assumption 15 requires λ_n to be small enough in order to make the bias resulting from the penalty term small. These assumptions pull the selection of λ_n in different directions. However, there exist a range of values for λ_n which can satisfy them. Given μ 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 λ_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

$$\hat{\theta}_\lambda = \arg\min_{\theta\in\Theta}\big[\frac{1}{n}\hat{Q}_n(\theta) + \lambda\|\tilde{\zeta}_n\|^{-\mu}\|\zeta\|\big].$$

Let $\Lambda = [0, \lambda_{\max}]$ be an interval from which the tuning parameter λ is selected, where $\lambda_{\max}$ is a finite positive number. We propose to select the tuning parameter λ that minimizes the following information criterion:

$$h_n(\lambda) = \frac{1}{n}\hat{Q}_n(\hat{\theta}_\lambda) - f(\hat{\zeta}_\lambda)\Gamma_n, \tag{19}$$

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 Γ_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 ζ 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 16. *$\Gamma_n > 0$, $\Gamma_n \to 0$ 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 16, we may take $\Gamma_n = O(n^{-1/4})$. Assumption 16 shows that the information criterion in (19) 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 13–15, e.g., $\bar{\lambda}_n = n^{-1/2-\mu/8}$. Define $\Lambda_n = \{\lambda \in \Lambda : \hat{\zeta}_\lambda = 0 \text{ if } \zeta_0 \neq 0, \text{ and } \hat{\zeta}_\lambda \neq 0 \text{ if } \zeta_0 = 0\}$, which collects non-favorable λ's.

Proposition 4.6. *Under Assumptions 1–3, 4′, 6, 9, 12 and 16,* $\mathrm{P}(\inf_{\lambda\in\Lambda_n} h_n(\lambda) > h_n(\bar{\lambda}_n)) \to 1$ *as $n \to \infty$.*

Proposition 4.6 does not mean that the tuning parameter chosen by minimizing the information criterion in (19) must be $\bar{\lambda}_n$ because $h_n(\hat{\lambda}_n) \leq h_n(\bar{\lambda}_n)$. Instead, it means that any λ that fails to identify the true model cannot be selected asymptotically as the optimal tuning parameter by the information criterion in (19), 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 α and ζ, the AGLASSO estimator of δ is

$$\hat{\delta}_n(\alpha,\zeta) = (X_n'\hat{H}_nX_n)^{-1}X_n'\hat{H}_n[e^{\alpha W_n}y_n - (W_nX_n^{**}, Z_n)\zeta]. \tag{20}$$

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 $L_{n1}(\alpha,\zeta) \equiv \frac{1}{n}\hat{Q}_n(\alpha,\hat{\delta}_n(\alpha,\zeta),\zeta) = \frac{1}{n}[\hat{\mathcal{M}}_n e^{\alpha W_n} y_n - \hat{\mathcal{M}}_n(W_n X_n^{**}, Z_n)\zeta]'[\hat{\mathcal{M}}_n e^{\alpha W_n} y_n - \hat{\mathcal{M}}_n(W_n X_n^{**}, Z_n)\zeta]$ with $\hat{\mathcal{M}}_n = \hat{\Pi}_n^{-1/2} F_n'[I_n - X_n(X_n'\hat{H}_n X_n)^{-1}X_n'\hat{H}_n]$. Note that $L_n(\alpha,\zeta)$ is an AGLASSO criterion function in the least squares framework for a given α, then we can directly apply the algorithms for computing the usual group LASSO.[20] Let $\hat{\zeta}_n(\alpha)$ be the AGLASSO estimator of ζ for a given α. 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 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:

$$e^{\alpha W_n} Y_n = X_{n1}\beta_1 + l_n\beta_2 + W_n X_{n1}\beta_3 + Z_n\beta_4 + V_n. \tag{21}$$

We conduct Monte Carlo studies with two row-normalized spatial weights matrices: one is based on the queen criterion and the other on the rook criterion. The $X_{n1} = (x_{n1,i})$ contains an exogenous variable drawn independently from $N(0,1)$. The standard deviation of v_{ni} is set to be proportional to $|x_{n1,i}|$ so that $v_{ni} = |x_{n1,i}|\epsilon_{ni}$, where $\epsilon_{ni} \sim N(0,\sigma_0^2)$. We consider 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 (u_{ni},ϵ_{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 σ_0^2 is chosen such that $R^2 = 0.2$ or 0.8, for $R^2 = \mathrm{var}(X_{n1}\beta_{10} + W_n X_{n1}\beta_{30})/[\mathrm{var}(X_{n1}\beta_{10} + W_n X_{n1}\beta_{30}) + \bar{\sigma}^2]$ with $\bar{\sigma}^2$ being the average variance $\frac{\sigma_0^2}{n}\sum_{i=1}^n x_{n1,i}^2$ of v_{ni}'s. 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 α_0 is set to either -0.2 or -1, and ς_0 is set to -0.2.[21] 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. The nominal size of a test is set to 0.05. For the investigation of powers of test statistics, the data are generated by MESS models with ζ_0 values being $(1,0.5)'$, $(1,1)'$, $(1,1.5)'$, $(1,2)'$, $(1,2.5)'$, or $(1,3)'$. The tuning parameter λ for the AGLASSO is selected by minimizing the information criterion (19) with $\Gamma_n = 0.14n^{-1/4}$.[22] We consider two sample sizes: $n = 144$ or 400. The number of Monte Carlo repetitions is $2,000$.

[20] 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).

[21] The MESS model and the SAR model have similar interpretations in some sense. The chosen values -0.2 and -1 for the spatial parameter α in (21) 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).

[22] In theory, the information criterion (19) can achieve model selection consistency as long as Γ_n satisfies the order requirement in Assumption 16. But the finite sample performance depends on the choice of Γ_n. From the proof of Proposition 4.6, when $\zeta_0 = 0$, Γ_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, $n = 144$, $R^2 = 0.8$ and $\alpha_0 = -0.2$, 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 $c = 0.14$. We then set $\Gamma_n = 0.14n^{-1/2}$ in all cases and for all sample sizes.

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 queen matrix, $R^2 = 0.2$, $\alpha_0 = -1$, and $n = 400$.[23] 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 β_1 and β_4 are close to the standard normal, those for α, β_2 and β_3 show obvious deviations from the normal distribution. In particular, the density estimate for α and β_3 have shown bimodal behaviors. 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.]

Table 1 presents the probabilities that the AGLASSO estimator selects the right model, i.e., the proportions of Monte Carlo repetitions where the AGLASSO estimates $\hat{\zeta}_n = 0$ when $\zeta_0 = 0$, or $\hat{\zeta}_n \neq 0$ when $\zeta_0 \neq 0$. All probabilities are higher than 96%. For $\zeta_0 = 0$, while the AGLASSO estimates ζ as nonzero with small probabilities when $n = 144$, it can correctly estimate ζ as zero with probabilities equal to or very close to 1 as the sample size n increases to 400. With a nonzero ζ_0, for cases where the AGLASSO does not always select the right model when $n = 144$, the correct model selection probabilities also increase as n increases to 400.

[Table 1 about here.]

To investigate relevant ratios of convergence 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 α, β_2 and β_3 are only $n^{1/4}$-consistent, but those of β_1 and β_4 are $n^{1/2}$-consistent, and the AGLASSO and N2SLS-r estimators of α, β_1 and β_2 are $\sqrt{n}$-consistent. In this case, Table 2 shows that, for the N2SLS, the ratios of α, β_2 and β_3 fluctuate around 1.29 and are significantly smaller than 1.67, and those of β_1 and β_4 are around 1.67; for the N2SLS-r, the ratios of α, β_1 and β_2 are close to 1.67; for the AGLASSO, the ratios for β_1 are around 1.67, and those for α and β_2 are slightly 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

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

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 $n = 144$.[24] Table 3 shows the results in the irregular case with $\zeta_0 = 0$. We first focus on the N2SLS. Biases and SEs for α, β_2 and β_3 are relatively larger than those for β_1 and β_4.[25] Specifically, while the biases for β_1 and β_4 are all smaller or equal to 0.036 and 0.007 respectively, and the SEs for β_1 and β_4 are all smaller or equal to 0.157 and 0.172 respectively, the biases and SEs for α, β_2 and β_3 are usually several times larger than those of β_1 and β_4. The impact of R^2, α_0 and spatial weights matrices on biases and SEs are ambiguous. The CPs are all close to 95% except those for β_2 with the queen matrix, which are around 85%. Since the N2SLS-r estimator has imposed the right restriction, it has much smaller bias and smaller SE than those of the N2SLS estimator in almost all cases. For the AGLASSO, due to a small 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. The CPs for the N2SLS-r and AGLASSO are similar and close to 95% 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 0.055 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 β_4, the SEs of the N2SLS estimators for α, β_2 and β_3 are significantly smaller, while those for β_1 and β_4 have similar magnitudes. Since the N2SLS-r estimator has imposed the wrong restriction $\zeta = 0$, it has relatively large bias in all cases. As the AGLASSO will estimate ζ as nonzero with probabilities close to one, it has almost 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 are similar to those for Table 4.

[Table 4 about here.]

[Table 5 about here.]

[24]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. 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.

[25]Recall that the N2SLS estimators of α, β_2 and β_3 only have the $n^{1/4}$-rate of convergence.

The empirical size and power properties of the distance difference test and gradient test are summarized in Table 6. The two tests have small size distortions, with the largest size distortion being 2.4 percentage points. For $R^2 = 0.2$, all powers for the two tests are 100%. For cases with $R^2 = 0.8$ and $n = 144$, we observe powers to be around 70% when $\zeta_0 = (1, 0.5)'$. The power generally increases as the sample size and β_{40} in the DGP increase. Cases with the rook matrix have larger powers than those with the queen matrix for the distance difference test, but it is the other way round for the gradient test. Cases with different α_0 values have similar power. Overall, both the distance difference test and gradient test are powerful. None of the two tests is observed to dominate the other one.

[Table 6 about here.]

6 Conclusion

In this paper, we consider estimation of the MESS model with 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. With given IVs, optimal N2SLS estimation is feasible with a HAC estimated covariance matrix of empirical moments. For the N2SLS estimation, parameters of the model 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. However, with those coefficients being zero, the N2SLS estimator becomes irregular as it has slower than the $\sqrt{n}$-rate of convergence and non-normal asymptotic distribution. 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 tests for 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 to the N2SLS estimation, we propose a simultaneous estimation and model selection procedure 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. 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, N2SLS estimators of the parameters with only the $n^{1/4}$-rate of convergence in the irregular case have large biases and SEs, but N2SLS estimators of all parameters in the regular case perform well. The AGLASSO estimators perform as well as the restricted N2SLS in the irregular case and as the unrestricted N2SLS in the regular case. The distance difference test and gradient test have small size distortions 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 tests, then a further estimation of the restricted model is needed if the null hypothesis of no Durbin and endogenous regressors is not rejected.

Acknowledgments

We are grateful to the editors Zongwu Cai, Yongmiao Hong and Cheng Hsiao of this special issue 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 at Huazhong University of Science and Technology, LEO at University of Orleans, 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, No. 71471108), Program for Innovative Research Team of Shanghai University of Finance and Economics, 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:

$$\frac{\partial^2 Q_n(\theta)}{\partial\alpha^2} = 2Y_n' e^{\alpha W_n'} W_n'^2 H_n(e^{\alpha W_n}Y_n - D_n\beta) + 2Y_n' e^{\alpha W_n'} W_n' H_n W_n e^{\alpha W_n} Y_n, \tag{22}$$

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

$$\frac{\partial^2 Q_n(\theta)}{\partial\beta\partial\beta'} = 2D_n' H_n D_n. \tag{24}$$

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

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

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

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

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

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

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 ω, any additional derivatives for this derivative are zero. Hence, by Lemma 2 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 ω.

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

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

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.

Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. New York, Oxford University Press.

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.

Elhorst, J.P., Lacombe, D.J., Piras, G., 2012. On model specification and parameter space definitions in higher order spatial econometric models. Regional Science and Urban Economics 42, 211–220.

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.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica: Journal of the Econometric Society 50, 1029–1054.

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.

Koch, W., 2013. Does the financial structure matter for economic development? Working paper.

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., 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.

Pfaffermayr, M., 2012. Spatial convergence of regions revisited: A spatial maximum likelihood panel approach. Journal of Regional Science 52, 857–873.

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.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

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.

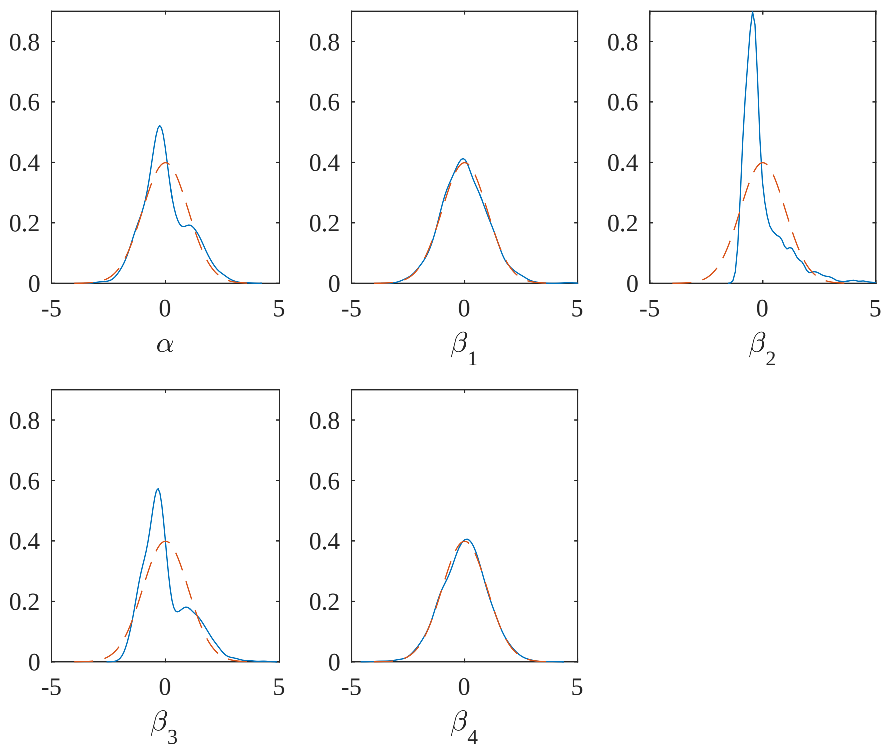

Figure 1: Kernel density estimates of the N2SLS estimators with $\zeta_0 = 0$ [Solid line: kernel density estimate; dashed line: standard normal PDF]

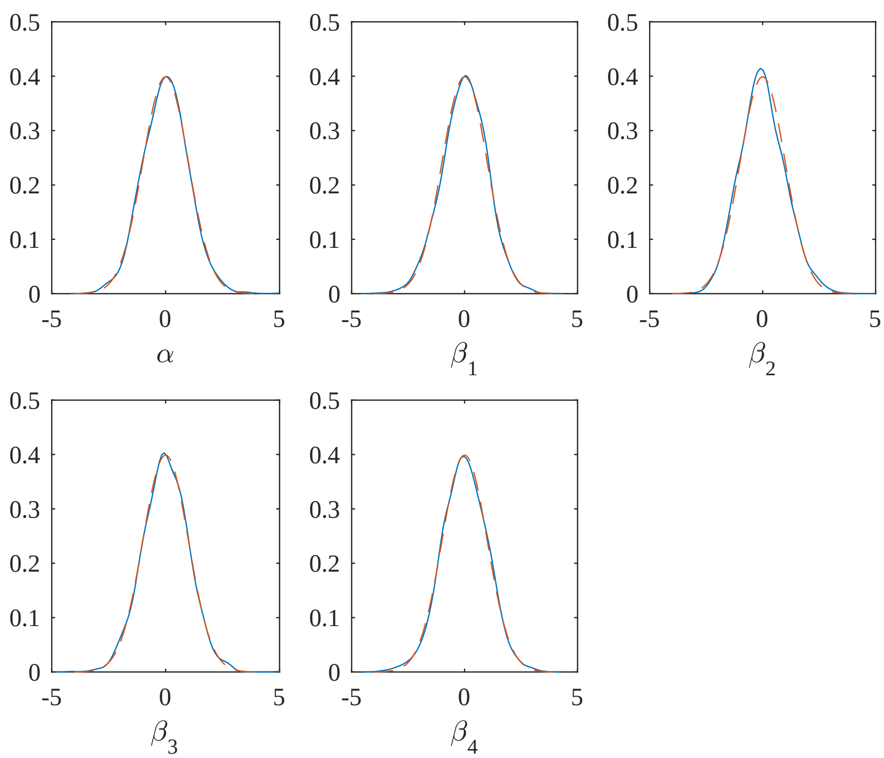

Figure 2: Kernel density estimates of the N2SLS estimators with $\zeta_0 = (1, 1)'$ [Solid line: kernel density estimate; dashed line: standard normal PDF]

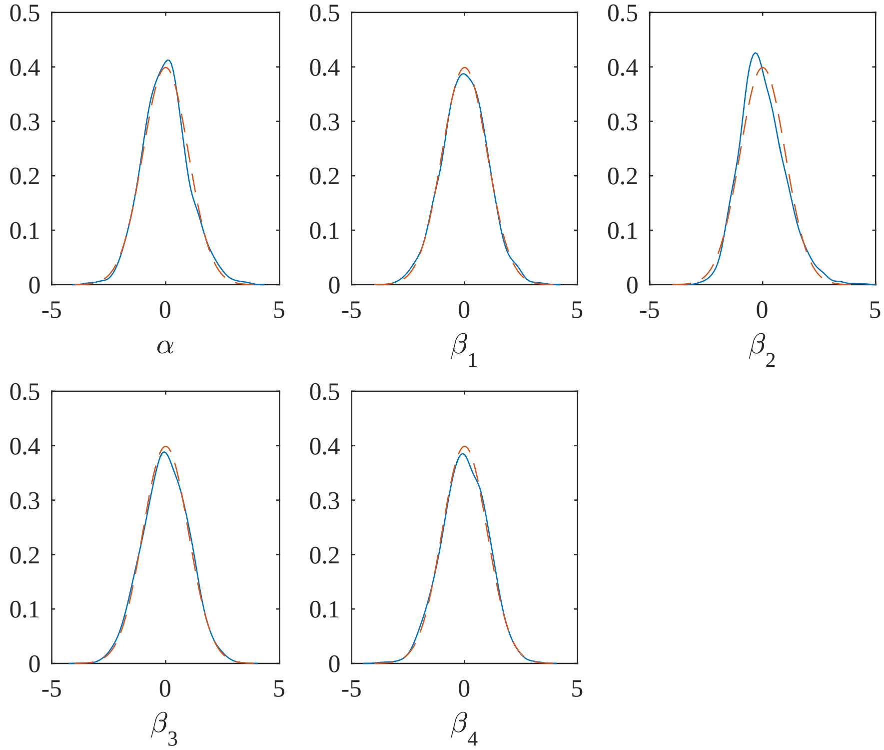

Figure 3: Kernel density estimates of the N2SLS estimators with $\zeta_0 = (0, 1)'$ [Solid line: kernel density estimate; dashed line: standard normal PDF]

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, $R^2 = 0.2$, $\alpha_0 = -0.2$	0.964	1.000	1.000	1.000	1.000	1.000
queen, $R^2 = 0.2$, $\alpha_0 = -1$	0.979	1.000	1.000	0.998	1.000	1.000
rook, $R^2 = 0.2$, $\alpha_0 = -0.2$	0.972	1.000	1.000	0.999	1.000	1.000
rook, $R^2 = 0.2$, $\alpha_0 = -1$	0.973	1.000	1.000	0.998	1.000	1.000
queen, $R^2 = 0.8$, $\alpha_0 = -0.2$	0.968	0.989	0.983	0.998	1.000	1.000
queen, $R^2 = 0.8$, $\alpha_0 = -1$	0.962	0.984	0.973	0.998	1.000	1.000
rook, $R^2 = 0.8$, $\alpha_0 = -0.2$	0.971	0.995	0.978	0.999	1.000	0.999
rook, $R^2 = 0.8$, $\alpha_0 = -1$	0.964	0.994	0.977	1.000	1.000	0.999

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 $n = 400$

	α	β_1	β_2	β_3	β_4
$\zeta_0 = 0$					
queen, $R^2 = 0.2$, $\alpha_0 = -0.2$	1.263[1.742]1.923	1.719[1.716]1.723	1.236[1.756]1.958	1.280[—]—	1.697[—]—
queen, $R^2 = 0.2$, $\alpha_0 = -1$	1.212[1.623]1.722	1.634[1.647]1.650	1.283[1.711]1.737	1.207[—]—	1.680[—]—
rook, $R^2 = 0.2$, $\alpha_0 = -0.2$	1.298[1.594]1.841	1.660[1.671]1.677	1.310[1.609]1.928	1.331[—]—	1.653[—]—
rook, $R^2 = 0.2$, $\alpha_0 = -1$	1.339[1.650]1.820	1.695[1.708]1.710	1.655[1.711]1.864	1.312[—]—	1.694[—]—
queen, $R^2 = 0.8$, $\alpha_0 = -0.2$	1.316[1.695]1.926	1.692[1.648]1.648	1.333[1.752]2.219	1.309[—]—	1.691[—]—
queen, $R^2 = 0.8$, $\alpha_0 = -1$	1.282[1.643]1.841	1.654[1.652]1.639	1.395[1.717]1.795	1.264[—]—	1.730[—]—
rook, $R^2 = 0.8$, $\alpha_0 = -0.2$	1.359[1.580]1.816	1.670[1.684]1.689	1.422[1.605]1.850	1.364[—]—	1.726[—]—
rook, $R^2 = 0.8$, $\alpha_0 = -1$	1.364[1.626]1.807	1.725[1.725]1.714	1.621[1.679]1.917	1.341[—]—	1.708[—]—
$\zeta_0 = (1, 1)$					
queen, $R^2 = 0.2$, $\alpha_0 = -0.2$	1.742[1.653]1.742	1.686[1.644]1.686	1.808[1.740]1.808	1.651[—]1.651	1.738[—]1.738
queen, $R^2 = 0.2$, $\alpha_0 = -1$	1.724[1.640]1.724	1.683[1.629]1.683	1.757[1.698]1.757	1.730[—]1.730	1.781[—]1.781
rook, $R^2 = 0.2$, $\alpha_0 = -0.2$	1.746[1.669]1.746	1.609[1.647]1.609	1.715[1.711]1.715	1.624[—]1.624	1.688[—]1.688
rook, $R^2 = 0.2$, $\alpha_0 = -1$	1.710[1.611]1.710	1.691[1.573]1.691	1.698[1.669]1.698	1.730[—]1.730	1.723[—]1.723
queen, $R^2 = 0.8$, $\alpha_0 = -0.2$	1.993[1.645]1.898	1.651[1.588]1.656	2.087[1.655]2.034	1.791[—]1.768	1.724[—]1.971
queen, $R^2 = 0.8$, $\alpha_0 = -1$	1.892[1.743]1.953	1.681[1.639]1.683	2.230[1.733]1.832	1.687[—]1.684	1.804[—]2.156
rook, $R^2 = 0.8$, $\alpha_0 = -0.2$	1.776[1.704]1.908	1.712[1.643]1.720	1.766[1.708]1.799	1.686[—]1.754	1.779[—]1.902
rook, $R^2 = 0.8$, $\alpha_0 = -1$	1.828[1.538]1.848	1.694[1.589]1.695	1.762[1.666]1.782	1.802[—]1.805	1.708[—]1.811
$\zeta_0 = (0, 1)$					
queen, $R^2 = 0.2$, $\alpha_0 = -0.2$	1.771[1.714]1.771	1.687[1.666]1.687	1.743[1.818]1.743	1.648[—]1.648	1.710[—]1.710
queen, $R^2 = 0.2$, $\alpha_0 = -1$	1.712[1.638]1.712	1.638[1.655]1.638	1.749[1.738]1.749	1.749[—]1.749	1.718[—]1.718
rook, $R^2 = 0.2$, $\alpha_0 = -0.2$	1.775[1.651]1.775	1.678[1.666]1.678	1.739[1.652]1.739	1.760[—]1.760	1.687[—]1.687
rook, $R^2 = 0.2$, $\alpha_0 = -1$	1.744[1.669]1.744	1.656[1.636]1.656	1.799[1.709]1.799	1.709[—]1.709	1.670[—]1.670
queen, $R^2 = 0.8$, $\alpha_0 = -0.2$	1.524[1.688]1.480	1.705[1.662]1.714	1.453[1.966]1.657	1.482[—]1.440	1.676[—]1.962
queen, $R^2 = 0.8$, $\alpha_0 = -1$	1.549[1.656]1.484	1.741[1.642]1.758	1.681[2.110]1.462	1.516[—]1.427	1.705[—]2.187
rook, $R^2 = 0.8$, $\alpha_0 = -0.2$	1.640[1.631]1.570	1.681[1.576]1.690	1.956[1.750]1.842	1.591[—]1.536	1.685[—]1.997
rook, $R^2 = 0.8$, $\alpha_0 = -1$	1.680[1.661]1.585	1.808[1.676]1.822	2.458[1.900]1.787	1.624[—]1.570	1.721[—]2.059

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 β_3 and β_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 β_3 and β_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 $n = 144$

	α	β_1	β_2	β_3	β_4
queen, $R^2 = 0.2$, $\alpha_0 = -0.2$					
N2SLS	-0.261[0.679]0.959	0.035[0.157]0.923	-0.033[0.736]0.821	-0.169[0.691]0.966	0.000[0.042]0.952
N2SLS-r	-0.025[0.230]0.937	-0.000[0.149]0.909	0.000[0.251]0.933	0.000[0.000]—	0.000[0.000]—
AGLASSO	-0.038[0.262]0.933	0.002[0.149]0.911	-0.006[0.283]0.926	-0.012[0.176]—	0.001[0.016]—
queen, $R^2 = 0.2$, $\alpha_0 = -1$					
N2SLS	-0.096[0.737]0.944	0.030[0.152]0.944	0.212[1.164]0.880	-0.013[0.756]0.947	0.001[0.041]0.959
N2SLS-r	-0.016[0.219]0.944	-0.005[0.142]0.922	0.007[0.244]0.928	0.000[0.000]—	0.000[0.000]—
AGLASSO	-0.026[0.236]0.942	-0.005[0.142]0.919	0.000[0.249]0.925	-0.007[0.103]—	0.000[0.014]—
rook, $R^2 = 0.2$, $\alpha_0 = -0.2$					
N2SLS	0.095[0.467]0.968	0.027[0.153]0.927	0.225[0.609]0.947	0.091[0.475]0.989	0.002[0.041]0.958
N2SLS-r	0.001[0.157]0.948	-0.003[0.148]0.901	0.013[0.181]0.952	0.000[0.000]—	0.000[0.000]—
AGLASSO	0.000[0.181]0.939	-0.002[0.148]0.899	0.017[0.216]0.944	0.000[0.109]—	0.001[0.016]—
rook, $R^2 = 0.2$, $\alpha_0 = -1$					
N2SLS	0.185[0.512]0.934	0.029[0.152]0.933	0.387[0.886]0.961	0.174[0.478]0.984	0.001[0.041]0.958
N2SLS-r	0.002[0.160]0.941	-0.006[0.145]0.912	0.017[0.188]0.944	0.000[0.000]—	0.000[0.000]—
AGLASSO	0.004[0.181]0.935	-0.005[0.145]0.914	0.023[0.218]0.939	0.001[0.100]—	0.001[0.014]—
queen, $R^2 = 0.8$, $\alpha_0 = -0.2$					
N2SLS	-0.233[0.690]0.954	0.036[0.154]0.929	0.005[0.811]0.842	-0.148[0.690]0.965	0.007[0.172]0.954
N2SLS-r	-0.019[0.222]0.936	0.002[0.145]0.908	0.005[0.254]0.926	0.000[0.000]—	0.000[0.000]—
AGLASSO	-0.025[0.259]0.929	0.003[0.145]0.908	0.010[0.324]0.919	-0.003[0.155]—	0.008[0.059]—
queen, $R^2 = 0.8$, $\alpha_0 = -1$					
N2SLS	-0.080[0.756]0.931	0.031[0.155]0.935	0.246[1.205]0.876	0.000[0.775]0.941	0.004[0.170]0.945
N2SLS-r	-0.012[0.225]0.941	-0.005[0.145]0.909	0.010[0.257]0.932	0.000[0.000]—	0.000[0.000]—
AGLASSO	-0.026[0.255]0.939	-0.004[0.144]0.908	0.003[0.276]0.926	-0.010[0.132]—	0.010[0.060]—
rook, $R^2 = 0.8$, $\alpha_0 = -0.2$					
N2SLS	0.083[0.494]0.953	0.030[0.151]0.933	0.227[0.660]0.936	0.080[0.493]0.989	0.001[0.167]0.953
N2SLS-r	0.002[0.157]0.949	-0.003[0.146]0.897	0.015[0.178]0.941	0.000[0.000]—	0.000[0.000]—
AGLASSO	-0.002[0.184]0.945	-0.001[0.146]0.900	0.015[0.209]0.941	-0.004[0.101]—	0.006[0.056]—
rook, $R^2 = 0.8$, $\alpha_0 = -1$					
N2SLS	0.182[0.514]0.927	0.026[0.154]0.926	0.379[0.863]0.964	0.168[0.481]0.990	0.005[0.166]0.945
N2SLS-r	0.007[0.157]0.947	-0.008[0.146]0.904	0.018[0.182]0.947	0.000[0.000]—	0.000[0.000]—
AGLASSO	0.013[0.174]0.942	-0.007[0.145]0.906	0.028[0.208]0.950	0.007[0.091]—	0.008[0.061]—

"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 $n = 144$

	α	β_1	β_2	β_3	β_4
queen, $R^2 = 0.2$, $\alpha_0 = -0.2$					
N2SLS	0.003[0.112]0.940	0.002[0.147]0.917	0.007[0.149]0.937	0.005[0.271]0.940	-0.002[0.040]0.949
N2SLS-r	-0.100[0.290]0.999	-0.015[0.361]0.875	-0.058[0.427]0.953	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	0.003[0.112]0.940	0.002[0.147]0.917	0.007[0.149]0.937	0.005[0.271]0.940	-0.002[0.040]0.949
queen, $R^2 = 0.2$, $\alpha_0 = -1$					
N2SLS	-0.000[0.113]0.941	-0.006[0.148]0.919	0.007[0.147]0.949	0.003[0.282]0.936	-0.002[0.042]0.940
N2SLS-r	-0.099[0.292]0.999	-0.035[0.361]0.874	-0.063[0.429]0.951	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.000[0.113]0.942	-0.006[0.148]0.919	0.007[0.147]0.949	0.003[0.282]0.936	-0.002[0.042]0.940
rook, $R^2 = 0.2$, $\alpha_0 = -0.2$					
N2SLS	-0.000[0.072]0.941	-0.005[0.142]0.928	0.004[0.114]0.931	0.005[0.176]0.946	-0.000[0.037]0.945
N2SLS-r	-0.072[0.218]1.000	-0.025[0.385]0.872	-0.037[0.400]0.945	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.000[0.072]0.941	-0.005[0.142]0.928	0.004[0.114]0.931	0.005[0.176]0.946	-0.000[0.037]0.945
rook, $R^2 = 0.2$, $\alpha_0 = -1$					
N2SLS	0.000[0.073]0.957	-0.004[0.144]0.923	0.001[0.114]0.945	-0.004[0.182]0.930	-0.002[0.039]0.944
N2SLS-r	-0.068[0.221]0.998	-0.037[0.369]0.876	-0.049[0.400]0.947	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	0.000[0.073]0.956	-0.004[0.144]0.923	0.001[0.114]0.945	-0.004[0.182]0.930	-0.002[0.039]0.944
queen, $R^2 = 0.8$, $\alpha_0 = -0.2$					
N2SLS	-0.053[0.416]0.946	0.007[0.151]0.936	0.026[0.461]0.938	-0.045[0.550]0.948	-0.012[0.165]0.941
N2SLS-r	-0.594[0.332]0.709	-0.061[0.231]0.866	-0.420[0.217]0.554	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.059[0.421]0.942	0.006[0.151]0.935	0.020[0.452]0.933	-0.052[0.550]0.939	-0.018[0.189]0.932
queen, $R^2 = 0.8$, $\alpha_0 = -1$					
N2SLS	-0.055[0.410]0.940	0.010[0.154]0.935	0.029[0.509]0.933	-0.047[0.551]0.946	-0.012[0.168]0.947
N2SLS-r	-0.602[0.347]0.680	-0.055[0.233]0.874	-0.418[0.226]0.568	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.069[0.423]0.927	0.010[0.154]0.934	0.015[0.418]0.920	-0.062[0.550]0.927	-0.021[0.201]0.934
rook, $R^2 = 0.8$, $\alpha_0 = -0.2$					
N2SLS	-0.013[0.245]0.947	0.013[0.158]0.916	0.014[0.249]0.946	-0.019[0.286]0.966	-0.010[0.158]0.946
N2SLS-r	-0.494[0.319]0.684	-0.092[0.247]0.842	-0.361[0.223]0.624	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.019[0.262]0.943	0.013[0.158]0.916	0.011[0.253]0.941	-0.024[0.297]0.960	-0.014[0.173]0.940
rook, $R^2 = 0.8$, $\alpha_0 = -1$					
N2SLS	-0.020[0.240]0.957	-0.003[0.153]0.930	0.006[0.243]0.956	-0.033[0.297]0.958	-0.014[0.156]0.936
N2SLS-r	-0.475[0.313]0.718	-0.108[0.236]0.847	-0.350[0.232]0.638	-1.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.022[0.240]0.953	-0.004[0.154]0.930	0.004[0.244]0.951	-0.034[0.295]0.953	-0.017[0.170]0.930

"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 $n = 144$

	α	β_1	β_2	β_3	β_4
queen, $R^2 = 0.2$, $\alpha_0 = -0.2$					
N2SLS	0.001[0.119]0.947	0.003[0.146]0.921	0.006[0.152]0.943	0.005[0.252]0.944	-0.002[0.043]0.945
N2SLS-r	0.024[0.308]0.999	0.002[0.343]0.884	0.076[0.491]0.971	0.000[0.000]—	-1.000[0.000]—
AGLASSO	0.001[0.119]0.947	0.003[0.146]0.922	0.006[0.152]0.943	0.005[0.252]0.944	-0.002[0.043]0.944
queen, $R^2 = 0.2$, $\alpha_0 = -1$					
N2SLS	-0.002[0.119]0.942	0.006[0.146]0.924	0.007[0.153]0.947	-0.003[0.262]0.938	-0.002[0.044]0.940
N2SLS-r	0.014[0.314]0.998	0.005[0.355]0.865	0.068[0.476]0.969	0.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.002[0.119]0.942	0.006[0.146]0.924	0.007[0.153]0.947	-0.003[0.262]0.938	-0.002[0.044]0.940
rook, $R^2 = 0.2$, $\alpha_0 = -0.2$					
N2SLS	0.002[0.085]0.944	-0.000[0.144]0.917	0.004[0.120]0.947	0.000[0.185]0.942	-0.002[0.043]0.940
N2SLS-r	0.011[0.224]0.999	-0.008[0.354]0.872	0.046[0.386]0.970	0.000[0.000]—	-1.000[0.000]—
AGLASSO	0.002[0.085]0.944	-0.000[0.144]0.917	0.004[0.120]0.947	0.000[0.185]0.942	-0.002[0.043]0.939
rook, $R^2 = 0.2$, $\alpha_0 = -1$					
N2SLS	-0.000[0.086]0.934	-0.000[0.142]0.924	0.006[0.126]0.936	-0.007[0.183]0.940	-0.002[0.043]0.949
N2SLS-r	0.010[0.223]1.000	0.002[0.353]0.871	0.038[0.391]0.971	0.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.000[0.086]0.934	-0.000[0.142]0.924	0.006[0.126]0.936	-0.007[0.183]0.940	-0.002[0.043]0.949
queen, $R^2 = 0.8$, $\alpha_0 = -0.2$					
N2SLS	-0.052[0.464]0.946	0.015[0.151]0.929	0.060[0.681]0.929	-0.011[0.509]0.947	-0.011[0.177]0.942
N2SLS-r	-0.022[0.360]0.947	-0.001[0.224]0.864	0.041[0.481]0.928	0.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.046[0.445]0.946	0.013[0.152]0.928	0.060[0.696]0.931	-0.009[0.491]0.930	-0.021[0.211]0.927
queen, $R^2 = 0.8$, $\alpha_0 = -1$					
N2SLS	-0.041[0.475]0.943	0.011[0.153]0.933	0.085[0.708]0.928	-0.008[0.517]0.955	-0.012[0.178]0.952
N2SLS-r	-0.007[0.355]0.951	-0.010[0.222]0.885	0.065[0.517]0.938	0.000[0.000]—	-1.000[0.000]—
AGLASSO	-0.049[0.451]0.946	0.010[0.155]0.931	0.058[0.596]0.928	-0.019[0.481]0.929	-0.027[0.228]0.928
rook, $R^2 = 0.8$, $\alpha_0 = -0.2$					
N2SLS	0.011[0.354]0.940	0.009[0.149]0.933	0.080[0.451]0.940	0.014[0.368]0.959	-0.018[0.178]0.953
N2SLS-r	0.013[0.253]0.951	-0.012[0.214]0.889	0.051[0.312]0.946	0.000[0.000]—	-1.000[0.000]—
AGLASSO	0.010[0.339]0.939	0.007[0.149]0.929	0.073[0.425]0.940	0.012[0.354]0.938	-0.030[0.218]0.933
rook, $R^2 = 0.8$, $\alpha_0 = -1$					
N2SLS	0.022[0.362]0.927	0.015[0.157]0.916	0.103[0.595]0.937	0.013[0.375]0.948	-0.021[0.179]0.951
N2SLS-r	0.012[0.264]0.951	0.002[0.225]0.877	0.051[0.355]0.950	0.000[0.000]—	-1.000[0.000]—
AGLASSO	0.009[0.339]0.931	0.013[0.158]0.915	0.071[0.429]0.935	0.000[0.360]0.924	-0.034[0.221]0.929

"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, α_0	n=144													
queen, 0.2, -0.2	0.072	1.000	1.000	1.000	1.000	1.000	1.000	0.052	1.000	1.000	1.000	1.000	1.000	1.000
queen, 0.2, -1	0.065	1.000	1.000	1.000	1.000	1.000	1.000	0.053	1.000	1.000	1.000	1.000	1.000	0.999
rook, 0.2, -0.2	0.073	1.000	1.000	1.000	1.000	1.000	1.000	0.047	1.000	1.000	1.000	1.000	1.000	1.000
rook, 0.2, -1	0.066	1.000	1.000	1.000	1.000	1.000	1.000	0.057	1.000	0.999	1.000	1.000	1.000	1.000
queen, 0.8, -0.2	0.056	0.662	0.953	0.994	1.000	1.000	1.000	0.045	0.714	0.975	0.996	1.000	1.000	1.000
queen, 0.8, -1	0.074	0.668	0.952	0.996	0.999	0.999	0.999	0.053	0.713	0.972	0.999	0.999	1.000	1.000
rook, 0.8, -0.2	0.050	0.789	0.958	0.994	0.999	0.999	0.999	0.044	0.637	0.953	0.993	0.999	1.000	0.998
rook, 0.8, -1	0.058	0.798	0.968	0.992	0.998	0.999	1.000	0.051	0.647	0.952	0.993	1.000	1.000	1.000
	n=400													
queen, 0.2, -0.2	0.062	1.000	1.000	1.000	1.000	1.000	1.000	0.043	1.000	1.000	1.000	0.999	1.000	0.999
queen, 0.2, -1	0.053	1.000	1.000	1.000	1.000	1.000	1.000	0.044	1.000	1.000	1.000	1.000	1.000	1.000
rook, 0.2, -0.2	0.059	1.000	1.000	1.000	1.000	1.000	1.000	0.050	1.000	1.000	1.000	1.000	0.999	1.000
rook, 0.2, -1	0.052	1.000	1.000	1.000	1.000	1.000	1.000	0.045	1.000	1.000	1.000	1.000	1.000	1.000
queen, 0.8, -0.2	0.064	0.971	1.000	1.000	1.000	1.000	1.000	0.054	0.978	1.000	1.000	1.000	1.000	1.000
queen, 0.8, -1	0.070	0.972	1.000	1.000	1.000	1.000	1.000	0.053	0.982	1.000	1.000	1.000	1.000	1.000
rook, 0.8, -0.2	0.059	0.978	1.000	1.000	1.000	1.000	1.000	0.043	0.959	1.000	1.000	1.000	1.000	1.000
rook, 0.8, -1	0.055	0.982	1.000	1.000	1.000	1.000	1.000	0.043	0.958	1.000	1.000	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 $\beta_{20} = 1$.