JACKKNIFE LAGRANGE MULTIPLIER TEST WITH MANY WEAK INSTRUMENTS

YUKITOSHI MATSUSHITA AND TAISUKE OTSU

Abstract. This paper proposes a jackknife Lagrange multiplier (JLM) test for instrumental variable regression models, which is robust to (i) many instruments, where the number of instruments may increase proportionally with the sample size, (ii) arbitrarily weak instruments, and (iii) heteroskedastic errors. To the best of our knowledge, currently there is no asymptotically size correct test in this setting. Our idea is to modify the score statistic by jackknifing and to construct its heteroskedasticity robust variance estimator. Compared to Hansen, Hausman and Newey's (2008) modification for many instruments on the LM test by Kleibergen (2002) and Moreira (2001), our JLM test is robust for heteroskedastic errors and may circumvent possible decrease in the power function. Simulation results illustrate the desirable size robustness and power properties of the proposed method.

1. Introduction

Statistical inference procedures in instrumental variable regression models can be crucially affected by the quality and number of the instrumental variables. It has been known that when instruments are only weakly correlated with the endogenous regressors, the standard asymptotic approximations to the finite sample distributions of the conventional estimators and statistics can be poor. If the number of instruments is large, efficiency of the estimators or associated tests may be improved, but it also makes the finite sample properties of usual inference procedures poor, too (see, e.g., Andrews and Stock, 2007a, for a review).

In order to overcome the weak instruments problem, several test statistics have been proposed. Kleibergen (2002) and Moreira (2001) proposed a Lagrange multiplier (LM) type statistic, while Moreira (2003) proposed a conditional likelihood ratio statistic, both of which are shown to be robust to the strength of the instruments. There have been a lot of studies on the properties of these tests and their extensions (see, e.g., Kleibergen, 2005, and Andrews, Moreira and Stock, 2006). We note that these tests were developed mainly in response to the weak instruments problem, and a priori not clear how well (or how poor) these tests would perform with many instruments.

There also have been many works investigating the effects of many instruments. Andrews and Stock (2007b) showed that Anderson-Rubin, LM, and conditional likelihood ratio statistics are robust to many weak instruments, where the instruments are arbitrarily weak and the number of instruments K satisfies $K^3/n \to 0$ for the sample size n. See also Newey and Windmeijer (2009) for the GMM theory including the LM statistic under the many weak moments asymptotics. Hansen, Hausman and Newey (2008) studied the case where K may be proportional to n and

We are grateful to Naoto Kunitomo for helpful comments. Matsushita acknowledges financial support from the JSPS KAKENHI (18K01541).

the error term is homoskedastic, and developed a many instruments robust standard error and modification for the LM test. Hausman *et al.* (2012) proposed a Wald type test based on heteroskedasticity and many instrument robust versions of the limited information maximum likelihood and Fuller (1977) estimators. However, all of them make assumptions on the rates of the concentration parameter or the number of instruments and/or homoskedasticity.

In this paper, we propose a jackknife Lagrange multiplier (JLM) test for instrumental variable regression models, which is robust to (i) many instruments, where the number of instruments may increase proportionally with the sample size, (ii) arbitrarily weak instruments, and (iii) heteroskedastic errors. To the best of our knowledge, currently there is no asymptotically size correct test in this setting. Our idea is to modify the score statistic by jackknifing and to construct its heteroskedasticity robust variance estimator. In particular, by applying the leave-one-out method introduced by Angrist, Imbens and Krueger (1999), we recenter a score type vector in the presence of many weak instruments and heteroskedasticity. Compared to Hansen, Hausman and Newey's (2008) modification for many instruments on the LM test by Kleibergen (2002) and Moreira (2001), our JLM test is robust to heteroskedastic error terms. Also, the Wald statistic by Hausman *et al.* (2012) is not fully robust to weak instruments because it relies on the consistency of the heteroskedastic limited information maximum likelihood estimator. Simulation results illustrate the desirable size robustness and power properties of the proposed method.

The paper is organized as follows. Section 2 presents our main results. After introducing our basic setup in Section 2.1, Section 2.2 proposes the JML statistic and studies its asymptotic property for a simple case, where there is no included exogenous regressor, and then Section 2.3 discusses a general case. In Section 3, we conduct a simulation study.

2. Main results

2.1. **Setup.** We first introduce our basic setup. Consider a single structural equation

$$y_{1i} = y_{2i}'\beta + z_{1i}'\gamma + u_i, \tag{2.1}$$

for $i = 1, \dots, n$, where y_{1i} is a scalar dependent variable, y_{2i} is a G-dimensional vector of endogenous regressors, z_{1i} is a K_1-dimensional vector of (included) exogenous regressors in (2.1), β and γ are G- and K_1-dimensional vectors of unknown parameters, respectively, and u_i is an error term. We assume that (2.1) is the first equation in a simultaneous system of $G+1$ linear stochastic equations relating $G+1$ endogenous variables $y_i = (y_{1i}, y_{2i}')'$, and $K = K_1 + K_2$ exogenous variables $z_i = (z_{1i}', z_{2i}')'$, where z_{2i} is a K_2-dimensional vector of instrumental variables for (2.1). The number of instruments $K_2 = K_{2n}$ may grow with the sample size n. We also assume that $(u_1, \dots, u_n)$ are mutually independent with $E(u_i|z_i) = 0$ for $i = 1, \dots, n$. The reduced form of y_i is defined as

$$y_i = \Pi' z_i + v_i = \begin{pmatrix} \pi_1' \\ \Pi_2' \end{pmatrix} z_i + \begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix}, \tag{2.2}$$

where π_1 is a K-dimensional vector and Π_2 is a $K \times G$ matrix of the reduced form coefficients, and $v_i = (v_{1i}, v_{2i}')'$ is a $(1+G)$-dimensional vector of the disturbances. $(v_1, \ldots, v_n)$ are mutually independent with $E(v_i|z_i) = 0$.

In this setup, we are interested in the following testing problem

$$H_0 : \beta = b \quad \text{against} \quad H_1 : \beta \neq b,$$

for a given b. In particular, we focus on the situation, where (i) the number of instruments may increase proportionally with the sample size (i.e., $K/n \to \alpha \in [0, 1)$ as $n \to \infty$), (ii) the instruments are arbitrarily weak (i.e., Π_2 may be zero), and (iii) the error term u_i may be heteroskedastic and non-normal, and develop a new robust test statistic for this scenario.

2.2. **Simple case: No exogenous regressor.** To present the basic idea, we begin with a simple case, where there is no included exogenous regressor, i.e., $y_{1i} = y_{2i}'\beta + u_i$. A general case will be considered in the next subsection.

We introduce some notation to define our test statistic. Let $Y_2 = (y_{21}, \ldots, y_{2n})'$, $Z = (z_1, \ldots, z_n)'$, and $V_2 = (v_{21}, \ldots, v_{2n})'$ be matrices for the endogenous regressors, instruments, and reduced form errors, respectively. Also define the observables $u_{0i} = y_{1i} - y_{2i}'b$ and $u_0 = (u_{01}, \ldots, u_{0n})'$. Based on the projection matrix $P = Z(Z'Z)^{-1}Z'$, we define an $n \times n$ matrix P^* such that $P^*_{ij} = P_{ij}$ for $i \neq j$ and $P^*_{ii} = 0$.

We note that under the null hypothesis $H_0 : \beta = b$, the score type vector $Y_2'Pu_0$ is not centered, i.e., $E(Y_2'Pu_0) = E(V_2'Pu_0) \neq 0$. This is due to the covariance between V_{2i} and u_{0i} multiplied by the diagonal element P_{ii} of P. Thus, we propose to construct our test statistic based on the jackknife version of the score type vector $Y_2'P^*u_0$, which satisfies $E(Y_2'P^*u_0) = 0$. The (conditional) variance of $Y_2'P^*u_0$ is written as

$$\begin{aligned}\Psi &= Var(Y_2'P^*u_0|Z) \\ &= \sum_{i,j,k,i\neq k,j\neq k}^{n} \sigma_k^2 \Pi_2' z_i P_{ik} P_{kj} z_j' \Pi_2 + \sum_{i\neq j}^{n} P_{ij}^2 \{E(v_{2i}v_{2i}'|Z)\sigma_j^2 + E(v_{2i}u_i|Z)E(v_{2j}'u_j|Z)\},\end{aligned}$$

where $\sigma_i^2 = E(u_i^2|Z)$. Note that this variance formula allows heteroskedastic errors. We estimate the variance Ψ by

$$\hat{\Psi} = Y_2'P^*\Sigma_0 P^* Y_2 + \sum_{i,j=1}^{n} y_{2i}y_{2j}'u_{0i}u_{0j}P_{ij}^{*2}, \tag{2.3}$$

where $\Sigma_0 = \text{diag}(u_{01}^2, \ldots, u_{0n}^2)$. By standardizing the jackknife score vector by this variance estimator, our JLM test statistic is defined as

$$JLM = (u_0'P^*Y_2)\hat{\Psi}^{-1}(Y_2'P^*u_0). \tag{2.4}$$

Compared to the standard LM statistic, $(u_0'PY_2)\left[\hat{\sigma}^2(Y_2'PY_2)^{-1}\right]^{-1}(Y_2'Pu_0)$, for some homoskedastic error variance estimator $\hat{\sigma}^2$ (Wang and Zivot, 1998), the major differences of our approach are uses of the jackknife score $Y_2'P^*u_0$ and heteroskedasticity robust variance estimator $\hat{\Psi}$.

To study the asymptotic property of JLM, we impose the following assumptions.

Assumption 1. *(i) Z is of full column rank, and there exists a constant $c \in [P_{ii}, 1)$ for all $i = 1, \ldots, n$. (ii) Conditional on Z, $\{(u_i, v_{2i}')\}_{i=1}^n$ are independent with $E(u_i|Z) = 0$ and $E(v_{2i}|Z) = 0$. (iii) There exists a positive constant C such that $\sup_i E(u_i^4|Z) \leq C$, $\sup_i E(|v_{2i}|^4|Z) \leq C$, almost surely, and $\sup_i E(|z_i'\pi_{2s}|^4) < C$ for all $s = 1, \ldots, G$, where π_{2s} is the s-th column of Π_2. (iv) $P(|corr(c'v_{2i}, u_i)| = 1|Z) < 1$ for any $c \neq 0$.*

Assumptions (i)-(iii) are standard in the literature of instrumental variable regression. Assumption (iv) is a mild condition to guarantee positive definiteness of Ψ (see Lemma 1 in Appendix). Under these assumptions, the limiting null distribution of the JLM statistic is obtained as follows.

Theorem 1. *Suppose Assumption 1 holds true, and $K \to \infty$ and $K/n \to \alpha \in [0, 1)$ as $n \to \infty$. Then under H_0,*

$$JLM \xrightarrow{d} \chi^2_G.$$

This theorem shows asymptotic pivotalness of the JLM statistic under the arbitrarily weak instruments, many instruments in the sense of $K/n \to \alpha \in [0, 1)$, and heteroskedasticity. An asymptotically valid confidence set for β can be constructed by inverting JLM.

We note that to the best of our knowledge, the JLM test is the only asymptotically size correct test in the setup of this theorem. The LM test by Kleibergen (2002) and Moreira (2001) is not robust to many instruments in the sense of $\alpha > 0$. Hansen, Hausman and Newey's (2008) modified version is robust to the case of $\alpha > 0$, but not robust to heteroskedastic errors. The Wald test by Hausman *et al.* (2012) is also robust to the case of $\alpha > 0$, but not fully robust to weak instruments.

Furthermore, we note that the LM statistic by Kleibergen (2002) and Moreira (2001) and its modification by Hansen, Hausman and Newey (2008) may have spurious decline of power in some regions for the alternative hypotheses. This spurious decline of power is caused by the fact that those LM statistics are equal to zero at the maximum as well as the minimum of the concentrated log-likelihood since both the Kleibergen's LM statistics and its modification are quadratic forms of the score of the concentrated likelihood (see, p. 1788 of Kleibergen, 2002). On the other hand, the jackknife score for our JLM statistic is different from the score of the concentrated likelihood, and our simulation study indicates that the JLM statistic does not exhibit such decline of power.

2.3. **Models with exogenous regressors.** In this subsection, we extend our analysis to models with exogenous regressors in (2.1). Under the null hypothesis $H_0 : \beta = b$, the slope parameters γ for the exogenous regressors can be estimated by

$$\hat{\gamma}(b) = (Z_1'Z_1)^{-1}Z_1'(y_1 - Y_2 b),$$

where $Z_1 = (z_{11}, \ldots, z_{1n})'$. We can construct the JLM test statistic in the same way as in the previous section. Define $\hat{u}_{0i} = y_{1i} - y_{2i}'b - z_{1i}'\hat{\gamma}(b)$ and $\hat{u}_0 = (\hat{u}_{01}, \ldots, \hat{u}_{0n})$. Based on the projection matrix $P_1 = Z_1(Z_1'Z_1)^{-1}Z_1'$, we define $n \times n$ matrices $P_2 = (I - P_1)Z_2(Z_2'(I - P_1)Z_2)^{-1}Z_2'(I - P_1)$ and $P^{\#}$ such that $P_{ij}^{\#} = P_{2,ij}$ for $i \neq j$ and $P_{ii}^{\#} = 0$.

We can show that under H_0,

$$(\hat{u}_0' P^{\#} Y_2)(\Psi^{\dagger})^{-1}(Y_2' P^{\#} \hat{u}_0) = (u_0' P^{\dagger} Y_2)(\Psi^{\dagger})^{-1}(Y_2' P^{\dagger} u_0) + o_p(1), \tag{2.5}$$

where $u_{0i} = y_{1i} - y_{2i}' b - z_{1i}' \gamma$ and $u_0 = (u_{01}, \ldots, u_{0n})'$, $P^{\dagger}$ is a $n \times n$ matrix such that $P_{ij}^{\dagger} = [P_2 + \text{diag}(P_2) P_1]_{i,j}$ for $i \neq j$ and $P_{ii}^{\dagger} = 0$, and

$$\Psi^{\dagger} = \sum_{i,j,k,i\neq k,j\neq k}^{n} \sigma_k^2 \Pi_2' z_i P_{ik}^{\dagger} P_{kj}^{\dagger} z_j' \Pi_2 + \sum_{i\neq j}^{n} P_{ij}^{\dagger 2} \{E(v_{2i} v_{2i}' | Z) \sigma_j^2 + E(v_{2i} u_i | Z) E(v_{2j}' u_j | Z)\}.$$

Thus, the score type vector $Y_2' P^{\#} \hat{u}_0$ can be a proxy for the mean zero vector $Y_2' P^{\dagger} u$, and we can construct the JML statistic for this general case by the quadratic form:

$$JLM = (\hat{u}_0' P^{\#} Y_2) \tilde{\Psi}^{-1} (Y_2' P^{\#} \hat{u}_0), \tag{2.6}$$

where

$$\tilde{\Psi} = Y_2' P^{\dagger} \hat{\Sigma}_0 P^{\dagger} Y_2 + \sum_{i,j=1}^{n} y_{2i} y_{2j}' \hat{u}_{0i} \hat{u}_{0j} (P_{ij}^{\dagger})^2, \tag{2.7}$$

and $\hat{\Sigma}_0 = \text{diag}(\hat{u}_{01}^2, \ldots, \hat{u}_{0n}^2)$.

The asymptotic property of this JLM statistic is obtained as follows, and similar comments to Theorem 1 apply.

Theorem 2. *Suppose Assumption 1 holds true, and $K \to \infty$ and $K/n \to \alpha \in [0,1)$ as $n \to \infty$. Then under H_0,*

$$JLM \xrightarrow{d} \chi_G^2.$$

3. Simulation

In this section, we conduct a simulation study to evaluate the finite sample properties of the proposed JML test. We consider the data generating process:

$$\begin{aligned} y_{1i} &= y_{2i}\beta_0 + z_{1i}\gamma_0 + u_i, \\ y_{2i} &= z_i' \pi_2 + v_{2i}, \end{aligned} \tag{3.1}$$

for $i = 1, \ldots, n$, where $\pi_2 = (d, \ldots, d)'$, $z_i = (z_{1i}, z_{2i}')'$, $z_{1i} = 1$, and $z_{2i} = (z_{21i}, z_{21i}^2, z_{21i}^3, z_{22i}')'$ with $z_{21i} \sim N(0,1)$ and $z_{22i} \sim N(0, I_{K-4})$. The error terms are generated by $(u_i, v_{2i}) = ((1 + \phi z_{21i}^2)\epsilon_{1i}, \rho u_i + \sqrt{1-\rho^2}\epsilon_{2i})$, where ϵ_{1i} and ϵ_{2i} are independent and drawn from $N(0,1)$.[1] We set $n = 100$ for the sample size in all cases, and set $\beta_0 = \gamma_0 = 1$, $\rho \in \{0.2, 0.6\}$, and $\phi \in \{0, 0.2\}$ for the cases of homoskedastic and heteroskedastic errors, respectively. For each Monte Carlo replication, we set the value of d to fix the value of the concentration parameter (given the realized values of $\{z_i\}$)

$$\delta^2 = \frac{\pi_2' \left[\sum_{i=1}^n z_{2i} z_{2i}' - \sum_{i=1}^n z_{2i} z_{1i}' \left(\sum_{i=1}^n z_{1i} z_{1i}'\right)^{-1} \sum_{i=1}^n z_{1i} z_{2i}'\right] \pi_2}{Var(v_{2i})}.$$

[1] In our preliminary simulation, we also consider the t_5 and χ_3 distributions as examples of fat-tailed and skewed errors, but the results are similar to the normal case.

We investigate the size properties of eight tests for $H_0 : \beta = \beta_0$: (i) the standard t-test with the two-stage least squares estimator (t_{TS}), (ii) the standard t-test with the limited information maximum likelihood estimator (t_{LI}), (iii) the t-test with the heteroskedasticity robust limited information maximum likelihood estimator by Hausman *et al.* (2012) (t_{HLI}), (iv) the Anderson-Rubin test (AR), (v) the conditional likelihood ratio test by Moreira (2003) (CLR), (vi) the Lagrange multiplier test by Kleibergen (2002) (KLM), (vii) the modified Lagrange multiplier test by Hansen, Hausman and Newey (2008) (mKLM), and (viii) the proposed JLM test (JLM). The number of Monte Carlo repetitions in each experiment is 10,000.

Tables 1 and 2 report the null rejection frequencies of the tests at the nominal 5% significance level for the cases of homoskedastic and heteroskedastic errors, respectively. Our findings are summarized as follows.

- **i):** The size distortions of both t_{TS} and t_{LI} are large except when δ^2 is large, K is small, and the errors are homoskedastic. The distortions tend to be quite large when δ^2 is small, and K and ρ are large (see the case of $\delta^2 = 2$, $K = 30$, and $\rho = 0.6$) even in the case of homoskedastic errors.
- **ii):** The size distortions of t_{HLI} are smaller for both the cases of homoskedastic and heteroskedastic errors compared to t_{TS} and t_{LI}. However, the distortions tend to be large when δ^2 is small (see the case of $\delta^2 = 2$). More precisely, t_{HLI} under-rejects when ρ is small and over-rejects when ρ is large.
- **iii):** AR, CLR, and KLM work well even when δ^2 is small in the case of homoskedastic errors. However, they tend to over-reject when K is large. The size distortions are severe in the case of heteroskedastic errors. These findings are consistent with lack of robustness of these tests against heteroskedastic errors and relatively large K as shown in Andrews and Stock (2007b).
- **iv):** mKLM works well for all the cases of homoskedastic errors. However, it tends to over-reject in the case of heteroskedastic errors. This result is also sensible because mKLM is derived under homoskedastic errors.
- **v):** Compared to the other tests we consider, the rejection frequencies of JLM are overall close to the nominal level for all cases. The JLM test is robust to many instruments, weak instruments and heteroskedastic errors, which agrees with our theoretical results in Section 2.

We also investigate the power properties of the tests for $H_0 : \beta = \beta_0$ under the alternative hypotheses $H_1 : \beta = \beta_0 + \Delta$. Figures 1 and 2 display the calibrated power curves of t_{HLI}, AR, CLR, KLM, mKLM and JLM at 5% significance level (i.e., the rejection frequencies of these tests, where the critical values are given by the Monte Carlo 95th percentiles of these test statistics under H_0). Among various cases tried in preliminary simulations, we present the cases of $K = 30$, $\delta^2 = 30$, and $\rho = 0.2$ as typical examples. From these figures, we can see that (i) t_{HLI}, CLR and AR are more powerful than JLM, and (ii) KLM and mKLM have similar local powers as JLM, but both tests have spurious decline of power at some points under the alternative hypotheses.

ρ	δ^2	K	t_{TS}	t_{LI}	t_{HLI}	AR	CLR	KLM	mKLM	JLM
0.2	30	5	0.044	0.047	0.055	0.057	0.055	0.054	0.051	0.044
	30	10	0.059	0.064	0.038	0.064	0.059	0.056	0.051	0.047
	30	30	0.122	0.148	0.028	0.096	0.093	0.080	0.048	0.051
	10	5	0.033	0.033	0.029	0.057	0.061	0.056	0.050	0.047
	10	10	0.057	0.054	0.017	0.062	0.061	0.055	0.045	0.048
	10	30	0.157	0.161	0.015	0.099	0.107	0.090	0.043	0.048
	2	5	0.017	0.013	0.009	0.057	0.058	0.053	0.038	0.048
	2	10	0.061	0.046	0.009	0.065	0.069	0.061	0.042	0.048
	2	30	0.182	0.156	0.011	0.094	0.111	0.097	0.042	0.048
0.6	30	5	0.089	0.054	0.061	0.062	0.054	0.053	0.052	0.045
	30	10	0.181	0.063	0.048	0.066	0.057	0.056	0.051	0.045
	30	30	0.694	0.113	0.059	0.100	0.082	0.072	0.046	0.050
	10	5	0.133	0.074	0.071	0.058	0.056	0.053	0.049	0.045
	10	10	0.312	0.103	0.069	0.065	0.061	0.058	0.048	0.046
	10	30	0.877	0.214	0.081	0.101	0.101	0.086	0.044	0.047
	2	5	0.207	0.112	0.079	0.059	0.061	0.053	0.040	0.039
	2	10	0.497	0.192	0.089	0.067	0.067	0.058	0.040	0.047
	2	30	0.941	0.337	0.110	0.100	0.118	0.098	0.040	0.049

Table 1. Empirical rejection frequencies at 5% significant level: Homoskedastic errors

ρ	δ^2	K	t_{TS}	t_{LI}	t_{HLI}	AR	CLR	KLM	mKLM	JLM
0.2	30	5	0.210	0.225	0.074	0.300	0.256	0.239	0.229	0.040
	30	10	0.198	0.219	0.062	0.245	0.229	0.202	0.187	0.042
	30	30	0.230	0.235	0.037	0.198	0.182	0.146	0.092	0.050
	10	5	0.162	0.153	0.048	0.296	0.273	0.217	0.192	0.042
	10	10	0.172	0.159	0.034	0.243	0.236	0.171	0.136	0.044
	10	30	0.281	0.228	0.024	0.198	0.209	0.146	0.071	0.048
	2	5	0.113	0.083	0.024	0.295	0.288	0.192	0.134	0.044
	2	10	0.177	0.119	0.020	0.245	0.244	0.148	0.096	0.048
	2	30	0.317	0.213	0.018	0.197	0.209	0.134	0.056	0.047
0.6	30	5	0.255	0.177	0.088	0.297	0.259	0.246	0.240	0.042
	30	10	0.365	0.170	0.074	0.237	0.214	0.197	0.185	0.044
	30	30	0.874	0.188	0.069	0.208	0.179	0.151	0.113	0.044
	10	5	0.377	0.222	0.112	0.300	0.272	0.228	0.211	0.032
	10	10	0.574	0.233	0.100	0.239	0.232	0.182	0.157	0.039
	10	30	0.971	0.310	0.108	0.198	0.202	0.144	0.078	0.042
	2	5	0.530	0.330	0.140	0.298	0.284	0.195	0.148	0.026
	2	10	0.761	0.372	0.140	0.241	0.239	0.150	0.102	0.037
	2	30	0.990	0.457	0.145	0.198	0.209	0.133	0.055	0.044

TABLE 2. Empirical rejection frequencies at 5% significant level: Heteroskedastic errors

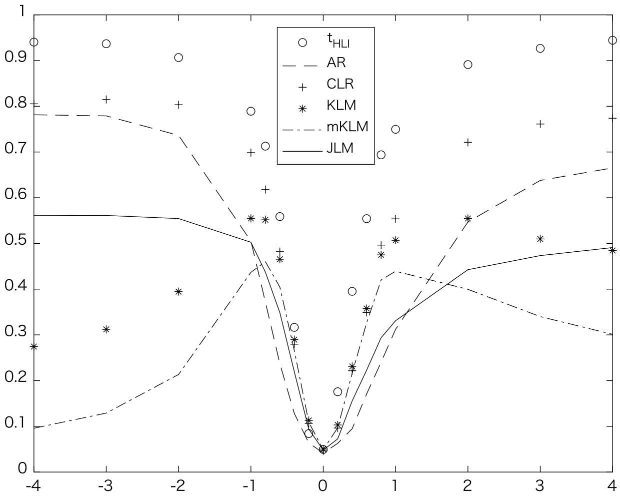

FIGURE 3.1. Calibrated power curves: $n = 100$, $K = 30$, $\rho = 0.2$, $\delta^2 = 30$, Homoskedastic errors

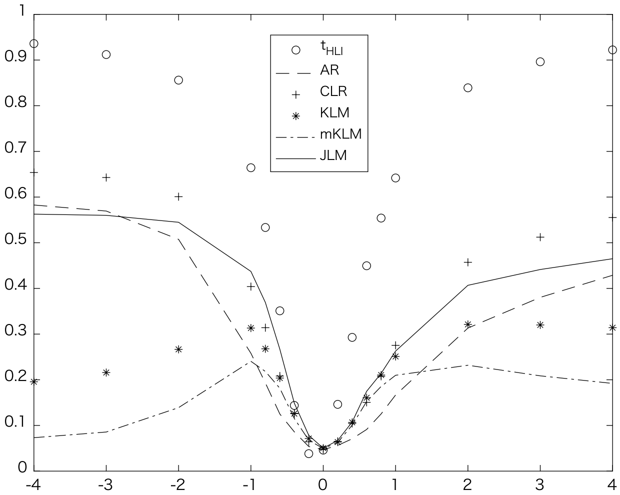

FIGURE 3.2. Calibrated power curves: $n = 100$, $K = 30$, $\rho = 0.2$, $\delta^2 = 30$, Heteroskedastic errors

Appendix A. Proof

Notation: Hereafter C mean a generic positive constant. Lemma 1 below guarantees that Ψ is positive definite almost surely. Thus, by the spectral decomposition, there exists an orthogonal matrix $Q = (q_1, \ldots, q_G)$ such that $QQ' = I$ and

$$Q'\Psi Q = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_G). \tag{A.1}$$

Also define

$$\begin{aligned} A_{ijk} &= \sigma_k^2 \Pi_2' z_i P_{ik} P_{kj} z_j' \Pi_2, \qquad B_{ij} = P_{ij}^2 \{E(v_{2i} v_{2i}'|Z)\sigma_j^2 + E(v_{2i} u_i|Z) E(v_{2j}' u_j|Z)\}, \\ \hat{A}_{ijk} &= u_k^2 y_{2i} P_{ik} P_{kj} y_{2j}', \qquad \hat{B}_{ij} = P_{ij}^2 (y_{2i} y_{2i}' u_j^2 + y_{2i} y_{2j}' u_i u_j), \end{aligned}$$

so that $\Psi = \sum_{i,j,k,i\neq k,j\neq k}^n A_{ijk} + \sum_{i\neq j}^n B_{ij}$ and $\hat{\Psi} = \sum_{i\neq j\neq k} \hat{A}_{ijk} + \sum_{i\neq j}^n \hat{B}_{ij}$ under H_0. Based on $Q = (q_1, \ldots, q_G)$, denote $\mu_g^2 = \sum_{i=1}^n q_g' \Pi_2' z_i z_i' \Pi_2 q_g$ for $g = 1, \ldots, G$.

A.1. **Proof of Theorem 1.**

By Lemma 1, Ψ is positive definite, and then we have

$$JLM = u' P^* Y_2 Q \Lambda^{-1} Q' Y_2' P^* u + o_p(1) \xrightarrow{d} \chi_G^2,$$

where the equality follows from Lemma 2 and the convergence follows from Lemma 3.

A.2. **Lemmas for Theorem 1.**

Lemma 1. *Under Assumption 1, Ψ is positive definite almost surely.*

Proof: Pick any nonzero G-dimensional vector c. Then $c'\Psi c = A + B$, where $A = \sum_{i,j,k,i\neq k,j\neq k}^n c' A_{ijk} c$ and $B = \sum_{i\neq j}^n c' B_{ij} c$. For A, note that

$$A = c' \sum_{k=1}^n \sigma_k^2 \left(\sum_{i\neq k}^n P_{ki} z_i' \Pi_2 \right)' \left(\sum_{j\neq k}^n P_{kj} z_j' \Pi_2 \right) c \geq 0.$$

For B, we have

$$\begin{aligned} B &= \sum_{i<j}^n P_{ij}^2 \{E[(c'v_{2i})^2|Z]\sigma_j^2 + E[(c'v_{2j})^2|Z]\sigma_i^2 + 2E(c'v_{2i}u_i|Z)E(c'v_{2j}u_j|Z)\} \\ &\geq \frac{1}{2} \sum_{i\neq j}^n P_{ij}^2 \{E[(c'v_{2i})^2|Z]\sigma_j^2 + E[(c'v_{2j})^2|Z]\sigma_i^2 - |2E(c'v_{2i}u_i|Z)E(c'v_{2j}u_j|Z)|\}. \end{aligned} \tag{A.2}$$

Also, Cauchy-Schwarz inequality combined with Assumption 1 (iv) implies

$$|E(c'v_{2i}u_i|Z)E(c'v_{2j}u_j|Z)| < \sqrt{E[(c'v_{2i})^2|Z]E[(c'v_{2j})^2|Z]\sigma_i^2\sigma_j^2},$$

almost surely. Thus, by $\frac{1}{2}(a^2 + b^2) \geq ab$, we have

$$E[(c'v_{2i})^2|Z]\sigma_j^2 + E[(c'v_{2j})^2|Z]\sigma_i^2 - |2E(c'v_{2i}u_i|Z)E(c'v_{2j}u_j|Z)| > 0. \tag{A.3}$$

Since $\sum_{i,j=1}^n P_{ij}^2 = \sum_{i=1}^n P_{ii} = K$, we have

$$\sum_{i\neq j} P_{ij}^2 = K - \sum_{i=1}^n P_{ii}^2 \geq K \left(1 - \max_{1\leq i\leq n} P_{ii} \right) > 0, \tag{A.4}$$

almost surely, where the last inequality follows from Assumption 1 (i).

Combining (A.2)-(A.4), we obtain $B > 0$ almost surely, and the conclusion follows.

Lemma 2. *Under Assumption 1,*

$$JLM = u' P^* Y_2 Q \Lambda^{-1} Q' Y_2' P^* u + o_p(1).$$

Proof: By Lemma 1 and (A.1), the LM statistic can be written as

$$\begin{aligned} JLM &= u_0' P^* Y_2 \{\Psi + (\hat{\Psi} - \Psi)\}^{-1} Y_2' P^* u_0 \\ &= u_0' P^* Y_2 \{QQ'\Psi QQ' + QQ'(\hat{\Psi} - \Psi)QQ'\}^{-1} Y_2' P^* u_0 \\ &= u_0' P^* Y_2 Q \Lambda^{-1/2} \{I + \Lambda^{-1/2} Q'(\hat{\Psi} - \Psi) Q \Lambda^{-1/2}\}^{-1} \Lambda^{-1/2} Q' Y_2' P^* u_0. \end{aligned}$$

Pick any $g, h = 1, \ldots, G$. Note that

$$\frac{1}{\sqrt{\lambda_g \lambda_h}} |q_g'(\hat{\Psi} - \Psi) q_h| \leq \sqrt{\frac{1}{\lambda_g} q_g'(\hat{\Psi} - \Psi) q_g} \sqrt{\frac{1}{\lambda_h} q_h'(\hat{\Psi} - \Psi) q_h} \xrightarrow{p} 0,$$

where the convergence follows from Lemma 4. Thus, we obtain $\Lambda^{-1/2} Q'(\hat{\Psi} - \Psi) Q \Lambda^{-1/2} \xrightarrow{p} 0$, and Slutsky's lemma yield the conclusion.

Lemma 3. *Under Assumption 1,*

$$u' P^* Y_2 Q \Lambda^{-1} Q' Y_2' P^* u \xrightarrow{d} \chi^2_G.$$

Proof: Without loss of generality, we assume that $\frac{K}{\mu_1^2}, \ldots, \frac{K}{\mu_g^2}$ are bounded and $\frac{K}{\mu_{g+1}^2}, \ldots, \frac{K}{\mu_G^2} \to \infty$ as $n \to \infty$. Pick any nonzero G_2-dimensional vector α, and define $S = \text{diag}(\mu_1, \ldots, \mu_g, \sqrt{K}, \ldots, \sqrt{K})$. Observe that

$$\begin{aligned} & (\alpha'\alpha)^{-1/2} \alpha' \Lambda^{-1/2} Q' Y_2' P^* u \\ = \quad & (\alpha'\alpha)^{-1/2} \left\{ \alpha' \Lambda^{-1/2} S S^{-1} Q' \sum_{i=1}^{n} \Pi_2' z_i (1 - P_{ii}) u_i + \alpha' \Lambda^{-1/2} \sqrt{K} \frac{1}{\sqrt{K}} Q' \sum_{i \neq j}^{n} v_{2i}' P_{ij} u_j \right\}. \end{aligned}$$

Here we apply Chao *et al.* (2012, Lemma A.2) by setting "U_i, ϵ_i, W_{in}, c_{1n}, and c_{2n}" in their notation as v_{2i}, u_i, $S^{-1} Q' \Pi_2' z_i (1 - P_{ii}) u_i$, $S \Lambda^{-1/2} \alpha$, and $\sqrt{K} \Lambda^{-1/2} \alpha$, respectively. It is straightforward to verify that the conditions of Chao *et al.* (2012, Lemma A.2) are satisfied. Thus, by the Cramér-Wold device, we have

$$\Lambda^{-1/2} Q' Y_2' P^* u \xrightarrow{d} N(0, I_G),$$

which implies the conclusion.

Lemma 4. *Under Assumption 1, it holds that*

$$\frac{1}{\lambda_g} q_g'(\hat{\Psi} - \Psi) q_g \xrightarrow{p} 0.$$

for each $g = 1, \ldots, G$.

Proof: Pick any $g = 1, \ldots, G$. Decompose

$$\begin{aligned} \frac{1}{\lambda_g} q_g'(\hat{\Psi} - \Psi) q_g &= \frac{1}{\lambda_g} \sum_{i \neq j \neq k} q_g'(\hat{A}_{ijk} - A_{ijk}) q_g + \frac{1}{\lambda_g} \sum_{i \neq j} q_g'(\hat{B}_{ij} - A_{iij} - B_{ij}) q_g \\ &\equiv M_1 + M_2. \end{aligned}$$

It is enough to show that $M_1, M_2 \xrightarrow{p} 0$. Based on (A.1), let $\lambda_g = A_g + B_g$, where $A_g = \sum_{i,j,k,i \neq k, j \neq k}^n q_g' A_{ijk} q_g$ and $B_g = \sum_{i \neq j}^n q_g' B_{ij} q_g$. Note that $A_g \geq 0$ and $B_g > 0$ from Lemma 1. We split into two cases: (I) B_g/A_g is bounded, and (II) $B_g/A_g \to \infty$.

Case (I). First, consider the case where B_g/A_g is bounded. It follows $M_1 \leq \frac{1}{A_g} \sum_{i \neq j \neq k} q_g'(\hat{A}_{ijk} - A_{ijk}) q_g \xrightarrow{p} 0$ from Chao *et al.* (2012, Lemma A4) by setting "W_i, Y_j, and η_k" in their notation as $\frac{1}{\sqrt{A_g}} q_g' y_{2i}$, $\frac{1}{\sqrt{A_g}} q_g' y_{2j}$, and u_k^2, respectively. Let $\mu_g^2 = \sum_{i=1}^n (q_g' \Pi_2' z_i)^2$. Note that $\mu_g^2 = O(A_g)$ because

$$\begin{aligned} \frac{\mu_g^2}{A_g} &\leq \frac{\mu_g^2}{C \sum_{i,j,k,i \neq k, j \neq k}^n q_g' \Pi_2' z_i P_{ik} P_{kj} z_j' \Pi_2 q_g} \\ &= \frac{\mu_g^2}{C \left\{ \sum_{i=1}^n (q_g' \Pi_2' z_i)^2 - 2 \sum_{i,k}^n P_{ik} P_{ki} (q_g' \Pi_2' z_i)^2 + \sum_{i=1}^n P_{ii}^2 (q_g' \Pi_2' z_i)^2 \right\}} \\ &= \frac{\mu_g^2}{C \left\{ \sum_{i=1}^n (1 - 2P_{ii} + P_{ii}^2)(g_q' \Pi_2' z_i)^2 \right\}} \leq C', \end{aligned}$$

where the first equality follows from $\sum_{i,j,k=1}^n z_i P_{ik} P_{kj} z_j' = \sum_{i=1} z_i z_i'$, the second equality follows from $\sum_{k=1}^n P_{ik} P_{ki} = P_{ii}$, and the second inequality follows from Assumption 1 (i). This allows us to verify the conditions in Chao *et al.* (2012, Lemma A4).

For M_2, we apply Chao *et al.* (2012, Lemma A3) by setting "W_i and Y_i" in their notation as $\frac{1}{A_g}(q_g' y_{2i})^2$ and u_i^2, respectively. Note that

$$E(W_i|Z) = \frac{1}{A_g}(q_g' \Pi_2' z_i)^2 + \frac{1}{A_g} q_g' E(v_{2i} v_{2i}'|Z) q_g,$$

so that

$$\max_{1 \leq i \leq n} |E(W_i|Z)| \leq C \left[\frac{1}{A_g} \max_{1 \leq i \leq n} \{(q_g' \Pi_2' z_i)^2\} + \frac{1}{A_g} \right],$$

almost surely. Moreover, for $\tilde{v}_{ig} = g_q' v_{2i}$, it holds

$$\begin{aligned} \max_{1 \leq i \leq n} Var(W_i|Z) &= \max_{1 \leq i \leq n} Var \left(\left(\frac{1}{\sqrt{A_g}} q_g' (\Pi_2' z_i + v_{2i}) \right)^2 \middle| Z \right) \\ &= \max_{1 \leq i \leq n} \frac{1}{A_g^2} \left\{ 4(q_g' \Pi_2' z_i)^2 E(\tilde{v}_{ig}^2|Z) + 4(q_g' \Pi_2' z_i) E(\tilde{v}_{ig}^3|Z) + Var(\tilde{v}_{ig}^2|Z) \right\} \\ &\leq \frac{C}{A_g^2} \max_{1 \leq i \leq n} \{(q_g' \Pi_2' z_i)^2 + (q_g' \Pi_2' z_i) + 1\}, \end{aligned}$$

almost surely. Thus, by applying Chao *et al.* (2012, Lemma A3) (with $W_i = \frac{1}{A_g}(q_g' y_{2i})^2$ and $Y_i = u_i^2$), we obtain

$$
\begin{aligned}
&\left\| \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_g - \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2\{q_g'\Pi_2' z_i z_i'\Pi_2 q_g \sigma_j^2 + q_g' E(v_{2i}v_{2i}'|z_i)\sigma_j^2 q_g\} \right\|_{L_2,Z}^2 \\
\leq\ & CK\left\{ \max_{1\leq i\leq n} Var(W_i|Z) \max_{1\leq i\leq n} Var(Y_i|Z) + \max_{1\leq i\leq n} Var(W_i|Z)\left(\max_{1\leq i\leq n} E(Y_i|Z)\right)^2 \right. \\
&\left. \quad + \left(\max_{1\leq i\leq n} E(W_i|Z)\right)^2 \max_{1\leq i\leq n} Var(Y_i|Z) \right\} \\
\leq\ & \frac{CK}{A_g^2} \max_{1\leq i\leq n}\{(q_g'\Pi_2' z_i)^4 + (q_g'\Pi_2' z_i)^2 + (q_g'\Pi_2' z_i) + 1\},
\end{aligned}
$$

almost surely. Taking expectation with respect to the distribution of Z and using Billingsley (1986, Theorem 16.1), we have

$$
\begin{aligned}
&E\left[\left\{ \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_g - \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2\{q_g'\Pi_2' z_i z_i'\Pi_2 q_g \sigma_j^2 + q_g' E(v_{2i}v_{2i}'|z_i)\sigma_j^2 q_g\} \right\}^2\right] \\
&= E_Z\left\| \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_g - \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2\{q_g'\Pi_2' z_i z_i'\Pi_2 q_g \sigma_j^2 + q_g' E(v_{2i}v_{2i}'|z_i)\sigma_j^2 q_g\} \right\|_{L_2,Z}^2 \\
&\leq \frac{CK}{A_g^2} E_Z\left[\max_{1\leq i\leq n}\{(q_g'\Pi_2' z_i)^4 + (q_g'\Pi_2' z_i)^2 + (q_g'\Pi_2' z_i) + 1\}\right] \\
&= O(K/A_g^2) = O(B_g/A_g^2) = o(1),
\end{aligned}
$$

where the third equality follows from the proof of Lemma 1. Thus, Markov inequality yields

$$
\frac{1}{A_g}\sum_{i\neq j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_g = \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2\{q_g'\Pi_2' z_i z_i'\Pi_2 q_g \sigma_j^2 + q_g' E(v_{2i}v_{2i}'|Z)\sigma_j^2 q_g\} + o_p(1). \tag{A.5}
$$

Now, by a similar argument using Chao *et al.* (2012, Lemma A3) by setting both "W_i and Y_i" in their notation as $\frac{1}{\sqrt{A_g}} q_g' y_{2i}$, we can show that

$$
\frac{1}{A_g}\sum_{i\neq j} P_{ij}^2 q_g' y_{2i} y_{2j}' u_i u_j q_g = \frac{1}{A_g}\sum_{i\neq j} P_{ij}^2 q_g' E(v_{2i}u_i|Z)E(v_{2j}'u_j|Z) q_g + o_p(1). \tag{A.6}
$$

Combining (A.5), (A.6) and the fact that $M_2 \leq \frac{1}{A_g}\sum_{i\neq j}(\hat{B}_{ij} - A_{iij} - B_{ij})$, we have $M_2 \xrightarrow{p} 0$.

Case (II). Next, we consider the case where $B_g/A_g \to \infty$. It follows $M_1 \leq \frac{1}{B_g}\sum_{i\neq j\neq k} q_g'(\hat{A}_{ijk} - A_{ijk}) q_g \xrightarrow{p} 0$ from Chao *et al.* (2012, Lemma A4) by setting "W_i, Y_j, and η_k" in their notation as $\frac{1}{\sqrt{B_g}} q_g' y_{2i}$, $\frac{1}{\sqrt{B_g}} q_g' y_{2j}$, and u_k^2, respectively. Note that $\frac{1}{B_g}\sum_{i\neq j\neq k} q_g' A_{ijk} q_g \xrightarrow{p} 0$ in this case.

For M_2, we apply Chao *et al.* (2012, Lemma A3) by setting "W_i and Y_i" in their notation with $\frac{1}{B_g}(q_g' y_{2i})^2$ and u_i^2, respectively, and it follows $M_2 \leq \frac{1}{B_g}\sum_{i\neq j}(\hat{B}_{ij} - A_{iij} - B_{ij}) \xrightarrow{p} 0$ by the same discussion as in Case (I).

A.3. **Proof of Theorem 2.** We show the theorem in the same way as in Theorem 1:

$$
\begin{aligned}
JLM &= (\hat{u}_0' P^{\#} Y_2)(\Psi^{\dagger})^{-1}(Y_2' P^{\#} \hat{u}_0) + o_p(1) \\
&= (u_0' P^{\dagger} Y_2)(\Psi^{\dagger})^{-1}(Y_2' P^{\dagger} u_0) + o_p(1) \\
&\xrightarrow{d} \chi^2_G.
\end{aligned} \tag{A.7}
$$

The first equality in (A.7) follows by the same argument as in Lemma 2, i.e., apply Chao *et al.* (2012, Lemmas A3 and A4) with P_{ij} replaced by $P^{\dagger}_{ij}$. Indeed, by noting that

$$
(P^{\dagger}_{ij})^2 = P^2_{2,ij} + 2P_{2,ii}P_{2,ij}P_{1,ij} + P^2_{2,ii}P^2_{1,ij} \leq P^2_{2,ij} + 2|P_{2,ij}P_{1,ij}| + P^2_{1,ij},
$$

and $y_{2i}P^{\dagger}_{ij} = y_{2i}P_{2,ij} + (y_{2i}P_{2,ii})P_{1,ij}$, we can show the same results as in Chao *et al.* (2012, Lemmas A3 and A4) with P_{ij} replaced by $P^{\dagger}_{ij}$.

The second equality in (A.7) follows from the relation in (2.5), which is shown as follows. Note that under H_0,

$$
\begin{aligned}
Y_2' P^{\#} \hat{u}_0 &= Y_2' P^{\#} (y_1 - Y_2 b_2 - Z_1 \hat{\gamma}(b)) \\
&= Y_2' P^{\#} u - Y_2' P^{\#} Z_1 (\hat{\gamma}(b) - \gamma) \\
&= Y_2' P^{\#} u - Y_2' P^{\#} P_1 u \\
&= Y_2' P^{\#} u + Y_2' \mathrm{diag}(P_2) P_1 u \\
&= Y_2' P^{\dagger} u + Y_2' \mathrm{diag}(P_2) \mathrm{diag}(P_1) u,
\end{aligned}
$$

where the fourth equality follows from $P_2 P_1 = 0$. Pick any G-dimensional vector c and let $\mu_c^2 = \sum_{i=1}^n (c'\Pi_2' z_i)^2$. Since

$$
\begin{aligned}
&Var\left(c'Y_2'\mathrm{diag}(P_2)\mathrm{diag}(P_1)u|Z\right) = \sum_{i=1}^n P^2_{2ii}P^2_{1ii} Var(c'y_{2i}u_i|Z) \leq C \sum_{i=1}^n P^2_{1ii} Var(c'y_{2i}u_i|Z) \\
\leq\ & C \sum_{i=1}^n P_{1ii} Var(c'y_{2i}u_i|Z) = O(K_1) = O(1),
\end{aligned}
$$

we have (I) $c'Y_2'\mathrm{diag}(P_2)\mathrm{diag}(P_1)u = o_p(\mu_c)$ when K/μ_c^2 is bounded, and (II) $c'Y_2'\mathrm{diag}(P_2)\mathrm{diag}(P_1)u = o_p(\sqrt{K})$ when $K/\mu_c^2 \to \infty$. Hence we have $c'Y_2'\mathrm{diag}(P_2)\mathrm{diag}(P_1)u = o_p(c'Y_2'P^{\dagger}u)$. Therefore the relation in (2.5) follows.

Finally, the convergence in (A.7) follows from the same argument as in Lemma 3. We note that

$$
c'Y_2'P^{\dagger}u = \sum_{j=1}^n \left\{ \sum_{i=1, i\neq j}^n c'\Pi_2' z_i (P_{2,ij} + P_{2ii}P_{1ij}) \right\} u_j + \sum_{i\neq j}^n c'v_{2i}P_{2,ij}u_j + \sum_{i\neq j}^n c'v_{2i}P_{2ii}P_{1ij}u_j
$$

and $\sum_{i\neq j}^n c'v_{2i}P_{2ii}P_{1ij}u_j = O_p(\sqrt{K_1}) = o_p(c'Y_2'P^{\dagger}u)$. Then we apply Chao *et al.* (2012, Lemma A2) by setting "U_i, ϵ_i, W_{in}, c_{1n}, and c_{2n}" in their notation as v_{2i}, u_i, $S^{-1}Q'\left\{\sum_{j=1,j\neq i}^n \Pi_2' z_j (P_{2,ji} + P_{2jj}P_{1ji})\right\} u_i$, $S\Lambda^{-1/2}\alpha$, and $\sqrt{K}\Lambda^{-1/2}\alpha$, respectively.

References

[1] Andrews, D. W. K., Moreira, M. J. and J. H. Stock (2006) Optimal two-sided invariant similar tests for instrumental variables regression with weak instruments, *Econometrica*, 74, 715-752.

[2] Andrews, D. W. K. and J. H. Stock (2007a) Inference with weak instruments, in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress*, vol. 3, Blundell, R., Newey, W. and T. Persson (eds.), Cambridge University Press, pp. 122-173.

[3] Andrews, D. W. K. and J. H. Stock (2007b) Testing with many weak instruments, *Journal of Econometrics*, 138, 24-46.

[4] Angrist, J. D., Imbens, G. W. and A. B. Krueger (1999) Jackknife instrumental variables estimation, *Journal of Applied Econometrics*, 14, 57-67.

[5] Billingsley, P. (1986) *Probability and Measure*, Wiley.

[6] Chao, J., Swanson, N. R., Hausman, J. A., Newey, W. K. and T. Woutersen (2012) Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments, *Econometric Theory*, 28, 42-86

[7] Fuller, W. A. (1977) Some properties of a modification of the limited information estimator, *Econometrica*, 45, 939-954.

[8] Hansen, C., Hausman, J. and W. K. Newey (2008) Estimation with many instrumental variables, *Journal of Business & Economic Statistics*, 26, 398-422.

[9] Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. and N. R. Swanson (2012) Instrumental variable estimation with heteroskedasticity and many instruments, *Quantitative Economics*, 3, 211-255.

[10] Kleibergen, F. (2002) Pivotal statistics for testing structural parameters in instrumental variables regression, *Econometrica*, 70, 1781-1803.

[11] Kleibergen, F. (2005) Testing parameters in GMM without assuming that they are identified, *Econometrica*, 73, 1103-1124.

[12] Moreira, M. J. (2001) Tests with correct size when instruments can be arbitrarily weak, Working paper.

[13] Moreira, M. J. (2003) A conditional likelihood ratio test for structural models, *Econometrica*, 71, 1027-1048.

[14] Newey, W. and F. Windmeijer (2009) GMM with many weak moment conditions, *Econometrica*, 77, 687-719.

[15] Wang, J. and E. Zivot (1998) Inference on structural parameters in instrumental variables regression with weak instruments, *Econometrica*, 66, 1389-1404.

Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan.

Email address: `matsushita.y@r.hit-u.ac.jp`

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: `t.otsu@lse.ac.uk`