A jackknife Lagrange multiplier test with many weak instruments*

Yukitoshi Matsushita† and Taisuke Otsu‡

August 15, 2022

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. In contrast to Crudu, Mellace and Sándor (2021) and Mikusheva and Sun (2021) who proposed jackknife Anderson-Rubin tests that are also robust to (i)-(iii), we modify a score statistic by jackknifing and construct its heteroskedasticity robust variance estimator. Compared to the Lagrange multiplier tests by Kleibergen (2002) and Moreira (2001) and their modification for many instruments by Hansen, Hausman and Newey (2008), our JLM test is robust to heteroskedastic errors and may circumvent a possible decrease in the power function. Simulation results illustrate the desirable size and power properties of the proposed method.

1 Introduction

In empirical applications of instrumental variable (IV) regression methods, researchers often face imprecise estimation results and so seek to employ many valid IVs to improve precision. However, statistical inference procedures in IV regression models can be crucially affected by the quality and number of the IVs. 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 test statistics can be poor. The use of many

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

†Graduate School of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan. Email: matsushita.y@r.hit-u.ac.jp

‡Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: t.otsu@lse.ac.uk.

instruments can improve efficiency of the estimators or their associated tests, but often leads the usual inference procedures to have poor finite sample properties (see, e.g., Andrews and Stock, 2007a, for a review).

In order to overcome the weak IV problem, several robust inference methods have been proposed. Kleibergen (2002) and Moreira (2001) proposed Lagrange multiplier (LM) type tests, while Moreira (2003) proposed a conditional likelihood ratio test, both of which are shown to be robust to the strength of the IVs.[1] 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 IV problem, and as such it is not clear how well (or how poorly) their tests perform with many instruments.

There have been many studies investigating the effects of many instruments. Linear models and asymptotics with many instruments were introduced by Kunitomo (1980) and Morimune (1983). Bekker (1994) pointed out that the many instruments asymptotic theory, where the number of instruments K may grow proportionally to the sample size n, may be suited better to applications, even when the number of instruments is moderate. Chao and Swanson (2005) generalized the many instruments asymptotic theory to allow for weaker instruments, where the concentration parameter may grow at a slower rate than n, and investigated conditions to achieve consistency for the k-class IV estimators. Han and Phillips (2006) further extended the many weak instruments asymptotic framework to study the asymptotic properties of the GMM estimator for possibly nonlinear models. 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 K satisfies $K^3/n \to 0$. We also refer the reader to 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 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 test based on heteroskedasticity and many instruments robust versions of the limited information maximum likelihood and Fuller (1977) estimators. These papers make assumptions on the rates of the concentration parameter or the number of instruments and/or homoskedasticity to achieve consistency of the point estimators and associated tests on parameter hypotheses. This paper complements these existing results by considering asymptotically valid tests under the null hypothesis with arbitrarily weak instruments, even though such tests may be inconsistent under

[1] Here robustness refers to size control under the null hypothesis on structural parameters. Under arbitrarily weak instruments, there is no consistent test in general (see Mikusheva and Sun (2021) and discussion below for the case of many and weak instruments).

the alternatives without further assumptions on the strength of instruments.

In this paper, we propose a jackknife Lagrange multiplier (JLM) test for IV 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. 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 Phillips and Hale (1977) and Angrist, Imbens and Krueger (1999), we re-center a score-type vector in the presence of many weak instruments and heteroskedasticity.[2] Compared to the LM tests by Kleibergen (2002) and Moreira (2001) and their modification for many instruments by Hansen, Hausman and Newey (2008), our JLM test is robust to heteroskedastic errors and may circumvent a possible decrease in the power function. In particular, the power of our test does not decline asymptotically in any region under an additional requirement on strength of the instruments. Furthermore, the Wald statistic introduced by Hausman *et al.* (2012) is not fully robust to weak instruments because it relies on the consistency of their heteroskedastic limited information maximum likelihood estimator. Our JLM test can be a useful complement to Hausman *et al.'s* (2012) Wald test if the researcher is primarily concerned with the size properties of the tests. The JLM test is asymptotically valid under the null hypothesis with arbitrarily weak instruments even though it is generally inconsistent under fixed alternatives without further requirements on the strength of the instruments. Simulation results illustrate the desirable size robustness properties of the proposed method.

Many papers in the econometrics literature have applied the idea of jackknifing for IV regression models. Phillips and Hale (1977), Angrist, Imbens and Krueger (1999), and Blomquist and Dahlberg (1999) proposed the jackknife IV estimator (JIVE), which aims at eliminating the correlation between the first stage fitted values and structural equation errors. Hahn, Hausman and Kuersteiner (2004) studied higher-order properties of the jackknife two-stage least squares estimator. Davidson and MacKinnon (2006) conducted an extensive simulation study on the finite sample performance of the JIVE. Ackerberg and Devereux (2009) proposed a bias-corrected JIVE and investigated its asymptotic properties under the many instruments asymptotics and heteroskedastic errors. Chao *et al.* (2012) studied asymptotic properties of the JIVE under the many-weak instruments asymptotics and heteroskedastic errors. Newey and Windmeijer (2009) extended the JIVE to the GMM context. Hansen and Kozbur (2014) proposed a regularized JIVE to deal with the case where the number of instruments may be larger than the sample size.

[2] In the context of overidentifying restriction testing, Chao *et al.* (2014) proposed a jackknife version of the conventional overidentifying restriction test statistic, which is robust to many instruments and heteroskedastic errors. In contrast, this paper is concerned with parameter hypothesis testing.

Recently and independently, Crudu, Mellace and Sándor (2021) and Mikusheva and Sun (2021) have proposed jackknife Anderson-Rubin tests, which are asymptotically size correct under (i)-(iii). These tests use jackknifing to re-center the Anderson-Rubin statistic in the presence of many weak instruments and heteroskedasticity. Furthermore, Mikusheva and Sun (2021) developed a novel variance estimator based on cross-fitting in the spirit of Kline, Saggio and Sølvsten (2020) to standardize the jackknifed Anderson-Rubin statistic. In contrast to these recent papers, we apply jackknifing to the score statistic. Thus the construction and theoretical developments for our statistic are different, and this paper may be considered as a complement to the Anderson-Rubin approach in Crudu, Mellace and Sándor (2021) and Mikusheva and Sun (2021). Simulation results indicate that our JLM statistic compares favorably with the jackknife Anderson-Rubin statistic, even though a formal analysis to compare these statistics is beyond the scope of this paper.

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 JLM 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. Section 3 conducts a simulation study and presents a real data example. Finally, Section 4 concludes.

2 Main results

2.1 Setup

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

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

for $i = 1, \ldots, n$, 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 (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 (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 IVs for (1). The number of instruments $K_2 = K_{2n}$ may grow with the sample size n, and thus the joint distribution of (y_i', z_i') is allowed to vary with n. We also assume that $(u_1, \ldots, u_n)$ are mutually independent conditional on $(z_1, \ldots, z_n)$ with $E(u_i|z_i) = 0$ almost surely for $i = 1, \ldots, n$. The reduced form of y_i is defined as

$$y_i = \Pi_n' z_i + v_i = \begin{pmatrix} \pi_{1n}' \\ \Pi_{2n}' \end{pmatrix} z_i + \begin{pmatrix} v_{1i} \\ v_{2i} \end{pmatrix}, \tag{2}$$

where π_{1n} is a K-dimensional vector and Π_{2n} 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. $(v_1, \dots, v_n)$ are mutually independent conditional on $(z_1, \dots, z_n)$ with $E(v_i|z_i) = 0$ almost surely.

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_{2n} may be zero), and (iii) the error term u_i may be heteroskedastic and non-normal. For this setting, we develop a new robust test statistic.

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 $Y_2 = (y_{21}, \dots, y_{2n})'$, $Z = (z_1, \dots, z_n)'$, and $V_2 = (v_{21}, \dots, v_{2n})'$ be matrices for the endogenous regressors, instruments, and reduced form errors, respectively. Although the number of columns K of Z grows with the sample size n, we suppress the dependence and denote Z_{K_n} by Z. We also define the observables $u_{0i} = y_{1i} - y_{2i}'b$ and $u_0 = (u_{01}, \dots, u_{0n})'$. Finally, we define the matrix P^* by $P_{ij}^* = P_{ij}$ for $i \neq j$ and $P_{ii}^* = 0$ for all i, where P is the projection matrix $Z(Z'Z)^{-1}Z'$.

We note that under the null hypothesis $H_0 : \beta = b$, the score-type vector $Y_2'Pu_0$ is not necessarily centered, i.e., $E(Y_2'Pu_0) = E(V_2'Pu_0)$ may not be zero. This is due to the fact that $E(V_{2i}P_{ii}u_{0i})$ may not be zero. 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 $E(Y_2'P^*u_0) = 0$.

By inserting the reduced form $y_{2i} = \Pi_{2n}'z_i + v_{2i}$, the (conditional) variance of $Y_2'P^*u_0$ is written as

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

where $\sigma_i^2 = E(u_i^2|Z)$. The first and second terms come from the components $\Pi_{2n}' z_i$ and v_{2i} in the reduced form, respectively. Under the conventional asymptotic framework with a fixed number of strong instruments, the first term dominates. On the other hand, under the many and weak instruments setup as in this paper, both terms may be of the same order. Note that this variance formula allows for heteroskedastic errors. Observe that Ψ_n can be alternatively written as

$$\Psi_n = E\left[\sum_{i,j,k,i\neq k,j\neq k}^{n} y_{2i} P_{ik} u_{0k}^2 P_{kj} y_{2j}' + \sum_{i\neq j}^{n} y_{2i} y_{2j}' u_{0i} u_{0j} P_{ij}^2 \middle| Z\right].$$

Based on this expression, we estimate the variance Ψ_n by

$$\hat{\Psi}_n = Y_2' P^* \Sigma_0 P^* Y_2 + \sum_{i,j=1}^{n} y_{2i} y_{2j}' u_{0i} u_{0j} P_{ij}^{*2}, \tag{3}$$

where $\Sigma_0 = \mathrm{diag}(u_{01}^2, \ldots, u_{0n}^2)$. By standardizing the jackknife score vector by this variance estimator, our JLM test statistic for testing $H_0 : \beta = b$ is defined as

$$JLM(b) = (u_0' P^* Y_2) \hat{\Psi}_n^{-1} (Y_2' P^* u_0). \tag{4}$$

Compared to the standard LM statistic, $(u_0' P Y_2) \left[\hat{\sigma}^2 (Y_2' P Y_2)\right]^{-1} (Y_2' P u_0)$, for some homoskedastic error variance estimator $\hat{\sigma}^2$ (Wang and Zivot, 1998), the major differences of our approach are the use of the jackknife score $Y_2' P^* u_0$ instead of $Y_2' P u_0$ and the use of the heteroskedasticity robust variance estimator $\hat{\Psi}_n$ instead of $\hat{\sigma}^2 (Y_2' P Y_2)$. Note that $Y_2' P^* u_0$ and $Y_2' P u_0$ are asymptotically equivalent under the conventional asymptotics with a fixed number of strong instruments.

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

Assumption 1. *(i) For each n, Z is of full column rank almost surely, and there exists a constant $c \in [P_{ii}, 1)$ almost surely for all $i = 1, \ldots, n$. (ii) For each n, conditional on Z, $\{(u_i, v_{2i}')\}_{i=1}^n$ are independent with $E(u_i|Z) = 0$ and $E(v_{2i}|Z) = 0$ almost surely. (iii) There exists a positive constant C (which is independent of n) such that for each n, $\max_{i=1,\ldots,n} E(u_i^4|Z) \leq C$, $\max_{i=1,\ldots,n} E(||v_{2i}||^4|Z) \leq C$, and $\max_{i=1,\ldots,n} |z_i' \pi_{2s}|^4 < C$ for all $s = 1, \ldots, G$ almost surely, where π_{2s} is the s-th column of Π_{2n}. (iv) There exists a positive constant C_1 (which is independent of n) such that for each n, $\max_{i=1,\ldots,n} |corr(c' v_{2i}, u_i|Z)| < C_1 < 1$ almost surely for any $c \neq 0$.*

We note that the distribution of the data (y_i', z_i') is allowed to vary with n. Assumption 1 (i)-(iii) are also imposed in existing papers on many weak IV regressions, such as Chao *et al.*, (2012) and Hausman *et al.* (2012). Assumption 1 (i) is on Z and implies $K < n$. Assumption 1

(ii) is a standard exogeneity condition for instruments, and Assumption 1 (iii) contains regularity conditions for the fourth conditional moments of the error terms, which are used to apply central limit theorems. Assumption 1 (iv) is a mild condition which guarantees positive definiteness of Ψ_n (see Lemma 1 in Appendix).[3] Under these assumptions, the limiting null distribution of the JLM statistic is obtained as follows:

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

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

This theorem proves the asymptotic pivotalness of the JLM statistic under the conditions that allow for (i) arbitrarily weak instruments, (ii) many instruments in the sense that $K/n \to \alpha \in [0, 1)$, and (iii) heteroskedasticity. By inverting $JLM(b)$, the JLM-based $100(1-a)\%$ confidence set can be obtained as $\{b : JLM(b) \leq \chi^2_{G,a}\}$, where $\chi^2_{G,a}$ is the $(1-a)$-th quantile of the χ^2_G distribution.[4]

We note that this theorem does not cover the case where K is fixed. In this case, we can still obtain the same conclusion as far as the instruments are strong enough (in the sense that Π_{2n} is fixed or decays to zero slower than the $\sqrt{n}$-rate).

Furthermore, 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. Recently and independently, Crudu, Mellace and Sándor (2021) and Mikusheva and Sun (2021) have proposed jackknife Anderson-Rubin tests, which are asymptotically size correct under the setup of Theorem 1. It is beyond the scope of this paper to compare our JLM test with these tests under the many and weak instruments setup. However, under the conventional asymptotic framework with a fixed number of strong instruments, we can see that the limiting null distribution of their jackknife Anderson-Rubin statistics is $(\chi^2_K - K)/\sqrt{2K}$, instead of χ^2_G for the JLM statistic. Therefore, under the conventional asymptotics, the JLM statistic will exhibit better power properties when $K > G$.

[3] Although Ψ_n is shown to be positive definite for each n almost surely, its eigenvalues typically diverge. Therefore, the argument based on the continuous mapping theorem for separately taking the limits for $\hat{\Psi}_n$ and $Y_2' P^* u_0$ is not applicable here if we want to derive the limiting distribution of $JLM(b)$. As shown in Lemma 3, we take the limit for the whole quadratic part of the dominant term of $JLM(b)$, where the diverging eigenvalues of Ψ_n are internally normalized.

[4] We note that Theorem 1 only guarantees pointwise asymptotic validity of the JLM test and confidence set by using the χ^2 critical value (i.e., the limit is taken under each null distribution). Although it is beyond the scope of this paper, it is interesting to assess uniform asymptotic size or coverage properties based on our JLM statistic by applying the generic results in Andrews, Cheng and Guggenberger (2020).

We next study power properties of the JLM test. As indicated by the impossibility result in Dufour (1997), we cannot achieve consistency of the JLM test under fixed alternatives without further assumptions. Indeed based on Mikusheva and Sun (2021), there exists no consistent test for the null $H_0 : \beta = b$ unless some condition on the concentration parameters, which guarantees sufficiently strong instruments, is satisfied. As such, we derive the consistency of the JLM test under this additional requirement.

Furthermore, we note that the LM statistic by Kleibergen (2002) and Moreira (2001) and its modification by Hansen, Hausman and Newey (2008) may lose power in some regions for the alternative hypotheses. This lack 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 they 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 (conventional) score of the concentrated likelihood. Specifically, we can show that under an additional requirement on strength of instruments, the power curve of the JLM statistic shows monotonicity in an asymptotic sense.

The power properties of the JLM test discussed above are described as follows.

Theorem 2. *Suppose Assumption 1 holds true, $K \to \infty$, and $K/n \to \alpha \in [0, 1)$ as $n \to \infty$. Additionally, assume that $\frac{K}{\min\{\mu_1^4,\ldots,\mu_G^4\}} \to 0$, where $(\mu_1, \ldots, \mu_G)$ are the concentration parameters defined in (10). Then under the alternative $H_1 : \beta = b + \Delta$ for a fixed $\Delta \neq 0$,*

(i) *$P\{JLM(b) \geq \chi^2_{G,a}\} \to 1$ as $n \to \infty$, where $\chi^2_{G,a}$ is the $(1-a)$-th quantile of the χ^2_G distribution,*

(ii) *there exists some $C > 0$ such that*

$$P\{JLM(b) \geq C||b - \beta||^2 \quad \textit{for each } b\} \to 1 \quad \textit{as } n \to \infty.$$

The additional assumption $\frac{K}{\min\{\mu_1^4,\ldots,\mu_G^4\}} \to 0$ is on the strength of the instruments. For example, Chao and Swanson (2005) imposed this assumption to achieve consistency of point estimators under their many weak instruments asymptotics. This theorem says that the JLM test is consistent for any fixed $\Delta \neq 0$ and that the JLM statistic $JLM(b)$ for testing $H_0 : \beta = b$ increases monotonically as $||b - \beta||$ increases with probability approaching to 1.

We note that the consistency and power monotonicity results presented in Theorem 2 do not contradict the impossibility result given in Dufour (1997), which says that any valid confidence set with level $1 - \alpha$ must be unbounded with probability close to $1 - \alpha$ in the neighborhood of nonidentification regions. The reason is that these power properties of the JLM test are derived under the additional condition $\frac{K}{\min\{\mu_1^4,\ldots,\mu_G^4\}} \to 0$, which requires sufficiently strong instruments relative to the number of instruments K. Indeed, based on Mikusheva and Sun (2021), unless $\frac{K}{\min\{\mu_1^4,\ldots,\mu_G^4\}} \to 0$, we can conclude that there exists no consistent test for the null $H_0 : \beta = b$.

2.3 Models with exogenous regressors

In this subsection, we extend our analysis to models with exogenous regressors in (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,

$$(\hat{u}_0'P^{\#}Y_2)(\Psi_n^{\dagger})^{-1}(Y_2'P^{\#}\hat{u}_0) = (u_0'P^{\dagger}Y_2)(\Psi_n^{\dagger})^{-1}(Y_2'P^{\dagger}u_0) + o_p(1), \tag{5}$$

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

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

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 JLM statistic for this general case using the quadratic form:

$$JLM(b) = (\hat{u}_0'P^{\#}Y_2)\tilde{\Psi}_n^{-1}(Y_2'P^{\#}\hat{u}_0), \tag{6}$$

where

$$\tilde{\Psi}_n = Y_2'P^{\dagger}\hat{\Sigma}_0 P^{\dagger\prime}Y_2 + \sum_{i,j=1}^{n} y_{2i}y_{2j}'\hat{u}_{0i}\hat{u}_{0j}(P_{ij}^{\dagger})^2, \tag{7}$$

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

The asymptotic property of this JLM statistic is obtained as follows.

Theorem 3. *Suppose Assumption 1 holds true, $K \to \infty$, and $K/n \to \alpha \in [0, 1)$ as $n \to \infty$.*

Then under $H_0 : \beta = b$,

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

Similar comments to Theorem 1 apply. For example, the JLM-based $100(1-a)\%$ confidence set for β in this setup can be constructed as $\{b : JLM(b) \leq \chi^2_{G,a}\}$.

3 Numerical illustrations

3.1 Simulation

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

$$\begin{aligned} y_{1i} &= y_{2i}\beta_0 + z_{1i}\gamma_0 + u_i, \\ y_{2i} &= z_i'\pi_2 + v_{2i}, \end{aligned} \tag{8}$$

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 $(u_i, v_{2i}) = ((1 + \phi z_{21i})\epsilon_{1i}, \rho u_i + \sqrt{1-\rho^2}\epsilon_{2i})$, where ϵ_{1i} and ϵ_{2i} are independent and drawn from $N(0,1)$.[5] We set $n = 200$ 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 $\{z_i\}$)

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

We investigate size properties of eleven 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) using the asymptotic χ^2 critical value, (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), (viii)

[5] In our preliminary simulation, we also consider the cases of (a) the t_5 and χ_3 distributions for error terms as examples of fat-tailed and skewed errors, (b) truncated support of z_i to satisfy Assumption 1 (iii), (c) $\{z_i\}$ kept fixed throughout Monte Carlo replications, and (d) identification failure by setting $d = 0$. For (a)-(c), the results are overall similar to the ones in Table 2. For (d), the size properties are also similar to the ones in Table 2.

the heteroskedasticity robust version of KLM by Kleibergen (2005) (HKLM), (viiii) the heteroskedasticity robust version of CLR by Kleibergen (2005) (HCLR), (x) the heteroskedasticity robust version of AR by Mikusheva and Sun (2021) (HAR), and (xi) the proposed JLM test (JLM). The number of Monte Carlo repetitions in each experiment is 10,000.

Tables 2 and 3 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) HCLR, HKLM, and HAR work relatively well for both the cases of homoskedastic and heteroskedastic errors. However, HCLR and HKLM tend to under- or over-reject when δ^2 is small. HAR works best among these although it tends to mildly over-reject for all cases.

vi) 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, as we would expect from 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 $H_1 : \beta = \beta_0 + \Delta$. We focus on HAR, mKLM, and JLM since the size distortions of the other tests are severe when the errors are heteroskedastic and/or the number of instruments is large. Figures 1-3 display the calibrated power curves 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 $n = 200$, $K = 30$, and $\rho = 0.2$ for $\delta^2 = 60, 30, 10$ as typical examples. First, JLM and mKLM are more powerful when $|\Delta|$ is small but less powerful when $|\Delta|$ is large compared to HAR. Although it is beyond the scope of this paper, we conjecture that the power property of JLM for large $|\Delta|$ can be improved by using a cross-fit variance estimator as in Mikusheva and Sun (2021). Second, mKLM exhibits declines of power in some regions for the alternative hypotheses while the power curves of JLM and HAR are monotone for all cases. This result is consistent with our theoretical finding in Theorem 2.

3.2 Real data example

We compare the confidence sets of the effect of schooling on log weekly wage with the specification underlying Table VII Column (6) of Angrist and Krueger (1991) using their original data. We focus on the specification with 180 and 1,530 instruments as in Mikusheva and Sun (2021). Table 1 reports the 95% confidence sets based on the HAR statistic by Mikusheva and Sun (2021), and the proposed JLM statistic. In this application, the confidence sets based on JLM are narrower than those based on HAR. We conjecture that this is due to the better power property of the JLM test for small values of $|\Delta|$ as illustrated in the simulation study above.

	HAR	JLM
180 instruments	[0.008, 0.201]	[0.067, 0.133]
1530 instruments	[-0.047, 0.202]	[0.025, 0.123]

Table 1: 95% confidence sets of the effect of schooling on log weekly wage using Angrist and Krueger's (1991) data

4 Conclusion

By modifying the score statistic based on jackknifing combined with heteroskedasticity robust estimation for its variance component, we propose a new jackknife Lagrange multiplier test for parameter hypotheses on instrumental variable regression models. Our test is easy to implement and robust not only to many and arbitrarily weak instruments but also to heteroskedastic errors. Simulation results endorse desirable size and power properties of the proposed test. It is interesting to adapt our idea of jackknifing to other tests, such as Moreira's (2003) conditional likelihood ratio test and its heteroskedasticity and autocorrelation robust version by Moreira and

Moreira (2019), to robustify these tests under many instruments asymptotics. Such extensions are currently under investigation by the authors.

ρ	δ^2	K	t_{TS}	t_{LI}	t_{HLI}	AR	CLR	KLM	mKLM	HKLM	HCLR	HAR	JLM
0.2	30	5	0.045	0.048	0.048	0.056	0.055	0.053	0.052	0.040	0.048	0.085	0.044
	30	10	0.056	0.064	0.036	0.055	0.057	0.055	0.053	0.038	0.048	0.074	0.051
	30	30	0.123	0.132	0.027	0.075	0.066	0.063	0.051	0.032	0.047	0.078	0.050
	30	90	0.391	0.342	0.024	0.129	0.146	0.127	0.052	0.112	0.132	0.087	0.053
	10	5	0.033	0.034	0.023	0.051	0.053	0.054	0.051	0.031	0.035	0.080	0.047
	10	10	0.058	0.061	0.019	0.058	0.057	0.052	0.046	0.032	0.032	0.077	0.050
	10	30	0.161	0.156	0.014	0.071	0.076	0.068	0.047	0.030	0.040	0.076	0.050
	10	90	0.447	0.370	0.015	0.126	0.158	0.140	0.047	0.110	0.131	0.082	0.049
	2	5	0.018	0.015	0.008	0.054	0.056	0.055	0.048	0.016	0.013	0.082	0.049
	2	10	0.054	0.042	0.008	0.058	0.060	0.059	0.047	0.019	0.014	0.079	0.050
	2	30	0.179	0.152	0.009	0.067	0.073	0.072	0.047	0.026	0.029	0.072	0.051
	2	90	0.484	0.371	0.012	0.126	0.158	0.145	0.049	0.105	0.131	0.086	0.052
0.6	30	5	0.086	0.055	0.055	0.059	0.052	0.052	0.052	0.042	0.051	0.085	0.049
	30	10	0.176	0.062	0.049	0.057	0.055	0.053	0.052	0.039	0.049	0.075	0.050
	30	30	0.696	0.097	0.046	0.072	0.063	0.059	0.049	0.034	0.048	0.075	0.046
	30	90	0.999	0.253	0.069	0.128	0.136	0.119	0.049	0.112	0.139	0.088	0.051
	10	5	0.131	0.073	0.068	0.052	0.053	0.052	0.050	0.035	0.040	0.082	0.048
	10	10	0.312	0.099	0.064	0.055	0.055	0.055	0.050	0.034	0.036	0.078	0.049
	10	30	0.879	0.201	0.081	0.073	0.070	0.063	0.046	0.031	0.042	0.076	0.048
	10	90	1.000	0.349	0.097	0.129	0.154	0.136	0.046	0.109	0.143	0.086	0.050
	2	5	0.210	0.114	0.075	0.052	0.051	0.050	0.045	0.018	0.016	0.080	0.042
	2	10	0.497	0.197	0.092	0.055	0.058	0.050	0.042	0.024	0.016	0.075	0.042
	2	30	0.945	0.324	0.104	0.070	0.079	0.069	0.046	0.029	0.032	0.074	0.047
	2	90	1.000	0.448	0.125	0.124	0.161	0.148	0.050	0.110	0.151	0.082	0.051

Table 2: Empirical rejection frequencies at 5% significant level: Homoskedastic errors

ρ	δ^2	K	t_{TS}	t_{LI}	t_{HLI}	AR	CLR	KLM	mKLM	HKLM	HCLR	HAR	JLM
0.2	30	5	0.091	0.098	0.050	0.088	0.113	0.110	0.108	0.035	0.044	0.095	0.045
	30	10	0.096	0.107	0.044	0.080	0.099	0.097	0.093	0.037	0.040	0.081	0.045
	30	30	0.139	0.155	0.025	0.076	0.083	0.076	0.062	0.030	0.043	0.070	0.046
	30	90	0.408	0.351	0.024	0.124	0.152	0.132	0.051	0.114	0.136	0.082	0.048
	10	5	0.068	0.067	0.029	0.088	0.108	0.102	0.098	0.025	0.026	0.096	0.045
	10	10	0.083	0.085	0.019	0.077	0.091	0.087	0.078	0.025	0.026	0.087	0.050
	10	30	0.172	0.168	0.017	0.082	0.088	0.076	0.054	0.028	0.036	0.076	0.045
	10	90	0.458	0.378	0.016	0.129	0.162	0.147	0.049	0.104	0.128	0.082	0.050
	2	5	0.030	0.026	0.008	0.082	0.085	0.073	0.063	0.013	0.012	0.091	0.044
	2	10	0.066	0.052	0.008	0.078	0.081	0.067	0.054	0.018	0.013	0.081	0.047
	2	30	0.197	0.167	0.009	0.081	0.090	0.078	0.050	0.025	0.027	0.077	0.048
	2	90	0.493	0.380	0.012	0.128	0.161	0.146	0.047	0.107	0.127	0.083	0.053
0.6	30	5	0.128	0.097	0.064	0.091	0.116	0.112	0.111	0.040	0.049	0.098	0.046
	30	10	0.201	0.095	0.057	0.081	0.099	0.096	0.093	0.039	0.046	0.088	0.049
	30	30	0.708	0.130	0.056	0.082	0.088	0.081	0.068	0.031	0.044	0.075	0.050
	30	90	0.999	0.263	0.071	0.130	0.148	0.126	0.055	0.116	0.136	0.086	0.050
	10	5	0.170	0.112	0.076	0.092	0.107	0.104	0.100	0.028	0.033	0.098	0.044
	10	10	0.329	0.134	0.077	0.079	0.094	0.089	0.082	0.032	0.035	0.081	0.046
	10	30	0.883	0.213	0.080	0.081	0.087	0.076	0.057	0.029	0.043	0.075	0.047
	10	90	0.999	0.356	0.098	0.124	0.151	0.134	0.046	0.111	0.142	0.081	0.045
	2	5	0.248	0.149	0.091	0.087	0.095	0.086	0.077	0.014	0.014	0.092	0.032
	2	10	0.515	0.214	0.095	0.077	0.083	0.069	0.058	0.020	0.015	0.082	0.043
	2	30	0.948	0.338	0.107	0.083	0.089	0.075	0.052	0.026	0.031	0.077	0.050
	2	90	1.000	0.446	0.124	0.131	0.165	0.149	0.049	0.108	0.148	0.083	0.047

Table 3: Empirical rejection frequencies at 5% significant level: Heteroskedastic errors

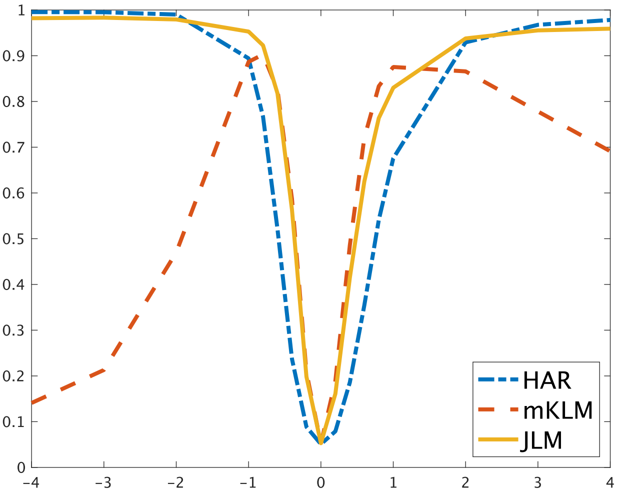

Figure 1: Calibrated power curves: $n = 200$, $K = 30$, $\rho = 0.2$, $\delta^2 = 60$, Heteroskedastic errors. Readers are referred to the online version of the paper for colored graphics.

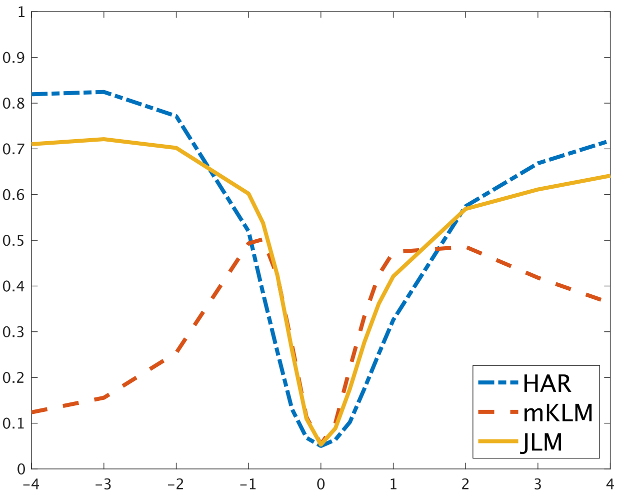

Figure 2: Calibrated power curves: $n = 200$, $K = 30$, $\rho = 0.2$, $\delta^2 = 30$, Heteroskedastic errors. Readers are referred to the online version of the paper for colored graphics.

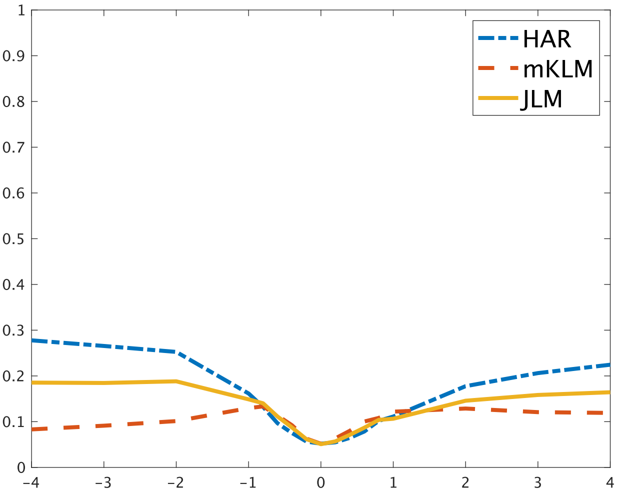

Figure 3: Calibrated power curves: $n = 200$, $K = 30$, $\rho = 0.2$, $\delta^2 = 10$, Heteroskedastic errors. Readers are referred to the online version of the paper for colored graphics.

A Proof

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

$$Q_n' \Psi_n Q_n = \Lambda_n = \mathrm{diag}(\lambda_1, \ldots, \lambda_G). \tag{9}$$

Also define $u_{0i} = y_{1i} - y_{2i}' b$, and

$$\begin{aligned} A_{ijk} &= \sigma_k^2 \Pi_{2n}' z_i P_{ik} P_{kj} z_j' \Pi_{2n}, \qquad B_{ij} = P_{ij}^2 \{E(v_{2i} v_{2i}' | Z) \sigma_j^2 + E(v_{2i} u_i | Z) E(v_{2j}' u_j | Z)\}, \\ \hat{A}_{ijk} &= u_{0k}^2 y_{2i} P_{ik} P_{kj} y_{2j}', \qquad \hat{B}_{ij} = P_{ij}^2 (y_{2i} y_{2i}' u_{0j}^2 + y_{2i} y_{2j}' u_{0i} u_{0j}), \end{aligned}$$

so that $\Psi_n = \sum_{i,j,k,i \neq k, j \neq k}^n A_{ijk} + \sum_{i \neq j}^n B_{ij}$ and $\hat{\Psi}_n = \sum_{i \neq j \neq k} \hat{A}_{ijk} + \sum_{i \neq j}^n \hat{B}_{ij}$ under H_0. Based on $Q_n = (q_1, \ldots, q_G)$, denote

$$\mu_g^2 = \sum_{i=1}^n q_g' \Pi_{2n}' z_i z_i' \Pi_{2n} q_g, \tag{10}$$

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

A.1 Proof of Theorem 1

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

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

under $H_0 : \beta = b$, 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_n is positive definite almost surely for each $n > K$.*

Proof: Pick any $n > K$ and nonzero G-dimensional vector c. Then $c' \Psi_n c = A + B$, 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

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

For B, we have

$$\begin{aligned} B &= \sum_{i<j}^{n} P_{ij}^2 \{E[(c'v_{2i})^2|Z]\sigma_j^2 + E[(c'v_{2j})^2|Z]\sigma_i^2 + 2E(c'v_{2i}u_i|Z)E(c'v_{2j}u_j|Z)\} \\ &\geq \frac{1}{2}\sum_{i\neq j}^{n} P_{ij}^2 \{E[(c'v_{2i})^2|Z]\sigma_j^2 + E[(c'v_{2j})^2|Z]\sigma_i^2 - |2E(c'v_{2i}u_i|Z)E(c'v_{2j}u_j|Z)|\}. \end{aligned} \quad (11)$$

Also, the 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

$$E[(c'v_{2i})^2|Z]\sigma_j^2 + E[(c'v_{2j})^2|Z]\sigma_i^2 - |2E(c'v_{2i}u_i|Z)E(c'v_{2j}u_j|Z)| > 0, \quad (12)$$

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

$$\sum_{i\neq j} P_{ij}^2 = K - \sum_{i=1}^n P_{ii}^2 \geq K\left(1 - \max_{1\leq i\leq n} P_{ii}\right) > 0, \quad (13)$$

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

Combining (11)-(13), we obtain $B > 0$ almost surely, and the conclusion follows.

Lemma 2. *Under Assumption 1 and $H_0 : \beta = b$,*

$$JLM(b) = u'P^*Y_2Q_n\Lambda_n^{-1}Q_n'Y_2'P^*u + o_p(1).$$

Proof: By Lemma 1 and (9), the LM statistic can be written as

$$\begin{aligned} JLM(b) &= u_0'P^*Y_2\{\Psi_n + (\hat{\Psi}_n - \Psi_n)\}^{-1}Y_2'P^*u_0 \\ &= u_0'P^*Y_2\{Q_nQ_n'\Psi_nQ_nQ_n' + Q_nQ_n'(\hat{\Psi}_n - \Psi_n)Q_nQ_n'\}^{-1}Y_2'P^*u_0 \\ &= u_0'P^*Y_2Q_n\Lambda_n^{-1/2}\{I + \Lambda_n^{-1/2}Q_n'(\hat{\Psi}_n - \Psi_n)Q_n\Lambda_n^{-1/2}\}^{-1}\Lambda_n^{-1/2}Q_n'Y_2'P^*u_0. \end{aligned}$$

Pick any $g, h = 1, \ldots, G$. Then it follows from Lemma 4 that

$$\frac{1}{\sqrt{\lambda_g\lambda_h}} q_g'(\hat{\Psi}_n - \Psi_n)q_h \xrightarrow{p} 0.$$

Thus, we obtain $\Lambda_n^{-1/2}Q_n'(\hat{\Psi}_n - \Psi_n)Q_n\Lambda_n^{-1/2} \xrightarrow{p} 0$, and Slutsky's lemma yields the conclusion.

Lemma 3. *Under Assumption 1,*

$$u'P^*Y_2Q_n\Lambda_n^{-1}Q_n'Y_2'P^*u \xrightarrow{d} \chi^2_G.$$

Proof: Without loss of generality, we assume that $\frac{K}{\mu_1^2} \nrightarrow \infty, \ldots, \frac{K}{\mu_{G_1}^2} \nrightarrow \infty$ and $\frac{K}{\mu_{G_1+1}^2} \to \infty, \ldots, \frac{K}{\mu_G^2} \to \infty$. Pick any nonzero G_2-dimensional vector ξ, and define $S_n = \mathrm{diag}(\mu_1, \ldots, \mu_g, \sqrt{K}, \ldots, \sqrt{K})$. Observe that

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

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, $S_n^{-1}Q_n'\Pi_{2n}'z_i(1-P_{ii})u_i$, $S_n\Lambda_n^{-1/2}\xi$, and $\sqrt{K}\Lambda_n^{-1/2}\xi$, respectively. It is straightforward to verify that the conditions of Chao *et al.* (2012, Lemma A.2) are satisfied.[6] Thus, by the Cramér-Wold device, we have

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

which implies the conclusion.

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

$$\frac{1}{\sqrt{\lambda_g\lambda_h}}q_g'(\hat{\Psi}_n - \Psi_n)q_h \xrightarrow{p} 0.$$

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

[6] Conditions (i)-(iii) of Chao *et al.* (2012, Lemma A2) are directly verified from Assumption 1. Condition (iv) of Chao *et al.* (2012, Lemma A2) can be verified as

$$\begin{aligned}
\sum_{i=1}^n E(||W_{in}||^4|Z) \quad &\leq \quad C\sum_{i=1}^n\left\{\sum_{g=1}^{G_1}\frac{1}{\mu_g^4}(z_i'\Pi_{2n}q_g)^4 + \sum_{g=G_1+1}^{G}\frac{1}{K^2}(z_i'\Pi_{2n}q_g)^4\right\} \\
&\leq \quad C\left\{\max_{1\leq i\leq n}(z_i'\Pi_{2n}q_g)^2\right\}\left\{\sum_{g=1}^{G_1}\frac{1}{\mu_g^2} + \sum_{g=G_1+1}^{G}\frac{\mu_g^2}{K^2}\right\} \to 0,
\end{aligned}$$

almost surely for some $C > 0$, where the first inequality follows from the definition of $W_{in} = S_n^{-1}Q_n'\Pi_{2n}'z_i(1-P_{ii})u_i$ and Assumption 1 (i) and (iii), and the convergence follows from $\mu_g^2 \to \infty$ for $g = 1, \ldots, G_1$ (because of $K \to \infty$), the assumption $\frac{K}{\mu_g^2} \to \infty$ for $g = G_1+1, \ldots, G$, and boundedness of $\max_{1\leq i\leq n}(z_i'\Pi_{2n}q_g)^2$ (by Assumption 1 (iii)). Furthermore, Chao *et al.* (2012, Lemma A2) require $||c_{1n}|| \leq C$ and $||c_{2n}|| \leq C$ for some $C > 0$. In our case, $c_{1n} = S_n\Lambda_n^{-1/2}\xi$ and $c_{2n} = \sqrt{K}\Lambda_n^{-1/2}\xi$ satisfy these requirements because of the fact that $\lambda_g = A_g + B_g \geq A_g$ and $\mu_g^2 = O(A_g)$ by the proof of Lemma 4.

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

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

It is enough to show that $M_1, M_2 \xrightarrow{p} 0$. Based on (9), 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 $B_g = \sum_{i\neq j}^n q_g' B_{ij} q_g$. We note that $\sqrt{\lambda_g \lambda_h} \geq \max\{\sqrt{A_g A_h}, \sqrt{B_g B_h}\}$ since $A_g, A_h \geq 0$ and $B_g, B_h > 0$ from Lemma 1. We consider two cases: (I) $\frac{K}{\mu_g^2 \mu_h^2} \to 0$, and (II) $\frac{K}{\mu_g^2 \mu_h^2} \nrightarrow 0$.

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

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

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

$$E(W_i|Z) = \frac{1}{\sqrt{A_g A_h}} (q_g' \Pi_{2n}' z_i z_i' \Pi_{2n} q_h) + \frac{1}{\sqrt{A_g A_h}} q_g' E(v_{2i} v_{2i}' |Z) q_h,$$

so that

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

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

$$
\begin{aligned}
& \max_{1\le i\le n} Var(W_i|Z) = \max_{1\le i\le n} Var\left(\frac{1}{\sqrt{A_g A_h}} \left\{ q_g'(\Pi_{2n}' z_i + v_{2i})(\Pi_{2n}' z_i + v_{2i})' q_h \right\} \middle| Z \right) \\
= \;& \max_{1\le i\le n} \frac{1}{A_g A_h} \left\{ (q_g'\Pi_{2n}' z_i)^2 E(\tilde{v}_{ig}^2|Z) + (q_h'\Pi_{2n}' z_i)^2 E(\tilde{v}_{ih}^2|Z) + 2|q_g'\Pi_{2n}' z_i z_i' \Pi_{2n} q_h| E(\tilde{v}_{ig}\tilde{v}_{ih}|Z) \right. \\
& \left. +2(q_g'\Pi_{2n}' z_i) E(\tilde{v}_{ig}\tilde{v}_{ih}^2|Z) + 2(q_h'\Pi_{2n}' z_i) E(\tilde{v}_{ig}^2\tilde{v}_{ih}|Z) + Var(\tilde{v}_{ig}\tilde{v}_{ih}|Z) \right\} \\
\le \;& \frac{C}{A_g A_h} \max_{1\le i\le n} \{ (q_g'\Pi_{2n}' z_i)^2 + (q_h'\Pi_{2n}' z_i)^2 + |q_g'\Pi_{2n}' z_i| + |q_h'\Pi_{2n}' z_i| + 1 \},
\end{aligned}
$$

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

$$
\begin{aligned}
& \left\| \frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_h - \frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 \{ q_g'\Pi_{2n}' z_i z_i' \Pi_{2n} q_h \sigma_j^2 + q_g' E(v_{2i} v_{2i}'|z_i) \sigma_j^2 q_h \} \right\|_{L_2,Z}^2 \\
\le \;& CK \left\{ \max_{1\le i\le n} Var(W_i|Z) \max_{1\le i\le n} Var(Y_i|Z) + \max_{1\le i\le n} Var(W_i|Z) \left(\max_{1\le i\le n} E(Y_i|Z) \right)^2 \right. \\
& \left. + \left(\max_{1\le i\le n} E(W_i|Z) \right)^2 \max_{1\le i\le n} Var(Y_i|Z) \right\} \\
\le \;& \frac{CK}{A_g A_h} \max_{1\le i\le n} \left\{ (q_g'\Pi_{2n}' z_i)^2 (q_h'\Pi_{2n}' z_i)^2 + (q_g'\Pi_{2n}' z_i)^2 + (q_h'\Pi_{2n}' z_i)^2 + |q_g'\Pi_{2n}' z_i| + |q_h'\Pi_{2n}' z_i| + 1 \right\},
\end{aligned}
$$

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

$$
\begin{aligned}
& E\left[\left\{ \frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_h - \frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 \{ q_g'\Pi_{2n}' z_i z_i' \Pi_{2n} q_h \sigma_j^2 + q_g' E(v_{2i} v_{2i}'|z_i) \sigma_j^2 q_h \} \right\}^2 \right] \\
& = E_Z \left\| \frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_h - \frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 \{ q_g'\Pi_{2n}' z_i z_i' \Pi_{2n} q_h \sigma_j^2 + q_g' E(v_{2i} v_{2i}'|z_i) \sigma_j^2 q_h \} \right\|_{L_2,Z}^2 \\
& \le \frac{CK}{A_g A_h} E_Z \left[\max_{1\le i\le n} \{ (q_g'\Pi_{2n}' z_i)^2 (q_h'\Pi_{2n}' z_i)^2 + (q_g'\Pi_{2n}' z_i)^2 + (q_h'\Pi_{2n}' z_i)^2 + |q_g'\Pi_{2n}' z_i| + |q_h'\Pi_{2n}' z_i| + 1 \} \right] \\
& = O\left(\frac{K}{A_g A_h} \right) = o(1),
\end{aligned}
$$

where the last equality follows from $\mu_g^2 = O(A_g)$ and $\mu_h^2 = O(A_h)$. Thus, the Markov inequality yields

$$
\frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 q_g' y_{2i} y_{2i}' u_j^2 q_h = \frac{1}{\sqrt{A_g A_h}} \sum_{i\ne j} P_{ij}^2 \{ q_g'\Pi_{2n}' z_i z_i' \Pi_{2n} q_h \sigma_j^2 + q_g' E(v_{2i} v_{2i}'|Z) \sigma_j^2 q_h \} + o_p(1). \tag{14}
$$

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

$$\frac{1}{\sqrt{A_g A_h}}\sum_{i\neq j} P_{ij}^2 q_g' y_{2i} y_{2j}' u_i u_j q_h = \frac{1}{\sqrt{A_g A_h}}\sum_{i\neq j} P_{ij}^2 q_g' E(v_{2i}u_i|Z)E(v_{2j}'u_j|Z)q_h + o_p(1). \tag{15}$$

Combining (14), (15) and the fact that $|M_2| \leq \frac{1}{\sqrt{A_g A_h}}\left|\sum_{i\neq j}(\hat{B}_{ij} - A_{iij} - B_{ij})\right|$, we have that $M_2 \xrightarrow{p} 0$.

Case (II). Next, we consider the case where $\frac{K}{\mu_g^2\mu_h^2} \nrightarrow 0$. It follows that $|M_1| \leq \frac{1}{\sqrt{B_g B_h}}\left|\sum_{i\neq j\neq k} q_g'(\hat{A}_{ijk} - A_{ijk})q_h\right| \xrightarrow{p} 0$ from Chao *et al.* (2012, Lemma A4) by setting "W_i, Y_j, and η_k" in their notation as $\frac{1}{\sqrt{B_g}}q_g' y_{2i}$, $\frac{1}{\sqrt{B_h}}q_h' y_{2j}$, and u_k^2, respectively. Note that $\frac{1}{\sqrt{B_g B_h}}\sum_{i\neq j\neq k} q_g' A_{ijk} q_h \xrightarrow{p} 0$ 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 $\frac{1}{\sqrt{B_g B_h}}(q_g' y_{2i} y_{2i}' q_h)$ and u_i^2, respectively (and $\frac{1}{\sqrt{B_g}}q_g' y_{2i}u_i$ and $\frac{1}{\sqrt{B_h}}q_h' y_{2i}u_i$, respectively), and it follows $|M_2| \leq \frac{1}{\sqrt{B_g B_h}}\left|\sum_{i\neq j}(\hat{B}_{ij} - A_{iij} - B_{ij})\right| \xrightarrow{p} 0$ by the same argument as in Case (I).

A.3 Proof of Theorem 2

Proof of (i)

To simplify the presentation, we consider the case of a single included endogenous regressor (i.e., $G = 1$). The case of multiple endogenous regressors is shown in a similar way using the spectral decomposition as in the proof of Theorem 1.

Pick any $b \neq \beta$. In the case of $G = 1$, the JLM statistic is written as $JLM(b) = (N_1 + N_2 + N_3)/\hat{\Psi}_n$, where

$$N_1 = \left(\sum_{i\neq j}^n P_{ij} y_{2i} y_{2j}\right)^2 (\beta - b)^2, \quad N_2 = 2\sum_{i\neq j}^n \sum_{k\neq l}^n P_{ij}P_{kl} y_{2i} u_j y_{2k} y_{2l}(\beta - b), \quad N_3 = \left(\sum_{i\neq j}^n P_{ij} y_{2i} u_j\right)^2.$$

For N_1, observe that

$$\begin{aligned}
N_1 &= \left\{\sum_{i=1}^n (1 - P_{ii})\Pi_{2n}' z_i z_i' \Pi_{2n} + \sum_{i\neq j}^n P_{ij} v_{2i} v_{2j} + 2\sum_{i=1}^n (1 - P_{ii})\Pi_{2n}' z_i v_{2i}\right\}^2 (\beta - b)^2 \\
&\geq \left\{C_1\mu_1^2 + \sum_{i\neq j}^n P_{ij} v_{2i} v_{2j} + 2\sum_{i=1}^n (1 - P_{ii})\Pi_{2n}' z_i v_{2i}\right\}^2 (\beta - b)^2 \\
&\equiv (N_{11} + N_{12} + N_{13})^2 (\beta - b)^2, \qquad (16)
\end{aligned}$$

for some $C_1 > 0$, where the inequality follows from Assumption 1 (i). For N_{12} and N_{13}, similar arguments as those in Chao *et al.* (2012, Lemma A2) yield

$$\begin{aligned} N_{12}^2 &= 2\sum_{i\neq j}^n P_{ij}^2 E[v_{2i}^2|Z]E[v_{2j}^2|Z]\{1+o_p(1)\}, \\ N_{13}^2 &= 4\sum_{i=1}^n (1-P_{ii})^2(\Pi_{2n}' z_i)^2 E[v_{2i}^2|Z]\{1+o_p(1)\}. \end{aligned}$$

Then $N_{12} = o_p(N_{11})$ follows from the facts that $\sum_{i\neq j}^n P_{ij}^2 \leq \sum_{i,j=1}^n P_{ij}^2 = \sum_{i=1}^n P_{ii} = K$ and the assumption $\frac{K}{\min\{\mu_1^4,\ldots,\mu_G^4\}} \to 0$, and $N_{13} = o_p(N_{11})$ follows from Assumption 1 (i) and (iii). Hence, by applying similar arguments to the cross terms, we have

$$N_1 \geq C_1\mu_1^4\{1+o_p(1)\}.$$

Similarly, for N_2 and N_3, we obtain

$$\begin{aligned} N_2 &= 4\left\{\sum_{i,j,k,i\neq j\neq k}^n P_{ik}P_{kj}(\Pi_{2n}' z_i)(\Pi_{2n}' z_j)E[v_{2k}u_k|Z](\beta-b) + \sum_{i\neq j}^n P_{ij}^2 E[v_{2i}^2|Z]E[v_{2j}u_j|Z](\beta-b)\right\}\{1+o_p(1)\} \\ &= o_p(N_1), \end{aligned}$$

and

$$\begin{aligned} N_3 &= \left\{\sum_{i,j,k,i\neq k,j\neq k}^n P_{ik}P_{kj}(\Pi_{2n}' z_i)(\Pi_{2n}' z_j)E[u_k^2|Z] + \sum_{i\neq j}^n P_{ij}^2 E[v_{2i}u_i|Z]E[v_{2j}u_j|Z] \right. \\ &\quad \left. + \sum_{i\neq j}^n P_{ij}^2 E[v_{2i}^2|Z]E[u_j^2|Z]\right\}\{1+o_p(1)\} \\ &= o_p(N_1). \end{aligned}$$

For the denominator, similar arguments as in the proof of Lemma 4 yield

$$
\begin{aligned}
\hat{\Psi}_n &= \sum_{i,j,k,i\neq k,j\neq k}^{n} y_{2i}P_{ik}\{u_k + y_{2k}(\beta - b)\}^2 P_{kj}y_{2j} + \sum_{i\neq j}^{n} P_{ij}^2 y_{2i}y_{2j}\{u_i + y_{2i}(\beta - b)\}\{u_j + y_{2j}(\beta - b)\} \\
&= \left(\sum_{i,j,k,i\neq k,j\neq k}^{n} y_{2i}P_{ik}y_{2k}^2 P_{kj}y_{2j} + \sum_{i\neq j}^{n} P_{ij}^2 y_{2i}^2 y_{2j}^2 \right)(\beta - b)^2 \\
&\quad +2\left(\sum_{i,j,k,i\neq k,j\neq k}^{n} y_{2i}P_{ik}u_k y_{2k}P_{kj}y_{2j} + \sum_{i\neq j}^{n} P_{ij}^2 y_{2i}y_{2j}u_i y_{2j} \right)(\beta - b) \\
&\quad + \left(\sum_{i,j,k,i\neq k,j\neq k}^{n} y_{2i}P_{ik}u_k^2 P_{kj}y_{2j} + \sum_{i\neq j}^{n} P_{ij}^2 y_{2i}y_{2j}u_i u_j \right) \\
&\leq C_3 \max\{\mu_1^2, K\}\{(\beta - b)^2 + 2(\beta - b) + 1\},
\end{aligned}
$$

for some $C_3 > 0$ by applying Chao *et al.* (2012, Lemmas A3 and A4) and Assumption 1 (iii).

By using the above results, it holds that

$$
\begin{aligned}
P\{JLM(b) < C\} &= P\left\{ \frac{N_1 + N_2 + N_3}{\hat{\Psi}_n} < C \right\} \\
&\leq P\left\{ \frac{C_1\mu_1^4(1 + o_p(1))}{C_3 \max\{\mu_1^2, K\}\{(\beta - b)^2 + 2(\beta - b) + 1\}} < C \right\} \\
&\to 0, \qquad (17)
\end{aligned}
$$

for any $C > 0$, where the convergence follows from the assumption that $K/\mu_1^4 \to 0$. Therefore, the conclusion follows.

Proof of (ii)

By using (17), there exists some $C' > 0$ such that

$$
\begin{aligned}
P\{JLM(b) < C'(\beta - b)^2\} &= P\left\{ \frac{N_1 + N_2 + N_3}{\hat{\Psi}_n} < C'(\beta - b)^2 \right\} \\
&\leq P\left\{ \frac{\mu_1^4}{\hat{\Psi}_n} < (1 - \epsilon)\frac{C'}{C_1}(\beta - b)^2 \right\} + o(1) \to 0,
\end{aligned}
$$

for all $\epsilon > 0$ small enough, where C_1 is a positive constant defined in (16), and the convergence follows from the assumption that $K/\mu_1^4 \to 0$. Therefore, the conclusion follows.

A.4 Proof of Theorem 3

We show the theorem in the same way as in Theorem 1:

$$\begin{aligned} JLM(b) &= (\hat{u}_0' P^{\#} Y_2)(\Psi_n^{\dagger})^{-1}(Y_2' P^{\#} \hat{u}_0) + o_p(1) \\ &= (u_0' P^{\dagger} Y_2)(\Psi_n^{\dagger})^{-1}(Y_2' P^{\dagger} u_0) + o_p(1) \\ &\xrightarrow{d} \chi^2_G, \end{aligned} \tag{18}$$

under $H_0 : \beta = b$, where the first equality in (18) 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

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

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 (18) follows from the relation in (5), which is shown as follows. Note that under H_0,

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

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

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

we have (I) $c'Y_2'\text{diag}(P_2)\text{diag}(P_1)u = o_p(\mu_c)$ when K/μ_c^2 is bounded, and (II) $c'Y_2'\text{diag}(P_2)\text{diag}(P_1)u = o_p(\sqrt{K})$ when $K/\mu_c^2 \to \infty$. Hence we have that $c'Y_2'\text{diag}(P_2)\text{diag}(P_1)u = o_p(c'Y_2'P^{\dagger}u)$. Therefore, the relation in (5) follows.

Finally, the convergence in (18) follows from the same argument as in Lemma 3. We note that

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

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, $S_n^{-1}Q_n'\left\{\sum_{j=1,j\neq i}^{n}\Pi_{2n}'z_j(P_{2,ji}+P_{2jj}P_{1ji})\right\}u_i$, $S_n\Lambda_n^{-1/2}\xi$, and $\sqrt{K}\Lambda_n^{-1/2}\xi$, respectively.

References

[1] Ackerberg, D. A. and P. J. Devereux (2009) Improved JIVE estimators for overidentified linear models with and without heteroskedasticity, *Review of Economics and Statistics*, 91, 351-362.

[2] Andrews, D. W. K., Cheng, X. and P. Guggenberger (2020) Generic results for establishing the asymptotic size of confidence sets and tests, *Journal of Econometrics*, 218, 496-531.

[3] 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.

[4] 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.

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

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

[7] Angrist, J. D. and A. B. Krueger (1991) Does Compulsory School Attendance Affect Schooling and Earnings?, *The Quarterly Journal of Economics*, 106, 979-1014.

[8] Bekker, P. A. (1994) Alternative approximation to the distribution of instrumental variable estimators, *Econometrica*, 62, 657-681.

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

[10] Blomquist S. and M. Dahlberg (1999) Small sample properties of LIML and jackknife IV estimators: experiments with weak instruments, *Journal of Applied Econometrics*, 14, 69-88.

[11] Chao, J. and N. R. Swanson (2005) Consistent estimation with a large number of weak instruments, *Econometrica*, 73, 1673-1692.

[12] 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.

[13] Chao, J., Hausman, J. A., Newey, W. K., Swanson, N. R. and T. Woutersen (2014) Testing overidentifying restrictions with many instruments and heteroskedasticity, *Journal of Econometrics*, 178, 15-21.

[14] Crudu, F., Mellace, G. and Z. Sándor (2021) Inference in instrumental variable models with heteroskedasticity and many instruments, *Econometric Theory*, 37, 281-310.

[15] Davidson, R. and J. G. MacKinnon (2006) The case against JIVE, *Journal of Applied Econometrics*, 21, 827-833.

[16] Dufour, J.-M. (1997) Some impossibility theorems in econometrics with applications to structural and dynamic models, *Econometrica*, 65, 1365-1387.

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

[18] Hahn, J., Hausman, J. and G. Kuersteiner (2004) Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations, *Econometrics Journal*, 7, 272-306.

[19] Han, C. and P. C. B. Phillips (2006) GMM with many moment conditions, *Econometrica*, 74, 147-192.

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

[21] Hansen, C. and D. Kozbur (2014) Instrumental variables estimation with many weak instruments using regularized JIVE,*Journal of Econometrics*, 182, 290-308.

[22] 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.

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

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

[25] Kline, P., Saggio, R., and M. Sølvsten (2020) Leave-out estimation of variance components, *Econometrica*, 88, 1859-1898.

[26] Kunitomo, N. (1980) Asymptotic expansions of distributions of estimators in a linear functional relationship and simultaneous equations, *Journal of the American Statistical Association*, 75, 693-700.

[27] Mikusheva, A. and L. Sun (2021) Inference with many weak instruments, forthcoming in *Review of Economic Studies*.

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

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

[30] Moreira, H. and M. J. Moreira (2019) Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors, *Journal of Econometrics*, 213, 398-433.

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

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

[33] Phillips G. D. A. and C. Hale (1977) The bias of instrumental variable estimators of simultaneous equation systems, *International Economic Review*, 18, 219-228.

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