CONDITIONAL LIKELIHOOD RATIO TEST WITH MANY WEAK INSTRUMENTS

SREEVIDYA AYYAR, YUKITOSHI MATSUSHITA, AND TAISUKE OTSU

Abstract. This paper extends validity of the conditional likelihood ratio (CLR) test developed by Moreira (2003) to instrumental variable regression models with unknown error variance and many weak instruments. In this setting, we argue that the conventional CLR test with estimated error variance loses exact similarity and is asymptotically invalid. We propose a modified critical value function for the likelihood ratio (LR) statistic with estimated error variance, and prove that this modified test achieves asymptotic validity under many weak instrument asymptotics. Our critical value function is constructed by representing the LR using four statistics, instead of two as in Moreira (2003). A simulation study illustrates the desirable properties of our test.

1. Introduction

Inference in regression models with endogenous structural variables and many, weakly relevant instrumental variables is of great importance in applied research. A canonical example is the paper by Angrist and Krueger (1991), in which the authors estimate the effect of educational attainment on wages by constructing up to 1,530 instruments for education through interacting quarter, state, and year of birth. A more recent literature that uses many weak instruments employs the "judge design" empirical strategy, which exploits random assignments of cases to judges (e.g., Dahl, Kostøl, and Mogstad 2014; Autor et al. 2019). Since each judge is an instrument in these settings, and judges can only process a certain number of cases, the number of instruments typically increases with the sample size. Many weak instrument settings are also found in papers estimating the new Keynesian Phillips curves (see Mavroeidis, Plagborg-Møller, and Stock 2014) or studies that employ the Fama-MacBeth method for pricing assets (Fama and MacBeth 1973).

In applications involving many weak instruments, researchers often rely on standard asymptotic approximations when conducting inference. However, asymptotic approximations to the finite sample distributions of conventional estimators and test statistics have been shown to be poor when instruments are weak. The use of many instruments can improve efficiency of estimators or their associated tests, but often causes the usual inference procedures to have poor finite sample properties, and several previous papers have noted this issue. Chao and Swanson (2005), Han and Phillips (2006), and Newey and Windmeijer (2009) generalize the many-instrument asymptotic theory to allow for weak instruments or moments. Andrews and Stock (2007) show that the Anderson-Rubin (AR), Lagrange multiplier (LM), and conditional likelihood ratio (CLR) tests are robust to many weak instruments, as long as the number of instruments k grows slower than the cube root of the sample size, $n^{1/3}$. For the case where k may be proportional to n, Hansen, Hausman, and Newey (2008) develop a many-instrument

We are grateful to Naoto Kunitomo for helpful comments.

robust standard error and a modification for the LM test, while Hausman et al. (2012) propose Wald tests with limited information maximum likelihood and Fuller estimators that are robust to heteroskedasticity and many instruments. More recent developments in conducting robust inference with many weak instruments include the jackknife AR tests by Crudu, Mellace, and Sándor (2021) and Mikusheva and Sun (2021), and the jackknife LM test by Matsushita and Otsu (2022).

For the weak (but fixed number of) instruments problem, the seminal work of Moreira (2003) sparked a growing literature on conditional inference. Moreira (2003) introduces a general conditional inference framework for instrumental variable regression models with homoskedastic errors and advocates the CLR test. Andrews, Moreira, and Stock (2006) establish a nearly optimal property of the CLR test, while Mills, Moreira, and Vilela (2014) propose approximately unbiased conditional Wald tests with comparable power to the CLR test. Moreira and Moreira (2019) extend the conditional inference framework to heteroskedastic and autocorrelated errors.

In this paper, we set out to investigate the finite sample performance of Moreira's (2003) CLR approach when k is relatively large, and may grow proportionally with n. Size robustness of the CLR test under $k = o(n^{1/3})$ has already been established by Andrews and Stock (2007). However, we show that in a setting with homoskedastic normal errors with unknown variance, the conventional CLR test loses exact similarity and is asymptotically invalid under many weak instrument asymptotics, where k may grow much faster than $n^{1/3}$. We propose a modified version of Moreira's (2003) CLR test, hereafter called the modified CLR (MCLR) test, which is robust to: (i) many instruments, where the number of instruments can grow at the same rate or slower than the sample size and (ii) weak instruments, to the extent that a consistent test may exist (Mikusheva and Sun 2021). It should be noted that we use the same test statistic as Moreira (2003) ("LR_1" in his paper), but our proposed test employs a different critical value function which is constructed by representing the likelihood ratio using four statistics, instead of two as in Moreira (2003). Our MCLR test retains asymptotic validity when there are many weak instruments, under the least restrictive condition possible on identification strength (Mikusheva and Sun 2021). This result holds even when we relax the assumption of normally distributed error terms, as long as we impose an additional moment condition.

The rest of this article is organized as follows. Section 2 introduces our setup and the likelihood ratio statistic (LR) when error variance is unknown, and discusses its representation by four statistics as well as the properties of those statistics. In Section 3, we propose our MCLR test by constructing a robust critical value function and establish its asymptotic validity in a many weak instruments setting. We also discuss the lack of validity of the conventional CLR critical value function in our setup. Section 4 illustrates the usefulness of our proposed method by a simulation study. All proofs are contained in Appendix A.

2. Setup and test statistics

2.1. **Setup.** Consider the following instrumental variable regression model:

$$\begin{aligned} y_1 &= Y_2\beta + u, \\ Y_2 &= Z\Pi_2 + V_2, \end{aligned} \tag{1}$$

where $y_1 = (y_{1i}, \dots, y_{1n})'$ is an $n \times 1$ vector of dependent variables, Y_2 is an $n \times l$ matrix of endogenous regressors, β is an $l \times 1$ vector of unknown structural parameters, u is an $n \times 1$ vector of mean-zero disturbances, Z is an $n \times k$ matrix of instruments, Π_2 is a $k \times l$ matrix of unknown parameters, and V_2 is an $n \times l$ matrix of mean-zero error terms. We assume without loss of generality that there are no exogenous regressors in (1) since one can always partial them out using standard projection methods. The reduced form system can be written as

$$Y = Z\Pi + V, \tag{2}$$

where $Y = (y_1, Y_2)$, $\Pi = (\pi_1, \Pi_2)$, and $V = (v_1, V_2)$ with $\pi_1 = \Pi_2\beta$ and $v_1 = V_2\beta + u$.

This paper is concerned with testing the null hypothesis $H_0 : \beta = \beta_0$ on the structural parameters, against the alternative $H_1 : \beta \neq \beta_0$, where the coefficients Π_2 are treated as nuisance parameters. In particular, this paper focuses on the situation wherein researchers wish to test such a hypothesis, but only have many weak instruments at their disposal.

To proceed, we impose the following assumptions.

Assumption.

1: *[Normal errors] The rows of V are independent and identically distributed, and follow $N(0, \Omega)$ with a positive definite matrix Ω. Ω is unknown to the researcher.*

2.: *[Many weak instruments] Z is non-random. The number of instruments $k = k_n$ may grow as n increases. One of the following two conditions hold*

(a) $\frac{k}{n} \to \alpha \in [0, 1)$ as $n \to \infty$, and the concentration parameter

$$\mu^2 = (A_0'\Omega^{-1}A_0)^{-1/2}A_0'\Omega^{-1}\Pi'Z'Z\Pi\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2},$$

with $A_0 = (\beta_0, I_l)'$ satisfies $\frac{\mu^2}{\sqrt{k}} \to \infty$ as $n \to \infty$; or

(b) $\frac{k}{n} \to 0$ as $n \to \infty$ (without any condition on μ^2).

The normality of reduced-form errors in Assumption 1 is useful to motivate our conditional inference approach, which is inspired by the exact similarity of the LR statistic with known Ω. Indeed, Moreira (2003) proves that conditional on the sufficient statistic Π_2 and when errors are normally distributed, the LR statistic with known Ω has a finite-sample distribution independent of nuisance parameters under H_0 and its quantiles can be used to construct a similar test (as long as this distribution is continuous). Since we maintain Moreira (2003)'s conditional inference framework, we begin with normally distributed error terms, although we will show that this assumption can be relaxed for the asymptotic analysis (see Theorem 3). Throughout this paper, we focus on the case where Ω is unknown to researchers. Assumption 2 concerns the instrumental variables. In this paper, we restrict Z to be non-random, which is equivalent to conditioning on Z. In order to allow for k to grow proportionally with n as in Assumption 2 (a), we need to impose an additional condition $\frac{\mu^2}{\sqrt{k}} \to \infty$, which imposes a lower bound on the strength of the instruments. Indeed, $\frac{\mu^2}{\sqrt{k}} \to \infty$ is the least restrictive characterization of weak identification that allows for a consistent test of $H_0 : \beta = \beta_0$ (Mikusheva and Sun 2021). Chao and Swanson (2005) also impose this condition to achieve consistency of point estimators under many weak instrument asymptotics. If k grows slower than n, as in Assumption 2 (b), there is no requirement on μ^2 , i.e., the instruments can be arbitrarily weak.

Note that Wald tests based on many-instrument robust standard errors (Hansen, Hausman, and Newey 2008; Hausman et al. 2012) are asymptotically valid under Assumption 2 (a), but not under Assumption 2 (b). Our MCLR test is asymptotically valid in both cases. Simulation studies in Section 4 illustrate this distinction numerically. Andrews and Stock (2007) show that the conventional CLR test is asymptotically valid under relatively small numbers of instruments, that is when $k^3/n \to 0$. Assumption 2 allows the number of instruments k to be much larger, and as illustrated in our simulation study, the MCLR test is preferred when k/n is large.

2.2. **Likelihood ratio statistic with known Ω.** We first introduce some notation. When the variance Ω of V is known, the LR statistic for testing H_0 against H_1 is written as

$$LR_0 = \frac{b_0' Y' P_Z Y b_0}{b_0' \Omega b_0} - \bar{\lambda}, \tag{3}$$

where $b_0 = (1, -\beta_0')'$, $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix with respect to Z, and $\bar{\lambda}$ is the smallest eigenvalue of $\Omega^{-1/2}Y'P_Z Y\Omega^{-1/2}$ (Moreira 2003).

To derive a more convenient expression for LR_0, note that $Z'Y$ is a sufficient statistic for the parameters (β, Π) under the assumption $V \sim N(0, \Omega)$ with known Ω. This implies that $Z'YD$ is also a sufficient statistic, for any nonsingular matrix D. So, we set $D = (b_0, \Omega^{-1}A_0)$ and obtain the partition $Z'YD = [S : T]$, where

$$S = Z'Yb_0, \qquad T = Z'Y\Omega^{-1}A_0.$$

This is a convenient partitioning because S and T are independent and only T depends on the nuisance parameters Π. Indeed, under the null hypothesis, T alone is a sufficient statistic for Π.

By using standardized versions of S and T:

$$\bar{S} = (Z'Z)^{-1/2}Z'Yb_0(b_0'\Omega b_0)^{-1/2}, \qquad \bar{T} = (Z'Z)^{-1/2}Z'Y\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2},$$

the LR statistic LR_0 can be alternatively written as

$$LR_0 = \bar{S}'\bar{S} - \bar{\lambda} \equiv \psi_0(\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T}), \tag{4}$$

where $\bar{\lambda}$ is the smallest eigenvalue of $(\bar{S}, \bar{T})'(\bar{S}, \bar{T})$. See Proposition 1 in Moreira (2003) for a proof. If Ω is known, we can apply the conventional CLR test by Moreira (2003) based on LR_0, even with many weak instruments. Since this paper focuses on the case of unknown Ω as stated in Assumption 1, the conventional test is infeasible. Its feasible counterpart, obtained by plugging in a consistent estimator of Ω, turns out to be invalid under the many weak instrument asymptotics (see, Remark 2 below).

2.3. **Likelihood ratio statistic with unknown Ω.** We now introduce our test statistic of interest, for the case of unknown Ω. The error variance matrix Ω can be estimated by

$$\hat{\Omega} = \frac{1}{n-k} Y' M_Z Y, \tag{5}$$

where $M_Z = I_n - P_Z$. By replacing Ω in (3) with the estimator $\hat{\Omega}$, the LR statistic for testing H_0 with unknown Ω is written as

$$\frac{LR_1}{n-k} = \frac{b_0' Y' P_Z Y b_0}{b_0' Y' M_Z Y b_0} - \hat{\lambda}, \tag{6}$$

where $\hat{\lambda}$ is the smallest eigenvalue of $\frac{1}{n-k} \hat{\Omega}^{-1/2} Y' P_Z Y \hat{\Omega}^{-1/2}$.

To obtain an analogous expression to (4) for LR_1, we introduce two more objects:

$$\tilde{S} = M_Z Y b_0 (b_0' \Omega b_0)^{-1/2}, \qquad \tilde{T} = M_Z Y \Omega^{-1} A_0 (A_0' \Omega^{-1} A_0)^{-1/2}.$$

Based on this notation, we obtain the following representation of the LR_1 statistic.

Proposition 1. *LR_1 can be written as a function of n, k, and $(\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T}, \tilde{S}'\tilde{S}, \tilde{S}'\tilde{T}, \tilde{T}'\tilde{T})$:*

$$\frac{LR_1}{n-k} = \psi_{1,n,k}(\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T}, \tilde{S}'\tilde{S}, \tilde{S}'\tilde{T}, \tilde{T}'\tilde{T}).$$

This proposition says that the LR statistic LR_1 depends on six objects, instead of three as for $LR_0 = \psi_0(\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T})$ in (3). In order to develop our conditional inference method based on LR_1, we first establish the following properties of those six objects.

Proposition 2. *Under Assumption 1 and the null hypothesis $H_0 : \beta = \beta_0$, it holds*

(i): $\bar{S} | \bar{T} = t \sim N(0, I_k)$ *and* $\bar{S}'\bar{T} | \bar{T} = t \sim N(0, t't)$,

(ii): $\left. \begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T} \end{pmatrix} \right| \bar{T} = t \sim Wishart(n-k, I_{l+1}, 0)$,

(iii): $\bar{S}$, $\bar{T}$, *and* $(\tilde{S}, \tilde{T})$ *are mutually independent.*

Remark 1. Moreira (2003) builds a conditional inference framework for the conventional CLR test based on two sufficient statistics, $\bar{S}$ and $\bar{T}$. We add two more statistics, $\tilde{S}$ and $\tilde{T}$, which are shown to be mutually independent of $\bar{S}$ and $\bar{T}$. We need to formally establish the properties of $\tilde{S}$ and $\tilde{T}$ because we explicitly focus on the case of unknown Ω, as stated in Assumption 1. On the other hand, Moreira (2003) defines the conventional CLR test using LR_0 and later establishes that using a plug-in consistent estimator for Ω is asymptotically valid under weak (but a fixed number of) instruments. Since this will not be the case under Assumption 2, we directly consider LR_1. Under many weak instrument asymptotics, the dimensions of all four of our statistics $\bar{S}$, $\bar{T}$, $\tilde{S}$ and $\tilde{T}$ grow to ∞, which explains why the six objects we focus on are expressed by inner-products. As we will see in Section 3, $\bar{T}$ will play the most important role in our conditional inference approach, since it is a sufficient statistic for Π. Moreover, $\bar{T}'\bar{T}$ is centered at the concentration parameter μ^2, and therefore is a measure of how strongly identified the first-stage is. We will use this fact in Section 4.

3. Conditional likelihood ratio test with many weak instruments

Based on the test statistic LR_1 and its properties presented in the last section, we now develop our conditional inference method. To begin with, recall that $\bar{T}$ is a sufficient statistic for Π, and consider the critical value function for given $\bar{T} = t$:

$$c_{1,\alpha}(t) = (1-\alpha)\text{-th quantile of } \psi_{1,n,k}(\mathcal{S}'\mathcal{S}, \mathcal{S}'t, t't, \mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_3),$$

where $\psi_{1,n,k}$ is defined in Proposition 1, and $\mathcal{S} \sim N(0, I_k)$ and $\begin{pmatrix} \mathcal{W}_1 & \mathcal{W}_2 \\ \mathcal{W}_2 & \mathcal{W}_3 \end{pmatrix} \sim Wishart(n-k, I_{l+1}, 0)$ are independent. Propositions 1 and 2 directly imply the following property of $c_{1,\alpha}(t)$.

Theorem 1. *Under Assumption 1 and the null hypothesis $H_0 : \beta = \beta_0$, it holds that*

$$\Pr\left\{\frac{LR_1}{n-k} \geq c_{1,\alpha}(\bar{T})\right\} = 1 - \alpha. \tag{7}$$

This theorem says that if $\bar{T}$ is observable, the LR test using $\frac{LR_1}{n-k}$ with the critical value $c_{1,\alpha}(\bar{T})$ is exactly similar. However, since $\bar{T}$ is unobservable for the case of unknown Ω, the test based on (7) is infeasible. To develop a feasible version, we estimate $\bar{T}$ by

$$\hat{T} = (Z'Z)^{-1/2} Z'Y\hat{\Omega}^{-1}A_0(A_0'\hat{\Omega}^{-1}A_0)^{-1/2},$$

where $\hat{\Omega}$ is as defined in (5). Based on this estimator, our proposed rejection rule is defined as:

$$\text{Reject } H_0 \text{ if } \frac{LR_1}{n-k} \geq c_{1,\alpha}(\hat{T}). \tag{8}$$

The next theorem is the main result of our paper, and it establishes asymptotic validity of the MCLR test in (8).

Theorem 2. *Consider the setup in Section 2.1. Under Assumptions 1 and 2, it holds that*

$$\Pr\left\{\frac{LR_1}{n-k} \geq c_{1,\alpha}(\hat{T})\right\} \to 1 - \alpha \quad \textit{as } n \to \infty. \tag{9}$$

This theorem is derived under the normality assumption (Assumption 1). For non-normal errors, as long as $k/n \to 0$, we can also establish asymptotic validity of the MCLR test by requiring an additional moment condition. Let P_{ii} be the (i,i)-th element of P_Z.

Theorem 3. *Consider the setup in Section 2.1. Under Assumption 2 (b) and $\frac{1}{k}\sum_{i=1}^n P_{ii}^2 \to 0$, (9) holds true.*

Specifically, as long as the number of instruments k grows slower than the sample size n, and $\frac{1}{k}\sum_{i=1}^n P_{ii}^2 \to 0$, our MCLR test is asymptotically valid even when reduced form errors are non-normal. Note, Assumption 2 (b) is still more general than $k = o(n^{1/3})$, which Andrews and Stock (2007) require for asymptotic validity of the CLR test with non-normal errors.

Remark 2. [Lack of similarity and validity of conventional CLR test] When Ω is known, the critical value function of the test statistic LR_0 in (4) for testing $H_0 : \beta = \beta_0$ can be obtained as

$$c_{0,\alpha}(t) = (1-\alpha)\text{-th quantile of } \psi_0(\mathcal{D}_1, \mathcal{D}_2, t't),$$

where $\mathcal{D}_1 \sim \chi^2(k-l)$ and $\mathcal{D}_2 \sim N(0, t't)$ are independent. As shown by Moreira (2003), the test $\mathbb{I}\{LR_0 \geq c_{0,\alpha}(\bar{T})\}$ is exactly similar for the case of known Ω (i.e., $\Pr\{LR_0 \geq c_{0,\alpha}(\bar{T})\} = 1-\alpha$). When Ω is unknown, Moreira (2003) suggested to plug-in the estimator $\hat{\Omega}$ to the test statistic LR_0 (which yields LR_1) and use $c_{0,\alpha}(\hat{T})$, that is:

$$\text{Reject } H_0 \text{ if } LR_1 \geq c_{0,\alpha}(\hat{T}). \tag{10}$$

However, since LR_1 is evidently different from LR_0, we cannot guarantee similarity for LR_1, i.e.,

$$\Pr\{LR_1 \geq c_{0,\alpha}(\bar{T})\} \neq \Pr\{LR_0 \geq c_{0,\alpha}(\bar{T})\} = 1-\alpha.$$

Therefore, even if we ignore the estimation error arising from using $\hat{T}$ instead of $\bar{T}$, the conventional CLR test in (10) is asymptotically invalid in our setup.

4. Numerical illustrations

In this section, we compare the critical value function $c_{1,\alpha}(t)$ of the MCLR test with $c_{0,\alpha}(t)$ of the conventional CLR test (Section 4.1), and use Monte Carlo simulations to evaluate the finite sample performance of our MCLR test relative to existing competitors.

4.1. **Critical value function.** The critical value function $c_{1,\alpha}(t)$ of our MCLR test does not have a closed form, as is the case with Moreira's (2003) critical value function $c_{0,\alpha}(t)$. Panel A of Table 1 presents the critical value function of the MCLR test for the 5% significance level. Critical values are shown for $n = 100$, using different values of $\tau = \bar{T}'\bar{T}$ and calculated using 10,000 Monte Carlo replications. Although $c_{1,\alpha}(t)$ takes $\bar{T}$ as input, we choose to focus on τ for simulations, which is indicative of identification strength. This also aids comparison with Moreira (2003), whose critical value function only depends on τ.

When $k = 1$, the critical value function of the MCLR test is a constant equal to 3.93 for all values of τ, with the slight variation in the final row being due to simulation error. Just as is the case for the CLR test, the critical value function of the MCLR test for any given k has an approximately exponential shape. Figure 4.1 illustrates this with a plot of the critical value function of our MCLR test when $k = 4$. When instruments are weak (i.e., τ is small), the critical values are larger. When τ is larger, such that instruments are stronger, the test behaves as if it were unconditional and the critical values are relatively stable around 3.93.

For comparison, in Panel B of Table 1, we present the critical value function $c_{0,\alpha}(t)$ of the conventional CLR test. As shown in Theorem 2, this test runs into size problems when there are many weak instruments. This is evident in the critical value function - once the number of instruments exceeds a tenth of the sample size, the critical values of the CLR test lie everywhere below those of our MCLR test. This suggests that the conventional CLR test would lead to over-rejection of the null hypothesis $H_0 : \beta = \beta_0$ when the number of instruments is large.

TABLE 1. Critical value functions

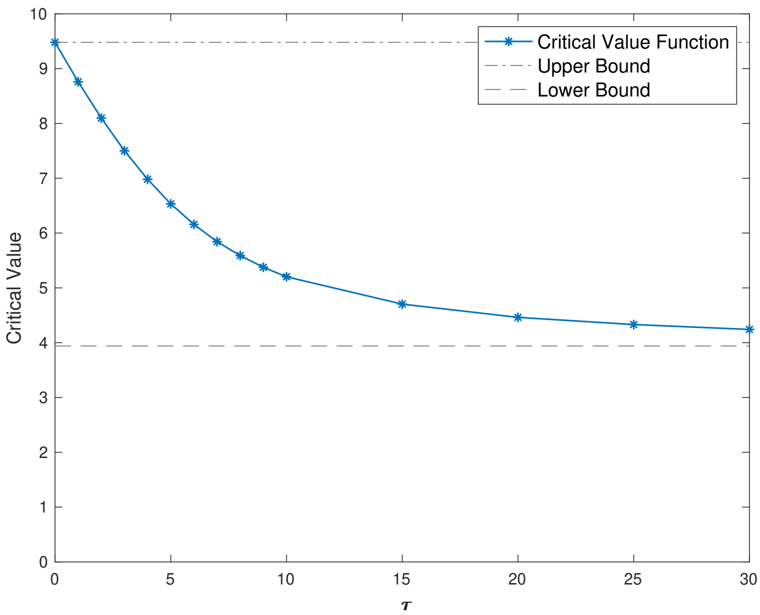
Panel A: MCLR

τ	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=10$	$k=20$	$k=50$
1	3.93	5.72	7.46	9.13	10.75	18.45	33.09	78.94
5	3.93	4.72	5.71	6.86	8.12	15.02	29.30	74.91
10	3.93	4.34	4.85	5.46	6.19	11.40	24.79	70.00
20	3.93	4.14	4.37	4.63	4.93	7.20	16.87	60.48
50	3.93	4.02	4.11	4.20	4.30	4.91	7.02	35.25
75	3.93	3.99	4.05	4.11	4.18	4.55	5.66	20.18
100	3.93	3.98	4.02	4.06	4.10	4.38	5.14	12.84
50000	3.94	3.94	3.94	3.94	4.10	3.94	3.96	4.04

Panel B: CLR

τ	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=10$	$k=20$	$k=50$
1	3.84	5.54	7.18	8.76	10.29	17.41	30.46	66.51
5	3.84	4.57	5.48	6.53	7.68	14.00	26.70	62.59
10	3.84	4.22	4.67	5.20	5.85	10.40	22.17	57.73
20	3.84	4.02	4.23	4.46	4.71	6.51	14.18	48.10
50	3.84	3.91	3.99	4.08	4.16	4.65	6.05	21.62
75	3.84	3.89	3.94	4.00	4.05	4.35	5.10	10.27
100	3.84	3.88	3.92	3.95	3.99	4.21	4.72	7.35
50000	3.84	3.84	3.84	3.84	3.99	3.84	3.84	3.84

FIGURE 1. Critical value function of MCLR test with $k=4$



4.2. **Simulation.** This subsection conducts a simulation study based on Design I of Staiger and Stock (1997) with $l=1$, and $\beta_0=0$. Z comprises of: a constant Z_1, Z_2 drawn from $N(0,1)$, Z_2^2, Z_2^3 and $k-4$ elements independently drawn from $N(0,1)$. Once drawn, Z is then held fixed across simulations. To vary the strength of instruments, we use a population version of Stock and Yogo's (2005) pre-test for weak instruments. We use three different values of Π_2 so that the population first-stage F-statistic $\delta^2 = \Pi_2' Z' Z \Pi_2 / \omega_{22}$ takes the values 2 (very weak instruments), 10 (weak instruments), and 30 (strong instruments), for different values of k. The rows of (u, V_2) are i.i.d. normal random vectors with unit variances and correlation ρ, which measures the degree of endogeneity of Y_2 in (1). The number of Monte Carlo replications is 10,000 for the size analysis and 2,500 for the power analysis.

Table 4.2 investigates the size properties of five tests for $n = 100$ and $H_0 : \beta = \beta_0$: (i) the t-test with the heteroskedasticity robust limited information maximum likelihood estimator by Hausman et al. (2012) (H-LIML), (ii) the conditional likelihood ratio test by Moreira (2003) (CLR), (iii) the modified Lagrange multiplier test by Hansen, Hausman, and Newey (2008) (mKLM), (iv) the jackknifed version of the Anderson–Rubin (AR) test proposed by Mikusheva and Sun (2021) (J-AR) and our proposed modified CLR test (MCLR). Firstly, note that the size distortions of H-LIML are large, except when δ^2 is large relative to k. The degree of endogeneity of Y_2 also seems to matter; when $\rho = 0.2$, the t-test tends to under-reject the null hypothesis, while when $\rho = 0.6$, the null is over-rejected. The distortions of the test are most severe when δ^2 is small relative to k, and k is large. As a result, we do not investigate power of this test.

The CLR test attains roughly the correct size when $k = 5$, even when identification is weak and the degree of endogeneity is high. However, size distortions can be observed when $k/n \geq 0.1$. Surprisingly, this is not visibly exacerbated when δ^2 is low, suggesting that it is the existence of many instruments that has the more severe empirical consequences on the conventional CLR test. Overall, even when the CLR test experiences little size distortion, it always has empirical rejection frequency farther from 5% than our proposed MCLR test.

Table 2. Empirical rejection frequencies at 5% significance level

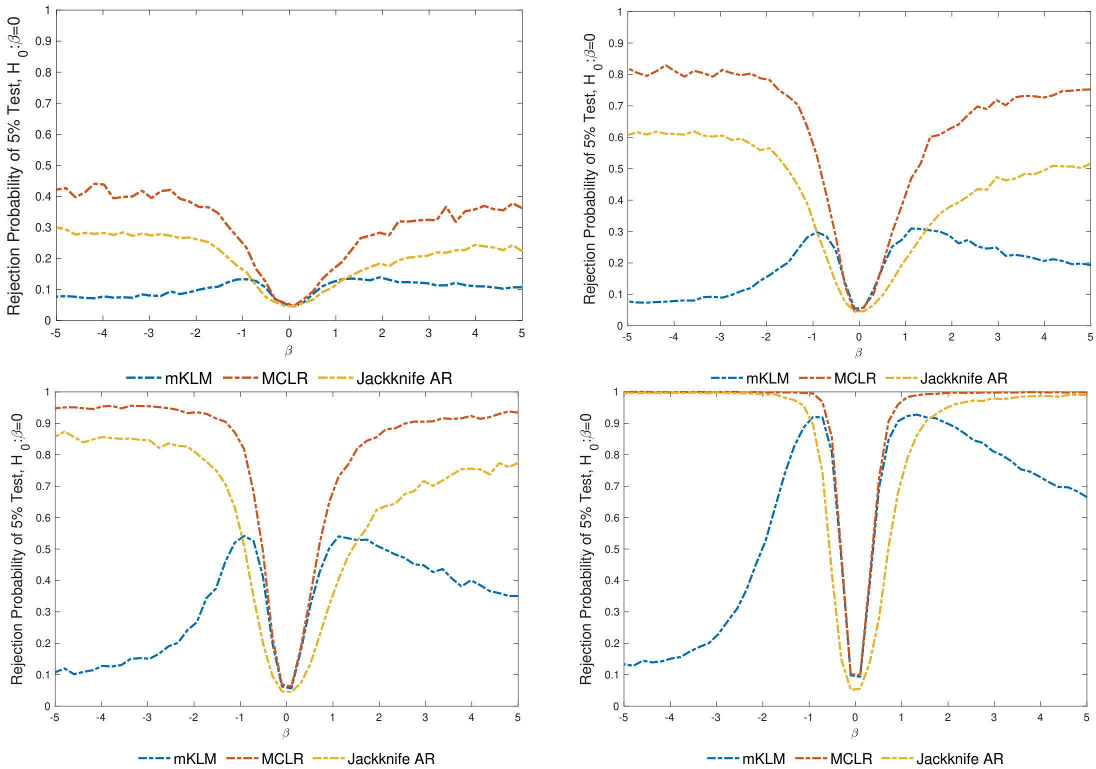
ρ	δ^2	k	H-LIML	CLR	mKLM	J-AR	M-CLR
0.2	30	5	0.038	0.055	0.051	0.063	0.048
0.2	30	10	0.030	0.064	0.050	0.064	0.054
0.2	30	30	0.028	0.092	0.048	0.064	0.055
0.2	10	5	0.022	0.046	0.046	0.060	0.042
0.2	10	10	0.018	0.059	0.045	0.056	0.047
0.2	10	30	0.016	0.093	0.044	0.060	0.048
0.2	2	5	0.008	0.048	0.041	0.062	0.042
0.2	2	10	0.007	0.057	0.039	0.052	0.039
0.2	2	30	0.013	0.087	0.046	0.058	0.050
0.6	30	5	0.058	0.053	0.055	0.062	0.048
0.6	30	10	0.058	0.062	0.053	0.061	0.052
0.6	30	30	0.052	0.076	0.048	0.056	0.048
0.6	10	5	0.070	0.049	0.047	0.062	0.045
0.6	10	10	0.069	0.062	0.050	0.056	0.056
0.6	10	30	0.086	0.092	0.047	0.057	0.048
0.6	2	5	0.084	0.050	0.042	0.065	0.046
0.6	2	10	0.092	0.058	0.040	0.059	0.037
0.6	2	30	0.106	0.088	0.045	0.060	0.057

The mKLM test works well. Although it tends to under-reject when δ^2 is relatively small, its size distortions never exceed 1%. Similarly, the J-AR test appears relatively robust to many weak instruments, although it tends to over-reject the null in all cases.

Compared to the other tests that we consider, the rejection frequencies of our proposed MCLR test are on average closest to the nominal level. As our theory in Section 3 suggests, the MCLR test is robust to weak instruments, many instruments, and many weak instruments.

Figure 4.2 presents calibrated power curves for the MCLR, J-AR and mKLM tests for $H_0 : \beta = \beta_0$, under the alternative hypotheses $H_1 : \beta = \beta_0 + \Delta$. These power curves are plotted with respect to the 5% significance level, i.e., the critical values for these three tests are given by the 95-th percentiles of these test statistics under H_0, computed via 10,000 Monte Carlo replications. Each curve is plotted for $n = 200$, $l = 1$, and $\beta_0 = 0$. We present four different cases, with different values of δ^2/k , holding $\rho = 0.2$. As we move from left to right, and top to bottom, the figures show the cases of $\delta^2/k = 1/3$, $1/2$, 1, and 2. Our MCLR test is uniformly more powerful at all values of δ^2/k, although is more pronounced for low δ^2/k. The mKLM test experiences spurious declines in power under alternative hypotheses that are further away from the null, with consistently low power when $\delta^2/k = 1/3$. When δ^2/k is high, i.e., identification is strong and/or the number of instruments is small, the J-AR has similar, although everywhere lower power, than our M-CLR test. While we do not present theoretical results on power, this result suggests that the MCLR test shares the superior power properties of the conventional CLR test, which has near optimal power with small k (Andrews, Moreira, and Stock 2006).

FIGURE 2. Calibrated Power Curves



Note: From left to right, and top to bottom, these figures plot power curves for: $\delta^2/k = 1/3$, $\delta^2/k = 1/2$, $\delta^2/k = 1$ and $\delta^2/k = 2$, with $\rho = 0.2$ and $n = 200$.

5. Conclusion

In this paper, we propose a modification of Moreira's (2003) conditional likelihood ratio (CLR) test, namely the MCLR test. We prove that in instrumental variable regression models with unknown error variance and many weak instruments, the MCLR test is asymptotically valid under many weak instrument asymptotics, unlike the CLR test. This is true even when the number of instruments grows proportionally to the sample size, and identification is as weak as possible to the extent that a consistent test may exist (Mikusheva and Sun 2021).

Our simulation study indicates that the MCLR test has superior size properties to the CLR test and is more powerful than competing tests that are robust to many weak instruments, including the modified Lagrange multiplier test by Hansen, Hausman, and Newey (2008) and the jackknife Anderson–Rubin test by Mikusheva and Sun (2021).

The size and power results presented in our theorems and simulation study lead us to recommend the MCLR test for general use in scenarios when instruments are many and potentially weak. This is based on the fact that the MCLR test is asymptotically valid with little size distortion in finite samples with many weak instruments, while also retaining the favorable power properties of the CLR test.

Appendix A. Mathematical appendix

A.1. **Proof of Proposition 1.** Let $(D_1, \ldots, D_6) = (\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T}, \tilde{S}'\tilde{S}, \tilde{S}'\tilde{T}, \tilde{T}'\tilde{T})$. Recall that $\frac{LR_1}{n-k} = \frac{b_0'Y'P_ZYb_0}{b_0'Y'M_ZYb_0} - \hat{\lambda}$, where $\hat{\lambda}$ is the smallest eigenvalue of $(n-k)^{-1}\hat{\Omega}^{-1/2}Y'P_ZY\hat{\Omega}^{-1/2}$. The numerator of the first term can be written as

$$b_0'Y'P_ZYb_0 = (b_0'\Omega b_0)\bar{S}'\bar{S} = (b_0'\Omega b_0)D_1,$$

where the first equality follows from the definition of $\bar{S}$. Similarly, the denominator of the first term of $\frac{LR_1}{n-k}$ can be written as

$$b_0'Y'M_ZYb_0 = (b_0'\Omega b_0)\tilde{S}'\tilde{S} = (b_0'\Omega b_0)D_4,$$

where the first equality follows from the definition of $\tilde{S}$. Thus the first term of $\frac{LR_1}{n-k}$ is written as $\frac{D_1}{D_4}$.

We now consider the second term of $\frac{LR_1}{n-k}$. Observe that $\hat{\lambda}$ is the minimum eigenvalue solution of $|\hat{\Omega}^{-1/2}Y'P_ZY\hat{\Omega}^{-1/2} - \hat{\lambda}I| = 0$, or equivalently

$$\left| F'Y'P_ZYF - \frac{\hat{\lambda}}{n-k}F'Y'M_ZYF \right| = 0,$$

for any nonsingular matrix F. By setting $F = [b_0(b_0'\Omega b_0)^{-1/2} : \Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}]$, the above equation can be written as

$$0 = \left| \begin{pmatrix} \bar{S}'\bar{S} & \bar{S}'\bar{T} \\ \bar{T}'\bar{S} & \bar{T}'\bar{T} \end{pmatrix} - \frac{\hat{\lambda}}{n-k} \begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T} \end{pmatrix} \right| = \left| \begin{pmatrix} D_1 & D_2 \\ D_2' & D_3 \end{pmatrix} - \frac{\hat{\lambda}}{n-k} \begin{pmatrix} D_4 & D_5 \\ D_5' & D_6 \end{pmatrix} \right|.$$

Therefore, $\hat{\lambda}$ can be solved for as a function of $(D_1, \ldots, D_6)$. Combining these results, we obtain the conclusion.

A.2. **Proof of Proposition 2.**

A.2.1. *Proof of (i).* As shown in Moreira (2003), $\bar{S} \sim N(0, I_K)$ and $\bar{S}$ and $\bar{T}$ are independent. Therefore, the conclusion follows.

A.2.2. *Proof of (ii).* Observe that

$$\begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T} \end{pmatrix} = W'M_ZW,$$

where $W = [Vb_0(b_0'\Omega b_0)^{-1/2} : V\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}]$. Since M_Z is an $n \times n$ non-random idempotent matrix with $\text{rank}(M_Z) = n-k$, it is sufficient for the conclusion to show that given $\bar{T} = t$, the rows of W are i.i.d. $N(0, I_{l+1})$. The i-th row of W can be written as

$$W_i' = [V_i'b_0(b_0'\Omega b_0)^{-1/2} : V_i'\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}], \tag{11}$$

where the i-th row V_i' of V satisfies $V_i \sim N(0, \Omega)$. Thus, we can see that $Var(V_i'b_0(b_0'\Omega b_0)^{-1/2}) = 1$, $Var(V_i'\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}) = I_l$, and $Cov(V_i'b_0(b_0'\Omega b_0)^{-1/2}, V_i'\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}) = 0$, and the conclusion follows.

A.2.3. *Proof of (iii).* Since $\bar{S}$ and $\bar{T}$ are independent, it is sufficient to show that $(\bar{S}, \bar{T})$ and $(\tilde{S}, \tilde{T})$ are independent. Note that $[\tilde{S} : \tilde{T}] = M_Z W$ whose i-th row W_i is defined in (11). Since

$$\bar{S} = (b_0'\Omega b_0)^{-1/2})^{-1/2}(Z'Z)^{-1/2}Z'(v_1 - V_2\beta_0),$$

we have $Cov(\bar{S}, [\tilde{S} : \tilde{T}]) = 0$. Also since both $\bar{S}$ and $[\tilde{S} : \tilde{T}]$ are normal, we obtain independence of $(\bar{S}, \bar{T})$ and $(\tilde{S}, \tilde{T})$.

A.3. **Proof of Theorem 2.** To simplify the presentation, we provide the proof for the case of $l = 1$. Analogous arguments hold for $l > 1$. Let

$$\mathcal{S} \sim N(0, I_k), \qquad \begin{pmatrix} \mathcal{W}_1 & \mathcal{W}_2 \\ \mathcal{W}_2 & \mathcal{W}_3 \end{pmatrix} \sim Wishart(n-k, I_2, 0),$$

be drawn independently, and define

$$\Psi(t) = \psi_{1,n,k}(\mathcal{S}'\mathcal{S}, \mathcal{S}'t, t't, \mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_3),$$

so that the critical value function for given $\bar{T} = t$ is given by $c_{1,\alpha}(t)$, the $(1-\alpha)$-th quantile of $\Psi(t)$.

A.3.1. *Proof under Assumption 2 (a).* For the conclusion in (9), it is sufficient to show that

$$(n-k)\frac{\mu^2}{k}\Psi(\bar{T}) \text{ converges to some non-degenerate distribution,} \tag{12}$$

$$(n-k)\frac{\mu^2}{k}\{\Psi(\hat{T}) - \Psi(\bar{T})\} \xrightarrow{p} 0. \tag{13}$$

We first show (12). By explicitly computing the smallest eigenvalue in $\Psi(t)$, $\Psi(\bar{T})$ can be written as

$$\Psi(\bar{T}) = \frac{\mathcal{S}'\mathcal{S}}{\mathcal{W}_1} + \frac{b + \sqrt{b^2 - 4ac}}{2a}, \tag{14}$$

where

$$\begin{aligned} a &= \frac{1}{(n-k)^2}(\mathcal{W}_1\mathcal{W}_3 - \mathcal{W}_2^2), \\ b &= \frac{1}{(n-k)^2}\{-\mathcal{W}_1(\bar{T}'\bar{T}) - (\mathcal{S}'\mathcal{S})\mathcal{W}_3 + 2(\mathcal{S}'\bar{T})\mathcal{W}_2\}, \\ c &= \frac{1}{(n-k)^2}\{(\mathcal{S}'\mathcal{S})(\bar{T}'\bar{T}) - (\mathcal{S}'\bar{T})^2\}. \end{aligned}$$

To proceed, we express a, b, and c by the following standardized objects

$$\begin{aligned} \mathcal{Z}_1 &= \frac{\mathcal{S}'\mathcal{S} - k}{\sqrt{k}}, \qquad \mathcal{Z}_2 = \frac{1}{\sqrt{k}}\mathcal{S}'\bar{T}, \qquad \mathcal{Z}_{\bar{T}} = \frac{\bar{T}'\bar{T} - k - \mu^2}{\sqrt{k}}, \\ \mathcal{Q}_1 &= \frac{\mathcal{W}_1 - (n-k)}{\sqrt{n-k}}, \qquad \mathcal{Q}_2 = \frac{\mathcal{W}_2}{\sqrt{n-k}}, \qquad \mathcal{Q}_3 = \frac{\mathcal{W}_3 - (n-k)}{\sqrt{n-k}}, \end{aligned}$$

where $\mu^2 = (A_0'\Omega^{-1}A_0)^{-1/2}A_0'\Omega^{-1}\Pi'Z'Z\Pi\Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}$ is the concentration parameter defined in Assumption 2 (a). Based on this notation, a, b, and c are written as

$$\begin{aligned}
a &= 1 + \frac{\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n-k}} + \frac{\mathcal{Q}_1\mathcal{Q}_3 - \mathcal{Q}_2^2}{n-k}, \\
b &= -\frac{k}{n-k}\left\{2 + \frac{\mu^2}{k} + \frac{\left(1+\frac{\mu^2}{k}\right)\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n-k}} + \frac{\mathcal{Z}_1 + \mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\mathcal{Z}_{\bar{T}}\mathcal{Q}_1 + \mathcal{Z}_1\mathcal{Q}_3 - 2\mathcal{Z}_2\mathcal{Q}_2}{\sqrt{k}\sqrt{n-k}}\right\}, \\
c &= \frac{k^2}{(n-k)^2}\left\{1 + \frac{\mu^2}{k} + \frac{\left(1+\frac{\mu^2}{k}\right)\mathcal{Z}_1 + \mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\mathcal{Z}_1\mathcal{Z}_{\bar{T}} - \mathcal{Z}_2^2}{k}\right\}.
\end{aligned}$$

Based on these expressions and by using $\frac{\mu^2}{\sqrt{k}} \to \infty$ (Assumption 2 (a)), we can expand the second term of $\Psi(\bar{T})$ in (14), that is

$$\begin{aligned}
&\frac{-b - \sqrt{b^2 - 4ac}}{2a} \\
&= \left[(A_0 + A_1 + A_2) - B_0\left\{1 + \frac{1}{2}(B_1 + B_2) - \frac{1}{8}B_1^2\right\}\right]\{1 - (C_1 + C_2) + C_1^2\} + o_p(n^{-1}) \\
&= (A_0 - B_0) + \left\{A_1 - \frac{1}{2}B_0B_1 - (A_0 - B_0)C_1\right\} \\
&\quad + \left\{A_2 - \frac{1}{2}B_0B_2 + \frac{1}{8}B_0B_1^2 + (A_0 - B_0)(-C_2 + C_1^2) - (A_1 - \frac{1}{2}B_0B_1)C_1\right\} + o_p(n^{-1}), \quad (15)
\end{aligned}$$

where

$$\begin{aligned}
A_0 &= \frac{1}{2}\frac{k}{n-k}\left(2 + \frac{\mu^2}{k}\right), \qquad A_1 = \frac{1}{2}\frac{k}{n-k}\left\{\frac{\mathcal{Z}_1 + \mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\left(1+\frac{\mu^2}{k}\right)\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n-k}}\right\}, \\
A_2 &= \frac{1}{2}\frac{k}{n-k}\left(\frac{\mathcal{Z}_{\bar{T}}\mathcal{Q}_1 + \mathcal{Z}_1\mathcal{Q}_3 - 2\mathcal{Z}_2\mathcal{Q}_2}{\sqrt{k}\sqrt{n-k}}\right), \\
B_0 &= \frac{1}{2}\frac{k}{n-k}\frac{\mu^2}{k}, \qquad B_1 = \left(\frac{\mu^2}{k}\right)^{-1}\left\{-\frac{2\mathcal{Z}_1}{\sqrt{k}} + \frac{2\mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{2\left(1+\frac{\mu^2}{k}\right)\mathcal{Q}_1}{\sqrt{n-k}} - \frac{2\mathcal{Q}_3}{\sqrt{n-k}}\right\}, \\
B_2 &= \left(\frac{\mu^2}{k}\right)^{-2}\left[\begin{array}{c}
\frac{\left(1+\frac{\mu^2}{k}\right)^2}{n-k}\mathcal{Q}_1^2 + \frac{4+4\frac{\mu^2}{k}}{n-k}\mathcal{Q}_2^2 + \frac{1}{n-k}\mathcal{Q}_3^2 - \frac{2+2\frac{\mu^2}{k}}{n-k}\mathcal{Q}_1\mathcal{Q}_3 \\
+\frac{1}{k}\mathcal{Z}_{\bar{T}}^2 + \frac{1}{k}\mathcal{Z}_1^2 + \frac{4}{k}\mathcal{Z}_2^2 - \frac{2}{k}\mathcal{Z}_1\mathcal{Z}_{\bar{T}} \\
+\frac{2+4\frac{\mu^2}{k}}{\sqrt{k}\sqrt{n-k}}\mathcal{Z}_{\bar{T}}\mathcal{Q}_1 - \frac{2}{\sqrt{k}\sqrt{n-k}}\mathcal{Z}_{\bar{T}}\mathcal{Q}_3 \\
-\frac{2+2\frac{\mu^2}{k}}{\sqrt{k}\sqrt{n-k}}\mathcal{Z}_1\mathcal{Q}_1 + \frac{2-2\frac{\mu^2}{k}}{\sqrt{k}\sqrt{n-k}}\mathcal{Z}_1\mathcal{Q}_3 - \frac{8+4\frac{\mu^2}{k}}{\sqrt{k}\sqrt{n-k}}\mathcal{Z}_2\mathcal{Q}_2
\end{array}\right], \\
C_1 &= \frac{\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n-k}}, \qquad C_2 = \frac{\mathcal{Q}_1\mathcal{Q}_3 - \mathcal{Q}_2^2}{n-k}.
\end{aligned}$$

Also an expansion of the first term of (14) is

$$\frac{\mathcal{S}'\mathcal{S}}{\mathcal{W}_1} = \frac{k}{n-k}\left\{1 + \frac{\mathcal{Z}_1}{\sqrt{k}}\right\}\left\{1 - \frac{\mathcal{Q}_1}{\sqrt{n-k}} + \frac{\mathcal{Q}_1^2}{n-k}\right\} + o_p(n^{-1}).$$

Combining these results with direct calculations under Assumption 2 (a), an expansion for $\Psi(\bar{T})$ is obtained as

$$
\begin{aligned}
(n-k)\frac{\mu^2}{k}\Psi(\bar{T}) &= \frac{k}{n-k}\mathcal{Q}_2^2 + \frac{\sqrt{k}}{n-k}\frac{\mu^2}{k}\mathcal{Z}_1\mathcal{Q}_1^2 - 2\sqrt{\frac{k}{n-k}}\mathcal{Z}_2\mathcal{Q}_2 + \mathcal{Z}_2^2 + o_p(1) \\
&= \left(\mathcal{Z}_2 - \sqrt{\frac{\alpha}{1-\alpha}}\mathcal{Q}_2\right)^2 + o_p(1), \qquad (16)
\end{aligned}
$$

where the second equality follows from $\frac{k}{n} \to \alpha$ and $\frac{\mu^2}{k} = O(1)$. Since $(\mathcal{Z}_2, \mathcal{Q}_2)$ converges to a non-degenerate distribution, we obtain (12).

We now show (13). A similar argument to derive (16) (by replacing $\bar{T}$ with $\hat{T}$) yields that

$$
(n-k)\frac{\mu^2}{k}\Psi(\hat{T}) = \left(\hat{\mathcal{Z}}_2 - \sqrt{\frac{\alpha}{1-\alpha}}\mathcal{Q}_2\right)^2 + o_p(1),
$$

where $\hat{\mathcal{Z}}_2 = \frac{1}{\sqrt{k}}\mathcal{S}'\hat{T}$. Thus it is sufficient for (12) to show that

$$
\hat{\mathcal{Z}}_2 = \mathcal{Z}_2 + o_p(1). \qquad (17)
$$

Let

$$
F = [b_0(b_0'\Omega b_0)^{-1/2} : \Omega^{-1}A_0(A_0'\Omega^{-1}A_0)^{-1/2}].
$$

Note that

$$
\begin{aligned}
\mathcal{S}'\hat{T} &= \mathcal{S}'(Z'Z)^{-1/2}Z'Y\hat{\Omega}^{-1}A_0(A_0'\hat{\Omega}^{-1}A_0)^{-1/2} \\
&= \sqrt{n-k}\mathcal{S}'(Z'Z)^{-1/2}Z'Y(Y'M_ZY)^{-1}A_0(A_0'(Y'M_ZY)^{-1}A_0)^{-1/2} \\
&= \sqrt{n-k}\mathcal{S}'(Z'Z)^{-1/2}Z'YF\begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T}\end{pmatrix}^{-1}F'A_0\left(A_0'F\begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T}\end{pmatrix}^{-1}F'A_0\right)^{-1/2} \\
&= \sqrt{n-k}(\mathcal{S}'\bar{S}, \mathcal{S}'\bar{T})\begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T}\end{pmatrix}^{-1}\begin{pmatrix} 0 \\ \left\{\frac{1}{(\tilde{S}'\tilde{S})(\tilde{T}'\tilde{T})-(\tilde{S}'\tilde{T})^2}(\tilde{S}'\tilde{S})\right\}^{-1/2}\end{pmatrix} \\
&= \sqrt{n-k}\left\{\frac{1}{(\tilde{S}'\tilde{S})(\tilde{T}'\tilde{T})-(\tilde{S}'\tilde{T})^2}\right\}^{1/2}\left\{-(\mathcal{S}'\bar{S})(\tilde{S}'\tilde{T})(\tilde{S}'\tilde{S})^{-1/2} + (\mathcal{S}'\bar{T})(\tilde{S}'\tilde{S})^{1/2}\right\}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\hat{\mathcal{Z}}_2 &= \frac{\sqrt{k}}{\sqrt{n-k}}\left[\frac{1}{\left\{\frac{\tilde{S}'\tilde{S}-(n-k)}{n-k}+1\right\}\left\{\frac{\tilde{T}'\tilde{T}-(n-k)}{n-k}+1\right\} - \frac{1}{n-k}\left(\frac{\tilde{S}'\tilde{T}}{\sqrt{n-k}}\right)^2}\right]^{1/2} \\
&\quad \times\left\{-\frac{1}{\sqrt{k}}\left(\frac{\mathcal{S}'\bar{S}}{\sqrt{k}}\right)\left(\frac{\tilde{S}'\tilde{T}}{\sqrt{n-k}}\right)\left(\frac{\tilde{S}'\tilde{S}-(n-k)}{n-k}+1\right)^{-1/2}\right. \\
&\quad \left.+\frac{\sqrt{n-k}}{\sqrt{k}}\left(\frac{\mathcal{S}'\bar{T}}{\sqrt{k}}\right)\left(\frac{\tilde{S}'\tilde{S}-(n-k)}{n-k}+1\right)^{1/2}\right\} \\
&= \mathcal{Z}_2 + o_p(1),
\end{aligned}
$$

i.e., (17) holds true. Therefore, we obtain (13). Since (12) and (13) are satisfied, the conclusion follows.

A.3.2. *Proof under Assumption 2 (b).* By an analogous argument in the proof of Moreira (2003, Theorem 2), we can see that the conclusion under Assumption 2 (b) follows by:

$$\hat{\mathcal{Z}}_2 = \mathcal{Z}_2 + o_p(1), \tag{18}$$

$$\hat{\mathcal{Z}}_{\hat{T}} = \mathcal{Z}_{\bar{T}} + o_p(1). \tag{19}$$

For (18), we can apply the same argument as the proof of (17) (since it does not use the condition on μ^2). For (19), note that

$$\begin{aligned}
\frac{1}{k}\hat{T}'\hat{T} &= (A_0'\hat{\Omega}^{-1}A_0)^{-1/2}A_0'\hat{\Omega}^{-1}Y'P_ZY\hat{\Omega}^{-1}A_0(A_0'\hat{\Omega}^{-1}A_0)^{-1/2} \\
&= (n-k)(A_0'(Y'M_ZY)^{-1}A_0)^{-1/2}A_0'(Y'M_ZY)^{-1}Y'P_ZY(Y'M_ZY)^{-1}A_0(A_0'(Y'M_ZY)^{-1}A_0)^{-1/2} \\
&= \frac{n-k}{k}\{(\tilde{S}'\tilde{S})(\tilde{T}'\tilde{T})-(\tilde{S}'\tilde{T})^2\}^{-1}\begin{pmatrix} -(\tilde{S}'\tilde{T})(\tilde{S}'\tilde{S})^{-1/2} \\ (\tilde{S}'\tilde{S})^{1/2}\end{pmatrix}'\begin{pmatrix} \bar{S}'\bar{S} & \bar{S}'\bar{T} \\ \bar{S}'\bar{T} & \bar{T}'\bar{T}\end{pmatrix}\begin{pmatrix} -(\tilde{S}'\tilde{T})(\tilde{S}'\tilde{S})^{-1/2} \\ (\tilde{S}'\tilde{S})^{1/2}\end{pmatrix} \\
&= \frac{n-k}{k}\{(\tilde{S}'\tilde{S})(\tilde{T}'\tilde{T})-(\tilde{S}'\tilde{T})^2\}^{-1}\left\{(\bar{S}'\bar{S})(\tilde{S}'\tilde{T})^2(\tilde{S}'\tilde{S})^{-1}-2(\bar{S}'\bar{T})(\tilde{S}'\tilde{T})+(\bar{T}'\bar{T})(\tilde{S}'\tilde{S})\right\} \\
&= \left[\left\{\frac{\tilde{S}'\tilde{S}-(n-k)}{n-k}+1\right\}\left\{\frac{\tilde{T}'\tilde{T}-(n-k)}{n-k}+1\right\}-\frac{1}{n-k}\left(\frac{\tilde{S}'\tilde{T}}{\sqrt{n-k}}\right)^2\right]^{-1} \\
&\quad\times\left\{\frac{1}{n-k}\left(\frac{\bar{S}'\bar{S}-k}{k}+1\right)\left(\frac{\tilde{S}'\tilde{T}}{\sqrt{n-k}}\right)^2\left(\frac{\tilde{S}'\tilde{S}-(n-k)}{n-k}+1\right)^{-1}\right. \\
&\quad\left.-\frac{2}{\sqrt{k}\sqrt{n-k}}\left(\frac{\bar{S}'\bar{T}}{\sqrt{k}}\right)\left(\frac{\tilde{S}'\tilde{T}}{\sqrt{n-k}}\right)+\left(\frac{\bar{T}'\bar{T}-\mu^2-k}{k}+\frac{\mu^2}{k}+1\right)\left(\frac{\tilde{S}'\tilde{S}-(n-k)}{n-k}+1\right)\right\} \\
&= \frac{\bar{T}'\bar{T}-\mu^2-k}{k}+\left(\frac{\mu^2}{k}+1\right)\frac{\tilde{S}'\tilde{S}-(n-k)}{n-k}+\left(\frac{\mu^2}{k}+1\right)+O_p(K^{-1}),
\end{aligned}$$

where the first five equalities follow from the definitions and direct algebra, and the last equality follows from $\frac{\tilde{S}'\tilde{T}}{\sqrt{n-k}}=O_p(1)$, $\frac{\bar{S}'\bar{S}-k}{\sqrt{k}}=O_p(1)$, $\frac{\tilde{S}'\tilde{S}-(n-k)}{\sqrt{n-k}}=O_p(1)$, and $\frac{\bar{S}'\bar{T}}{\sqrt{k}}=O_p(1)$. Therefore, (19) is verified as:

$$\hat{\mathcal{Z}}_{\hat{T}} = \mathcal{Z}_{\bar{T}} + \sqrt{\frac{k}{n-k}}\left(\frac{\mu^2}{k}+1\right)\frac{\tilde{S}'\tilde{S}-(n-k)}{\sqrt{n-k}} + o_p(1) = \mathcal{Z}_{\bar{T}} + o_p(1),$$

where the second equality follows from $k/n \to 0$ (Assumption 2 (b)).

A.4. **Proof of Theorem 3.** Under $k/n \to 0$ (Assumption 2 (b)), $\left(\frac{\tilde{S}'\tilde{S}-(n-k)}{\sqrt{n-k}}, \frac{\tilde{S}'\tilde{T}}{\sqrt{n-k}}, \frac{\tilde{T}'\tilde{T}-(n-k)}{\sqrt{n-k}}\right)$ are of smaller order than $(\bar{\mathcal{Z}}_1, \bar{\mathcal{Z}}_2, \bar{\mathcal{Z}}_{\bar{T}}) = \left(\frac{\bar{S}'\bar{S}-k}{\sqrt{k}}, \frac{\bar{S}'\bar{T}}{\sqrt{k}}, \frac{\bar{T}'\bar{T}-k-\mu^2}{\sqrt{k}}\right)$. Thus, it is enough for the conclusion to show that non-normality of the errors does not affect the limit of the variance, i.e.,

$$Var(\bar{\mathcal{Z}}_1, \bar{\mathcal{Z}}_2, \bar{\mathcal{Z}}_{\bar{T}}) \to \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

To see this, let $u_i = Y_i' b_0$, $w_i = Y_i' \Omega^{-1} A_0$, $\sigma_u^2 = Var(u_i)$, $\kappa_u = E[u_i^4]$, and P_{ij} be the (i,j)-th element of P_Z. We have

$$\begin{aligned}
Var(\bar{\mathcal{Z}}_1) &= \frac{1}{k\sigma_u^4}\left\{E\left[\left(\sum_{i=1}^n\sum_{j=1}^n u_i u_j P_{ij}\right)^2\right] - \left(E\left[\sum_{i=1}^n\sum_{j=1}^n u_i u_j P_{ij}\right]\right)^2\right\} \\
&= \frac{1}{k\sigma_u^4}\left\{E\left[\sum_{i=1}^n\sum_{j\neq i}^n u_i^2 u_j^2 (2P_{ij}^2 + P_{ii}P_{jj}) + \sum_{i=1}^n u_i^4 P_{ii}^2\right] - \left(E\left[\sum_{i=1}^n u_i^2 P_{ii}\right]\right)^2\right\} \\
&= \frac{1}{k\sigma_u^4}\left\{\sigma_u^4\left\{\sum_{i=1}^n\sum_{j=1}^n (2P_{ij}^2 + P_{ii}P_{jj}) - 3\sum_{i=1}^n P_{ii}^2\right\} + \kappa_u \sum_{i=1}^n P_{ii}^2 - \sigma_u^4\left(\sum_{i=1}^n P_{ii}\right)^2\right\} \\
&= \frac{1}{k\sigma_u^4}\left\{2k\sigma_u^4 + (\kappa_u - 3\sigma_u^4)\sum_{i=1}^n P_{ii}^2\right\} \to 2,
\end{aligned}$$

where the fourth equality follows from $\sum_{i=1}^n\sum_{j=1}^n P_{ij}^2 = \sum_{i=1}^n P_{ii} = k$, and the convergence follows from the assumption $\frac{1}{k}\sum_{i=1}^n P_{ii}^2 \to 0$. Similarly, letting $\sigma_w^2 = Var(w_i)$, we have

$$\begin{aligned}
Cov(\bar{\mathcal{Z}}_1, \bar{\mathcal{Z}}_2) &= \frac{1}{k\sigma_u^3\sigma_w}\left\{E\left[\left(\sum_{i=1}^n\sum_{j=1}^n u_i u_j P_{ij}\right)\left(\sum_{i=1}^n\sum_{j=1}^n u_i w_j P_{ij}\right)\right]\right. \\
&\qquad \left. - E\left[\sum_{i=1}^n\sum_{j=1}^n u_i u_j P_{ij}\right] E\left[\sum_{i=1}^n\sum_{j=1}^n u_i w_j P_{ij}\right]\right\} \\
&= \frac{1}{k\sigma_u^3\sigma_w} E\left[\left(\sum_{i=1}^n\sum_{j=1}^n u_i u_j P_{ij}\right)\left(\sum_{i=1}^n\sum_{j=1}^n u_i w_j P_{ij}\right)\right] \\
&= \frac{1}{k\sigma_u^3\sigma_w} E\left[\sum_{i=1}^n u_i^3 w_i P_{ii}^2\right] = \frac{E[u_i^3 w_i]}{\sigma_u^2\sigma_{uw}}\frac{1}{k}\sum_{i=1}^n P_{ii}^2 \to 0,
\end{aligned}$$

where the second equality follows from $E[u_i w_i] = 0$.

The limits of the other elements can be shown in the same manner, so we obtain the conclusion.

References

Andrews, Donald, Marcelo Moreira, and James Stock (2006). "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression". In: *Econometrica* 74.3, pp. 715–752.

Andrews, Donald and James Stock (2007). "Testing with many weak instruments". In: *Journal of Econometrics* 138.1, pp. 24–46.

Angrist, Joshua and Alan Krueger (1991). "Does compulsory school attendance affect schooling and earnings?" In: *The Quarterly Journal of Economics* 106.4, pp. 979–1014.

Autor, David et al. (2019). "Disability Benefits, Consumption Insurance, and Household Labor Supply". In: *American Economic Review* 109.7, pp. 2613–54. DOI: `10.1257/aer.20151231`. URL: `https://www.aeaweb.org/articles?id=10.1257/aer.20151231`.

Chao, John and Norman Swanson (2005). "Consistent estimation with a large number of weak instruments". In: *Econometrica* 73.5, pp. 1673–1692.

Crudu, Federico, Giovanni Mellace, and Zsolt Sándor (2021). "Inference in instrumental variable models with heteroskedasticity and many instruments". In: *Econometric Theory* 37.2, pp. 281–310.

Dahl, Gordon, Andreas Ravndal Kostøl, and Magne Mogstad (2014). "Family welfare cultures". In: *The Quarterly Journal of Economics* 129.4, pp. 1711–1752.

Fama, Eugene and James MacBeth (1973). "Risk, return, and equilibrium: Empirical tests". In: *Journal of political economy* 81.3, pp. 607–636.

Han, Chirok and Peter CB Phillips (2006). "GMM with many moment conditions". In: *Econometrica* 74.1, pp. 147–192.

Hansen, Christian, Jerry Hausman, and Whitney Newey (2008). "Estimation with many instrumental variables". In: *Journal of Business & Economic Statistics* 26.4, pp. 398–422.

Hausman, Jerry A et al. (2012). "Instrumental variable estimation with heteroskedasticity and many instruments". In: *Quantitative Economics* 3.2, pp. 211–255.

Matsushita, Yukitoshi and Taisuke Otsu (2022). "A jackknife Lagrange multiplier test with many weak instruments". In: *Econometric Theory*.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James Stock (2014). "Empirical evidence on inflation expectations in the New Keynesian Phillips Curve". In: *Journal of Economic Literature* 52.1, pp. 124–88.

Mikusheva, Anna and Liyang Sun (Dec. 2021). "Inference with Many Weak Instruments". In: *The Review of Economic Studies*. rdab097. ISSN: 0034-6527.

Mills, Benjamin, Marcelo Moreira, and Lucas Vilela (2014). "Tests based on t-statistics for IV regression with weak instruments". In: *Journal of Econometrics* 182.2, pp. 351–363.

Moreira, Humberto and Marcelo J Moreira (2019). "Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors". In: *Journal of Econometrics* 213.2, pp. 398–433.

Moreira, Marcelo (2003). "A conditional likelihood ratio test for structural models". In: *Econometrica* 71.4, pp. 1027–1048.

Newey, Whitney K and Frank Windmeijer (2009). "Generalized method of moments with many weak moment conditions". In: *Econometrica* 77.3, pp. 687–719.

Staiger, Douglas and James Stock (1997). "Instrumental Variables Regression with Weak Instruments". In: *Econometrica: Journal of the Econometric Society*, pp. 557–586.

Stock, James and Motohiro Yogo (2005). "Testing for Weak Instruments in Linear IV Regression". In: *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, p. 80.

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

Email address: s.ayyar1@lse.ac.uk

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