Econometric Theory, **0**, 2025, 1–27.
doi:10.1017/S0266466625000088

CONDITIONAL LIKELIHOOD RATIO TEST WITH MANY WEAK INSTRUMENTS

Sreevidya Ayyar
London School of Economics

Yukitoshi Matsushita
Hitotsubashi University

Taisuke Otsu
London School of Economics

This article extends the validity of the conditional likelihood ratio (CLR) test developed by Moreira (2003, *Econometrica* 71(4), 1027—1048) to instrumental variable regression models with unknown homoskedastic error variance and many weak instruments. We argue that the conventional CLR test with estimated error variance loses exact similarity and is asymptotically invalid in this setting. We propose a modified critical value function for the likelihood ratio (LR) statistic with estimated error variance, and prove that our modified test achieves asymptotic validity under many weak instruments asymptotics. Our critical value function is constructed by representing the LR using four statistics, instead of two as in Moreira (2003, *Econometrica* 71(4), 1027—1048). A simulation study illustrates the desirable finite sample properties of our test.

1. INTRODUCTION

Inference in regression models with endogenous variables and many weak instruments is becoming increasingly relevant in applied research. Researchers often rely on standard asymptotic approximations when conducting inference in the presence of many weak instruments. 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 the efficiency of estimators or their associated tests, but when instruments are weak it can also exacerbate the poor finite sample properties of standard inference procedures.

Several previous papers have noted this issue. Chao and Swanson (2005), Han and Phillips (2006), and Newey and Windmeijer (2009) generalize the many instruments asymptotic theory to allow for weak instruments or moments. Andrews and

We are grateful to Naoto Kunitomo, Anna Mikusheva, Peter Phillips, and anonymous referees for helpful comments. Y.M. acknowledges financial support from the JSPS KAKENHI (23K01331). Address correspondence to Department of Economics, London School of Economics, London, WC2A 2AE, UK, e-mail: t.otsu@lse.ac.uk

© The Author(s), 2025. Published by Cambridge University Press.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

Stock (2007b) 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 instruments robust standard error and a modification of the LM test, while Hausman et al. (2012) propose Wald tests with the limited information maximum likelihood (LIML) 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 (2022), and the jackknife LM test by Matsushita and Otsu (2024).

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 for 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 article, we set out to investigate the performance of Moreira's (2003) CLR approach when k is relatively large and is allowed to grow proportionally with the sample size, n. Size robustness of the CLR test under $k = o(n^{1/3})$ has already been established by Andrews and Stock (2007b). However, we show that in a setting with homoskedastic normal errors and unknown variance, if k is allowed to grow much faster than $n^{1/3}$, then the conventional CLR test loses exact similarity and is asymptotically invalid under many weak instruments asymptotics. 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 as (or slower than) the sample size and (ii) weak instruments. 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 (LR) using four statistics, instead of two as in Moreira (2003). Our MCLR test retains asymptotic validity when there are many weak instruments, under a mild condition on identification strength. This result holds even when we relax the assumption of normally distributed error terms, as long as we impose an additional moment condition.

A substantive limitation of our approach is that all theoretical results are derived under the assumption of homoskedastic errors. Several existing inference methods (e.g., Hausman et al. (2012), Crudu et al. (2021), Mikusheva and Sun (2022), and Matsushita and Otsu (2024)) are robust to error terms being heteroskedastic, which is admittedly the more relevant case for applied research. This article should be considered a building block toward further generalizations of the CLR approach. A key observation of our MCLR approach is that in the case of homoskedastic

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

errors, the LR statistic with many weak instruments can be written as a function of four statistics, instead of two for the conventional CLR statistic (see Proposition 1 below). In the case of general heteroskedastic errors, such a representation of the test statistic by a finite number of statistics is typically unavailable. While this is beyond the scope of this article, it is an interesting avenue for future research to extend our approach to allow certain patterns of heteroskedasticity, such as the Kronecker product structure studied by Moreira and Moreira (2019).

The rest of this article is organized as follows. Section 2 introduces our setup and the LR statistic when error variance is unknown. We discuss a representation of the LR statistic 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 why the conventional CLR critical value function lacks validity in our setting. Section 4 illustrates the usefulness of our proposed method through a simulation study and proposes a pre-test procedure using our MCLR test. It also outlines how an applied researcher may compute critical values for the MCLR test. 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 is an $n \times 1$ vector of dependent variables, y_2 is an $n \times 1$ vector of endogenous regressors, β is a scalar unknown structural parameter, u is an $n \times 1$ vector of mean-zero disturbances, Z is an $n \times k$ matrix of instruments, π_2 is a $k \times 1$ vector of unknown parameters, and v_2 is an $n \times 1$ vector 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. Throughout this article, we focus on the model with a single endogenous regressor, leaving the case of multiple endogenous regressors to future research (see Section 5 for some discussion).

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 article is concerned with testing the null hypothesis $H_0 : \beta = \beta_0$ on the structural parameter, against the alternative $H_1 : \beta \neq \beta_0$, where the coefficients π_2 are treated as nuisance parameters. We focus on the situation where researchers only have many weak instruments at their disposal for testing $H_0 : \beta = \beta_0$.

To proceed, we impose the following assumptions. Let $a_0 = (\beta_0, 1)'$.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

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 nonrandom. One of the following two conditions holds.
   (a) $\frac{k}{n} \to \alpha \in (0,1)$ as $n \to \infty$, and the concentration parameter

   $$\mu^2 = (a_0'\Omega^{-1}a_0)^{-1}a_0'\Omega^{-1}\Pi'Z'Z\Pi\Omega^{-1}a_0, \tag{3}$$

   satisfies $\mu^2 = O(n)$ and $\frac{\mu^2}{\sqrt{k}} \to \infty$ as $n \to \infty$; or
   (b) $\frac{k}{n} \to 0$ as $n \to \infty$ (without any condition on μ^2), where k is fixed or diverging.

Normality of the 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 a sufficient statistic for Π 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 the distribution is continuous). Since we maintain Moreira (2003)'s conditional inference framework, we begin with normally distributed error terms, although we show that this assumption can be relaxed in our asymptotic analysis (see Theorem 3). Throughout this article, we focus on the case where Ω is unknown to researchers.

Assumption 2 concerns the instrumental variables. In this article, we restrict Z to be nonrandom, which is equivalent to conditioning on Z. To allow 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. For the MCLR test we propose, the condition $\frac{\mu^2}{\sqrt{k}} \to \infty$ is required to control the asymptotic size. Note that this condition is not required for correct size of alternative tests, such as the jackknife AR tests by Crudu et al. (2021) and Mikusheva and Sun (2022), and the jackknife LM test by Matsushita and Otsu (2024). In particular, if $\frac{\mu^2}{\sqrt{k}} = O(1)$, the result in (A.4) will be satisfied with a different normalization. However, under such a normalization, the result in (A.5) is typically violated. The main reason for this is that the normalized statistic $\frac{\bar{T}'\bar{T}-k-\mu^2}{\sqrt{k}}$ with known Ω (see, (5)) is not asymptotically equivalent to $\frac{\hat{T}'\hat{T}-k-\mu^2}{\sqrt{k}}$ with estimated $\hat{\Omega}$ (see, (10)) due to a non-negligible contribution of the estimation error of Ω as shown in (A.17). See Appendix A.3.2 for a detailed discussion. If k grows slower than n, as in Assumption 2(b), there is no requirement on μ^2; that is, the instruments can be arbitrarily weak.

Note that Wald tests based on many-instrument robust standard errors (Hansen et al., 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 (2007b) show that the conventional CLR test is asymptotically valid for 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 simulations, the MCLR test is especially preferable 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

$$LR_0 = \frac{b_0' Y' P_Z Y b_0}{b_0' \Omega b_0} - \bar{\lambda}, \tag{4}$$

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_2. Indeed, under H_0, T alone is a sufficient statistic for π_2.

By using standardized versions of S and T:

$$\bar{S} = (Z'Z)^{-1/2} Z'Yb_0 (b_0'\Omega b_0)^{-1/2}, \qquad \bar{T} = (Z'Z)^{-1/2} Z'Y\Omega^{-1}a_0 (a_0'\Omega^{-1}a_0)^{-1/2}, \tag{5}$$

the LR statistic LR_0 can be alternatively expressed as

$$LR_0 = \bar{S}'\bar{S} - \bar{\lambda} \equiv \psi_0(\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T}), \tag{6}$$

where $\bar{\lambda}$ is the smallest eigenvalue of $(\bar{S}, \bar{T})'(\bar{S}, \bar{T})$. See the proof of Moreira (2003, Prop. 1). If Ω is known, we can apply the conventional CLR test by Moreira (2003) based on LR_0, even with many weak instruments. Notice that in contrast to the AR statistic, $\bar{S}'\bar{S}$, the nonlinearity of LR_0 in $(\bar{S}, \bar{T})$ is non-quadratic. Conditional inference is typically conducted by conditioning on $\bar{T}$, or an estimator of $\bar{T}$, which leads to the distribution of the test statistic becoming non-standard and the critical values must be computed by simulation. This is a common feature for both the CLR test by Moreira (2003) and our MCLR test.

This article focuses on the case of unknown Ω, as stated in Assumption 1, so the conventional CLR test is infeasible. Its feasible counterpart, obtained by plugging in a consistent estimator of Ω, turns out to be invalid under many weak instruments asymptotics (see Remark 2 below).

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

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

$$\hat{\Omega} = \frac{1}{n-k} Y' M_Z Y, \tag{7}$$

where $M_Z = I_n - P_Z$ and I_n is the $n \times n$ identity matrix. By replacing Ω in (4) with the estimator $\hat{\Omega}$, the LR statistic for testing H_0 with unknown Ω is written as

$$\frac{LR_1}{n-k} = \frac{b_0' Y' P_Z Y b_0}{b_0' Y' M_Z Y b_0} - \hat{\lambda}, \tag{8}$$

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

To obtain an analogous expression to (6) 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 $(\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T}, \tilde{S}'\tilde{S}, \tilde{S}'\tilde{T}, \tilde{T}'\tilde{T})$:*[2]

$$\frac{LR_1}{n-k} = \psi_1(\bar{S}'\bar{S}, \bar{S}'\bar{T}, \bar{T}'\bar{T}, \tilde{S}'\tilde{S}, \tilde{S}'\tilde{T}, \tilde{T}'\tilde{T}).$$

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 (4). 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 that*

(i) $\bar{S} | \bar{T} = t \sim N(0, I_k)$ *and* $\bar{S}'\bar{T} | \bar{T} = t \sim \mathrm{N}(0, t't)$,
(ii) $\bar{S}$, $\bar{T}$, *and* $(\tilde{S}, \tilde{T})$ *are mutually independent,*
(iii) $\left(\begin{array}{cc} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T} \end{array} \right) \Bigg| \bar{T} = t \sim \mathrm{Wishart}(n-k, I_2)$.

[1] We note that $\hat{\Omega}$ is a natural choice to estimate Ω because (i) it is unbiased and consistent, and (ii) it yields independence of the denominator and numerator in the first term of (8), which greatly simplifies our theoretical development. Other estimators or proxies for Ω may be employed, as long as an analogous representation in Proposition 1 can be obtained.

[2] More precisely, $\psi_1(d_1, \ldots, d_6) = \frac{d_1}{d_4} - \lambda(d_1, \ldots, d_6)$, where $\lambda(d_1, \ldots, d_6)$ is the solution of $\left| \left(\begin{array}{cc} d_1 & d_2 \\ d_2 & d_3 \end{array} \right) - \lambda \left(\begin{array}{cc} d_4 & d_5 \\ d_5 & d_6 \end{array} \right) \right| = 0$ for λ.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

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 we show are 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. Moreover, under many weak instruments 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 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_2. Finally, notice that $\bar{T}'\bar{T}$ is centered at the concentration parameter μ^2, and therefore is a measure of how strongly identified the coefficients on the instruments in the first-stage are. We will use this fact in Section 4.

3. CLR TEST WITH MANY WEAK INSTRUMENTS

Based on the test statistic LR_1 and its properties, we now develop our conditional inference method. To begin with, recall that $\bar{T}$ is a sufficient statistic for π_2, and consider the critical value function for given $\bar{T} = t$:

$$c_{1,\eta}(t) \equiv (1-\eta)\text{th quantile of } \psi_1(\mathcal{S}'\mathcal{S}, \mathcal{S}'t, t't, \mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_3),$$

where ψ_1 is defined in Proposition 1, and $\mathcal{S} \sim \mathrm{N}(0, I_k)$ and $\begin{pmatrix} \mathcal{W}_1 & \mathcal{W}_2 \\ \mathcal{W}_2 & \mathcal{W}_3 \end{pmatrix} \sim$ Wishart$(n-k, I_2)$ are independent. Propositions 1 and 2 directly imply the following property of $c_{1,\eta}(t)$.

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

$$\Pr\left\{\frac{LR_1}{n-k} \geq c_{1,\eta}(\bar{T})\right\} = \eta. \tag{9}$$

This theorem says that if $\bar{T}$ is observable, the LR test using $\frac{LR_1}{n-k}$ with the critical value $c_{1,\eta}(\bar{T})$ is exactly similar. Note that $c_{1,\eta}(t)$ depends only on $\tau = t't$. However, since $\bar{T}$ is unobservable for the case of unknown Ω, a test based on (9) is infeasible.

To develop a feasible version, we estimate $\bar{T}$ by

$$\hat{T} = (Z'Z)^{-1/2} Z'Y\hat{\Omega}^{-1} a_0 (a_0'\hat{\Omega}^{-1}a_0)^{-1/2}, \tag{10}$$

where $\hat{\Omega}$ is as defined in (7). Based on this estimator, our proposed rejection rule is defined as

$$\text{Reject } H_0 \text{ if } \frac{LR_1}{n-k} \geq c_{1,\eta}(\hat{T}). \tag{11}$$

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

The next theorem is the main result of our article, and it establishes asymptotic validity of the MCLR test in (11).

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

$$\Pr\left\{\frac{LR_1}{n-k} \geq c_{1,\eta}(\hat{T})\right\} \to \eta \quad \text{as } n \to \infty. \tag{12}$$

Compared to Theorem 1, this theorem requires an additional condition (Assumption 2). For the case of $k/n \to \alpha \in (0,1)$ (Assumption 2(a)), an additional condition on the concentration parameter, $\mu^2/\sqrt{k} \to \infty$, is required to obtain (A.5) in Appendix A, which guarantees that replacing $\bar{T}$ with $\hat{T}$ does not change the limiting distribution of the test statistic. For the case of $k/n \to 0$ (Assumption 2(b)), such a requirement on μ^2 is unnecessary because the key asymptotic equivalence in (A.20) is guaranteed without any requirement on μ^2.

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 $P_Z = Z(Z'Z)^{-1}Z'$.

THEOREM 3. *Consider the setup in Section 2.1. The rows of V are independent and identically distributed with finite fourth moments. Under Assumption 2(b) and* $\frac{1}{k}\sum_{i=1}^{n} P_{ii}^2 \to 0$, *(12) is true.*

Specifically, as long as the number of instruments k grows slower than the sample size n and the projection matrix of instruments satisfies $\frac{1}{k}\sum_{i=1}^{n} P_{ii}^2 \to 0$, our MCLR test is asymptotically valid even when the reduced form errors are non-normal. The condition $\frac{1}{k}\sum_{i=1}^{n} P_{ii}^2 \to 0$, which is termed the design balance assumption in Cattaneo, Jansson, and Ma (2019), is used to guarantee that the limiting variance in (A.21) becomes identical to the Gaussian case. Note that Assumption 2(b) is still more general than $k = o(n^{1/3})$, which is imposed by Andrews and Stock (2007b) to establish the 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 (6) for testing $H_0 : \beta = \beta_0$ can be obtained as

$$c_{0,\eta}(t) = (1-\eta)\text{th quantile of } \psi_0(\mathcal{S}'\mathcal{S}, \mathcal{S}'t, t't),$$

where $\mathcal{S} \sim N(0, I_k)$. As shown by Moreira (2003), the test $\mathbb{I}\{LR_0 \geq c_{0,\eta}(\bar{T})\}$ is exactly similar for the case of known Ω (i.e., $\Pr\{LR_0 \geq c_{0,\eta}(\bar{T})\} = \eta$). When Ω is unknown, Moreira (2003) suggested to plug-in the estimator $\hat{\Omega}$ to the test statistic

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

LR_0 (which yields LR_1) and use $c_{0,\eta}(\hat{T})$, that is:

$$\text{Reject } H_0 \text{ if } LR_1 \geq c_{0,\eta}(\hat{T}). \tag{13}$$

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

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

Therefore, even if we ignore the estimation error arising from using $\hat{T}$ instead of $\bar{T}$, the conventional CLR test in (13) is asymptotically invalid in our setup.

4. NUMERICAL ILLUSTRATIONS

In this section, we compare the critical value function of the MCLR test, $c_{1,\eta}(t)$, with $c_{0,\eta}(t)$ of the conventional CLR test (Section 4.1). We then use Monte Carlo simulations to evaluate the finite sample performance of our MCLR test relative to existing alternatives. Finally, we employ our MCLR test to propose a two-step pre-test for homoskedasticity and weak identification.

4.1. Critical Value Function

The critical value function of our MCLR test $c_{1,\eta}(t)$ does not have a closed form, as is the case with Moreira's (2003) CLR critical value function, $c_{0,\eta}(t)$. Panel A of Table 1 presents critical values $(n-k)c_{1,\eta}(t)$ of the MCLR test, for the 5% significance level. Critical values are calculated using 5,000 Monte Carlo replications with $n = 100$, for different values of $\tau = \bar{T}'\bar{T}$. We choose to vary τ because it is directly indicative of identification strength and aids comparison with Moreira (2003), who presents critical values which are a function of τ.

As shown in Panel A of Table 1, when $k = 1$ the critical value function of the MCLR test is constant at 3.93 for all values of τ; the slight variation in the final row is attributable to numerical error. Interestingly, 3.93 is the 95th percentile of $F(1,99)$. This is in contrast to the critical value of the CLR test for $k = 1$, which is 3.84 and equal to the 95th percentile of $\chi^2(1)$. We suspect this difference arises because we use the LR statistic with unknown Ω; when written in terms of sufficient statistics, LR_0 is the sum of chi-square variables, whose degrees of freedom sum to 1 for $k = 1$, while LR_1 sums across ratios of chi-square distributed random variables.

Similar to the CLR test, the critical value function of the MCLR test for any given k has an approximately exponential shape. Figure 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), critical values are larger. When τ is large, the test behaves as if it were unconditional with critical values stable around 3.93.

For comparison, in Panel B of Table 1, we present the critical value function of the conventional CLR test, $(n-k)c_{0,\eta}(t)$. As per Theorem 2, this test predictably runs into size problems when there are many weak instruments. This has

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

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
50,000	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.	3.99	4.21	4.72	7.35
50,000	3.84	3.84	3.84	3.84	3.99	3.84	3.84	3.84

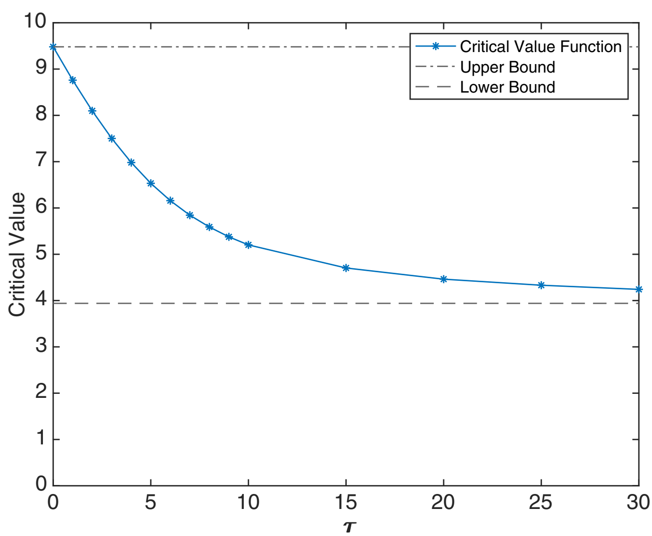

FIGURE 1. Critical value function of MCLR test with $k=4$.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

consequences for 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 over-reject the null hypothesis $H_0 : \beta = \beta_0$ when the number of weak instruments is large.

In practice, we suggest the following Algorithm 1 to compute the MCLR critical values, $c_{1,\eta}(\hat{t})$.

Algorithm 1 Computing MCLR critical values

```
Input Y,Z from data
Hypothesis H_0 : β = β_0
Compute t̂ = (Z'Z)^{-1/2} Z'YΩ̂^{-1} a_0 (a_0' Ω̂^{-1} a_0)^{-1/2} for Ω̂ =
1/(n-k) Y'(I_n - Z(Z'Z)^{-1} Z')Y and a_0 = (β_0, 1)'.
For m in 1,...,M

    (i) Independently draw S ~ N(0, I_k), a k×1
    vector of normal random variables, and
    W ~ Wishart(n-k, I_2). Define W_11 as the
    (1,1)-element of W.
    (ii) Compute λ̂ as the minimum eigenvalue of
    W^{-1} ( S'S  S't̂
             t̂'S  t̂'t̂ )
    (iii) Compute c_{1,s} = (n-k)[S'S/W_11 - λ̂]

Obtain {c_{1,m}}_{m=1}^M and set c_{1,η}(t̂) as the (1-η)-th
quantile of {c_{1,m}}_{m=1}^M. This is the critical value for
LR_1 at significance level η.
```

4.2. Simulation

We now turn to a simulation study, which is based on Design I of Staiger and Stock (1997). We allow for a single endogenous regressor and set $\beta_0 = 0$. Instruments are stochastic—Z comprises of a constant, Z_1, and i.i.d. draws from $\mathrm{N}(0, I_{k-1})$. In line with Assumption 1, the rows of (u, v_2) are i.i.d. normal random vectors with unit variances and correlation ρ. The latter parameter captures the degree of endogeneity of Y_2 in (1). Our sample size is $n = 100$ throughout.

Our simulations focus on the size and power performance of MCLR, relative to comparable hypothesis tests. We deviate from Staiger and Stock (1997)'s original design in two ways. First, we vary the number of instruments relative to the sample size to differentiate between cases which fall under Assumption 2(a) versus Assumption 2(b). Second, we vary the strength of our instruments. To do so, we use a population version of Stock and Yogo's (2005) pre-test for weak instruments. We use three different values of π_2 such that $\delta^2 = \pi_2' Z' Z \pi_2 / \omega_{22}$ takes the values 2

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

Table 2. Empirical rejection frequencies at 5% significance level.

ρ	δ^2	k	H-LIML	CLR	AR	mKLM	J-AR	M-CLR
0.2	30	5	0.041	0.057	0.028	0.052	0.050	0.051
0.2	30	10	0.030	0.060	0.036	0.057	0.054	0.057
0.2	30	30	0.031	0.089	0.047	0.051	0.057	0.052
0.2	10	5	0.032	0.060	0.025	0.049	0.052	0.047
0.2	10	10	0.018	0.064	0.031	0.051	0.052	0.051
0.2	10	30	0.016	0.084	0.036	0.060	0.049	0.049
0.2	2	5	0.008	0.050	0.026	0.049	0.048	0.041
0.2	2	10	0.007	0.057	0.030	0.050	0.050	0.041
0.2	2	30	0.013	0.100	0.044	0.055	0.051	0.055
0.6	30	5	0.054	0.057	0.029	0.062	0.051	0.051
0.6	30	10	0.058	0.057	0.034	0.060	0.053	0.051
0.6	30	30	0.052	0.081	0.042	0.051	0.049	0.057
0.6	10	5	0.077	0.059	0.027	0.053	0.044	0.053
0.6	10	10	0.069	0.053	0.034	0.046	0.052	0.049
0.6	10	30	0.081	0.093	0.047	0.049	0.054	0.059
0.6	2	5	0.091	0.042	0.024	0.045	0.049	0.035
0.6	2	10	0.092	0.057	0.034	0.047	0.055	0.047
0.6	2	30	0.086	0.089	0.040	0.052	0.052	0.046

(very weak instruments), 10 (weak instruments), and 30 (strong instruments), for different values of k. The population first-stage F-statistic corresponds to δ^2/k, and $\delta^2 = \frac{\mu^2}{\omega_{22}(a_0'\Omega^{-1}a_0)}$ is proportional to the concentration parameter μ^2.[3] The number of Monte Carlo replications is 5,000 for analyzing size and power, as well as for computing critical values.

For the null hypothesis $H_0 : \beta = 0$, Table 2 investigates the size properties of six tests: (i) the t-test with the heteroskedasticity robust LIML estimator by Hausman et al. (2012) (H-LIML), (ii) the CLR test by Moreira (2003), (iii) the homoskedastic AR test, (iv) the modified LM test by Hansen et al. (2008) (mKLM), (v) the jackknife version of the AR test by Mikusheva and Sun (2022) (J-AR) and (vi) our proposed modified CLR test (MCLR). We vary ρ, δ^2 and k across rows in Table 2.

We note that the size distortions of H-LIML are large, except when δ^2 and ρ^2 are large. 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

[3]Furthermore, note that in this design and under the null $H_0 : \beta = 0$, $\delta^2 = \frac{\mu^2}{\omega_{22}(a_0'\Omega^{-1}a_0)}$ can be written as $\mu^2 = \frac{\pi_2' Z' Z \pi_2}{1-\rho^2} = \frac{\delta^2}{1-\rho^2}$. In this article, ρ is treated as a constant in $(-1,1)$ so that the condition $\frac{\mu^2}{\sqrt{k}} \to \infty$ is equivalent to $\frac{\delta^2}{\sqrt{k}} \to \infty$. However, if we consider the case of $\rho^2 \to 1$, $\frac{\delta^2}{\sqrt{k}} \to \infty$ is sufficient, but not necessary, for $\frac{\mu^2}{\sqrt{k}} \to \infty$.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

over-rejected. The distortions of the test are most severe when δ^2 is small relative to k, and k is large.

The CLR test attains roughly the correct size when $k/n = 5/100$, even when identification is weak and the extent of endogeneity is high. However, size distortions can be observed when $k/n > 0.1$. Surprisingly, this is not visibly exacerbated by low δ^2, reaffirming that it is the existence of many instruments in the presence of some level of weak identification that has severe empirical consequences on the conventional CLR test. Overall, even when the CLR test experiences little size distortion, it always has an empirical rejection frequency farther from 5% than our proposed MCLR test.

The AR test consistently under-rejects the null hypothesis; while the distortions are not as severe as the H-LIML test, they are present across all combinations of δ^2, k, and ρ. We still investigate the power properties of this test, as we wish to investigate the power cost of ignoring the information in $\hat{\lambda}$ in the presence of many weak instruments.

The mKLM test works well. Although it tends to over-reject in some cases (e.g., high δ^2 and $\rho = 0.2$), its size distortions never exceed 2%. The J-AR test also appears robust to many weak instruments, and exhibits no distinct patterns of under- or over-rejection.

Compared to the other tests that we consider, the rejection frequencies of our MCLR test are on average closest to the nominal 5% level. As our theory in Section 3 suggests, the MCLR test is robust to weak instruments, many instruments, and many weak instruments.

Figure 2 presents calibrated (or size-adjusted) power curves for the MCLR, AR, J-AR, and mKLM tests for $H_0 : \beta = 0$, under the alternative hypotheses $H_1 : \beta = \Delta$. These power curves are plotted with respect to the 5% significance level; the critical values for these four tests are given by the 95th percentiles of their respective test statistics under $\mathbf{H_0}$, computed via 5,000 Monte Carlo replications. Each curve is plotted for $n = 100$ and $\beta_0 = 0$. We present four different cases, with $\rho = 0.2$ and different values of δ^2/k. 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. We set $k = 30$ in all cases.

Our MCLR test is uniformly more powerful at all values of δ^2/k, and the gain is more pronounced for low δ^2/k (that is, when instruments are weaker). The mKLM test experiences spurious declines in power under alternative hypotheses that are further away from the null, and has consistently low power when $\delta^2/k = 1/3$.[4] When δ^2/k is high, i.e., identification is the strongest among the cases we examine, the AR and J-AR have similar, although everywhere lower power, than our M-CLR test. Our simulations suggest that, in the presence of many weak instruments, the power cost of ignoring the information in $\hat{\lambda}$ (as the AR test statistic does) is greater

[4]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 both Kleibergen's LM statistics and its modification are quadratic forms of the score of the concentrated likelihood (see, p. 1788 of Kleibergen (2002)).

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

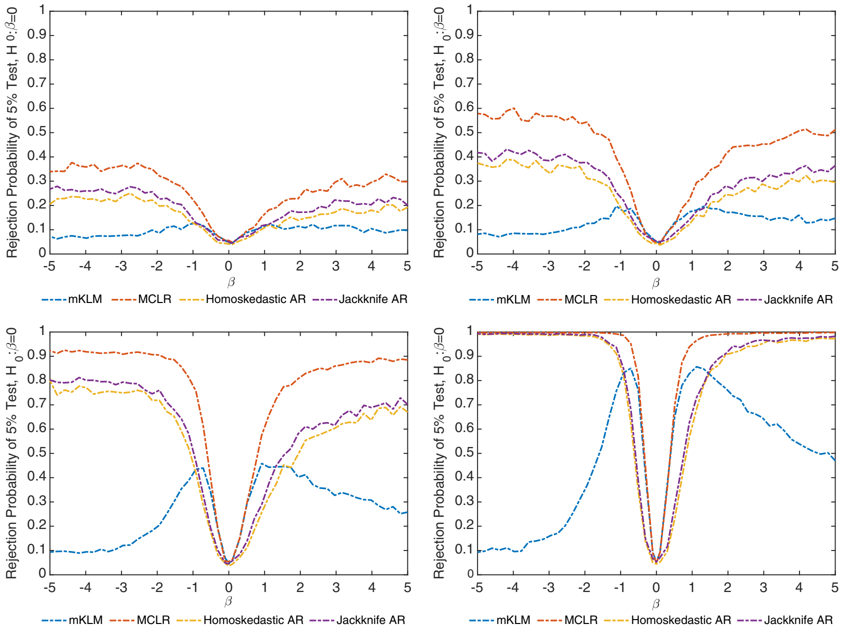

FIGURE 2. Calibrated power curves.
Note: From left to right, and top to bottom, these figures plot power curves for: $\delta^2/k = 1/3$, $\delta^2/k = 1/2$, $\delta^2/k = 1$ and $\delta^2/k = 2$, with $\rho = 0.2$, $k = 30$ and $n = 100$.

than the power cost of being robust to heteroskedasticity—indeed, the power of the AR test is much more comparable to that of the J-AR test than our MCLR test. While we do not present theoretical results on power, these findings suggest that the MCLR test shares the superior power properties of the conventional CLR test, which has near optimal power with small k Andrews et al., 2006.

4.3. Pre-Test for Homoskedasticity and Weak Identification

Based on our simulation study, in this subsection we propose a two-step pre-test for $H_0 : \beta = 0$.

The first step is to test whether $V = (v_1, v_2)$ is homoskedastic. For this, first we regress $Y = (Y_1, y_2)$ on the full matrix of instruments, retaining fitted values $\hat{Y} = (\hat{Y}_1, \hat{y}_2)$ and squared residuals $\hat{V}^2 = (\hat{v}_1^2, \hat{v}_2^2)$. Then, we construct a bivariate regression model of $\hat{v}_1^2$ on $\hat{Y}_1$ and $\hat{Y}_1^2$, and $\hat{v}_2^2$ on $\hat{y}_2$ and $\hat{y}_2^2$, and use a Wald test (at the 2.5% level, based on a Bonferroni correction) with the null hypothesis that the coefficients on $\hat{Y}_1$, $\hat{Y}_1^2$, $\hat{y}_2$, and $\hat{y}_2^2$ are equal to zero.

If the null of homoskedasticity is rejected, we apply the pre-test for weak identification outlined in Mikusheva and Sun (2022). For this, the authors propose the following test-statistic

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

$$\tilde{F} = \frac{1}{\sqrt{k\hat{\Upsilon}}} \sum_{i=1}^{n} \sum_{j \neq i} P_{ij} y_{2i} y_{2j},$$

where $\Upsilon = \frac{2}{k} \sum_{i=1}^{n} \sum_{j \neq i} \frac{P_{ij}^2}{M_{ii}M_{jj}+M_{ij}^2} y_{2i} M_i y_2 y_{2j} M_j y_2$, P_{ij} and M_{ij} are the (i,j)th element of P_Z and M_Z, respectively, and M_i is the ith column of M_Z. To achieve an overall size of 5% for this pre-test, the decision rule they propose is as follows: if $\tilde{F} > 9.98$, use the JIVE-Wald test, while for $\tilde{F} \leq 9.98$, use the J-AR test, both at the 2% significance level.[5]

If the null of homoskedasticity is not rejected, we also suggest to pre-test for weak identification but propose an alternative decision rule. If $\tilde{F} \leq 9.98$, use our MCLR test (at the 2% significance level). If $\tilde{F} > 9.98$, implement a t-test with the LIML estimator for β, also at the 2% significance level. For the t-test, we use the standard errors proposed by Hansen et al. (2008), which are robust to many instruments. Such a t-test has been shown to be powerful in the case of strong identification and homoskedastic errors Anderson, Kunitomo, and Matsushita, 2010.

To assess the performance of our two-step pre-test (henceforth, the MCLR pre-test), we compare its performance with the one-step pre-test of Mikusheva and Sun (2022) (the MS pre-test). Both pre-test procedures correspond to a nominal size of 5%, with a tolerance level of 5%. Table 3 reports the empirical size of both pre-tests for various combinations of ρ, δ^2, and, k. The empirical size of the MCLR pre-test is always within 2% of the nominal size, whereas for the MS pre-test, empirical size exceeds 7% when $\rho = 0.6$ and $\delta^2/k \leq 1$ (that is, when the degree of endogeneity of Y_2 is high and identification is relatively weak). For the other cases, there is little difference in empirical size between the two pre-testing procedures.

We assess the power of our MCLR pre-test in two ways. First, note that our results in Section 4.2 suggest that our pre-test should have a power advantage (in cases where $\tilde{F} < 9.98$) based on the uniformly greater power of the MCLR test compared to the J-AR test. To investigate whether this is the case, we analyze a one-step version of the MCLR pre-test (named MCLR one-step pre-test), which only tests for homoskedasticity—if the null of homoskedasticity is rejected, the J-AR test is used, and otherwise use the MCLR test. This will highlight any power gain from testing for homoskedasticity. To implement this analysis, we set (u, v_2) to be a function of the instruments (Z), instead of homoskedastic, with probability 0.5. We then compare the power of the MCLR one-step pre-test with the J-AR test. The latter would be the appropriate choice when there are many weak instruments and errors may be heteroskedastic.

Figure 3 presents the size-adjusted power curves from this comparison for $H_0 : \beta = 0$, under the alternative hypotheses $H_1 : \beta = \Delta$. Each curve is plotted at the 5% significance level for $n = 100$ and $\beta_0 = 0$. Analogous to Section 4.2,

[5] The relevant significance levels and critical values for achieving an overall size of 5% are given in Mikusheva and Sun (2022, Table 2).

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

TABLE 3. Empirical rejection frequencies of pre-test.

ρ	δ^2	k	M-CLR pre-test	MS pre-test
0.2	30	5	0.049	0.053
0.2	30	10	0.046	0.054
0.2	30	30	0.051	0.060
0.2	10	5	0.044	0.051
0.2	10	10	0.050	0.057
0.2	10	30	0.042	0.051
0.2	2	5	0.046	0.052
0.2	2	10	0.052	0.057
0.2	2	30	0.047	0.060
0.6	30	5	0.047	0.048
0.6	30	10	0.065	0.072
0.6	30	30	0.058	0.071
0.6	10	5	0.047	0.051
0.6	10	10	0.060	0.070
0.6	10	30	0.056	0.070
0.6	2	5	0.057	0.060
0.6	2	10	0.052	0.053
0.6	2	30	0.069	0.084

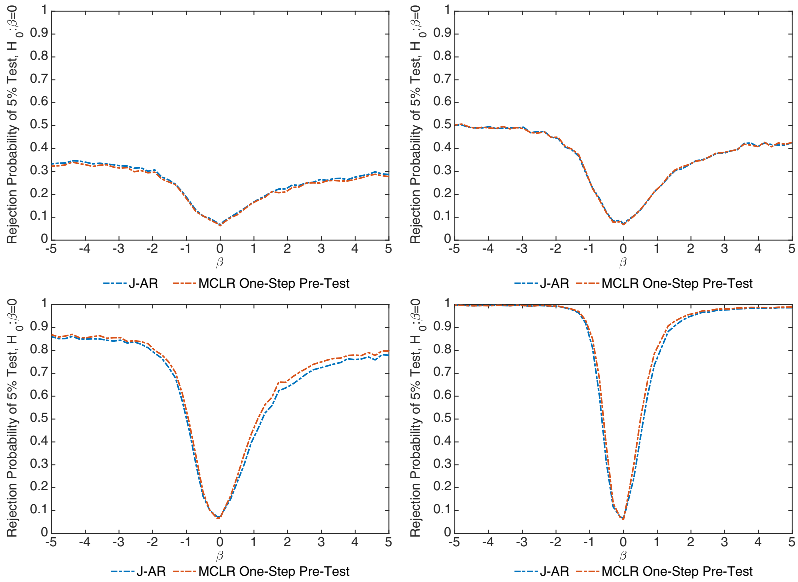

FIGURE 3. Power curves—MCLR one-step pre-test and J-AR test.
Note: From left to right, and top to bottom, these figures plot power curves for: $\delta^2/k = 1/3$, $\delta^2/k = 1/2$, $\delta^2/k = 1$ and $\delta^2/k = 2$, with $\rho = 0.2$, $k = 30$ and $n = 100$.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

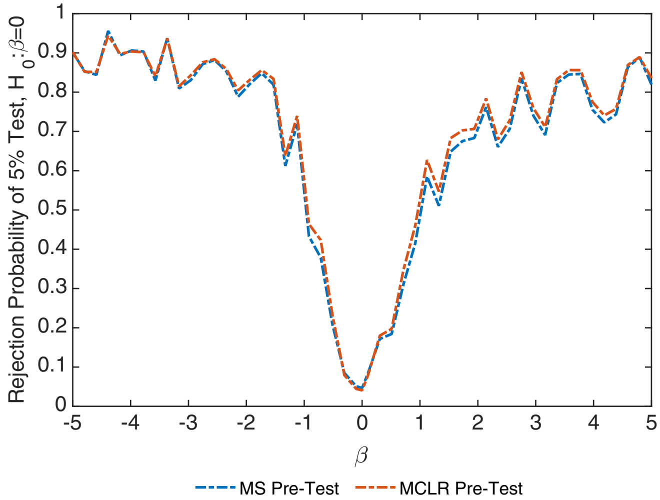

FIGURE 4. Power curves—MCLR and MS pre-tests.
Note: Power curves for the case of $k = 30$. We randomly select δ^2 from a uniform distribution $U[30,45]$ for each simulation draw.

we present four different cases, with $\rho = 0.2$, $k = 30$, and $\delta^2/k = 1/3, 1/2, 1, 2$. The most relevant cases are $\delta^2/k \leq 1$, which is when $\tilde{F} < 9.98$. In all cases, pre-testing for homoskedasticity results in at least as much power as simply using the J-AR test. Indeed, for all but the lowest value of δ^2/k, the MCLR one-step pre-test is slightly more powerful. For $\delta^2/k = 1/3$, the Wald test for homoskedasticity performs poorly in simulations, which is why the the MCLR one-step pre-test and the J-AR test have nearly identical statistical power.

Figure 4 demonstrates the overall power advantage of the MCLR pre-test, relative to the MS pre-test. Instead of comparing these two procedures for different levels of identification strength (δ^2/k), we set $k = 30$ and randomly select δ^2 from a uniform distribution $U[30,45]$ for each simulation draw.[6] As Figure 4 shows, the MCLR pre-test is at least as powerful as the MS pre-test, and is strictly more powerful for certain alternative hypotheses.

5. CONCLUSION

In this article, we propose a modification of Moreira's (2003) CLR test, namely the MCLR test. We prove that in instrumental variable regression models with unknown error variance, 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 weak.

[6]We decide on this range because it ensures some cases have $\tilde{F} > 9.98$, while others have $\tilde{F} \leq 9.98$, all without creating large discontinuities in the plotted curves. Comparable ranges produce nearly identical power curves.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

Our simulations suggest 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 LM test by Hansen et al. (2008) and jackknife AR test by Mikusheva and Sun (2022).

An important direction of future research is to extend our methodology to the case of multiple endogenous regressors. Although the model with a single endogenous regressor which we consider in this article covers many relevant examples in applied research (as mentioned in Andrews and Stock (2007a) and Andrews, Stock, and Sun (2019)), models with multiple endogenous regressors are also used in applied work. To this end, our exact similarity result in Theorem 1 under the normality assumption must first be extended to the case of multiple endogenous regressors. Phillips (1980) extends the results for a single endogenous regressor presented in Sawa (1969), providing results for the exact distribution of the instrumental variable regression estimator with multiple endogenous regressors. The finite sample analysis developed by Phillips (1980) allows arbitrarily weak and many instruments; thus, a promising direction for future work would be to adapt his analytical framework to our MCLR test statistic. The next step would be to drop the normality assumption and generalize the asymptotic results in Theorems 2 and 3. The asymptotic theory developed by Phillips (1989) establishes a key invariance principle (i.e., that exact distribution theory under the assumption of normality applies to a much wider class of errors by invoking a new central limit theory involving the projection matrix on the space of the instruments) for the arbitrarily weak and many instruments setup. By observing similarities between the statements of Phillips (1989)'s Theorem 2.4 and our main theorem (Theorem 3), we expect that his general theory can be adapted to drop the normality assumption for the case of multiple endogenous regressors.

As other directions of future research, it would be interesting to extend our MCLR test to be robust for the cases of heteroskedastic errors (i.e., a many instruments robust version of Moreira and Moreira (2019)), and many included exogenous regressors as studied in Anatolyev (2013).

A. MATHEMATICAL APPENDIX

Notation: Hereafter, let $n_1 = n - k$ and $\ell = \frac{k}{n_1}$.

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 $n_1^{-1} LR_1 = \frac{b_0' Y' P_Z Y b_0}{b_0' Y' M_Z Y b_0} - \hat{\lambda}$, where $\hat{\lambda}$ is the smallest eigenvalue of $n_1^{-1} \hat{\Omega}^{-1/2} Y' P_Z Y \hat{\Omega}^{-1/2}$. The numerator of the first term can be written as

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

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

where the first equality follows from the definition of $\bar{S}$. Similarly, the denominator of the first term of $n_1^{-1}LR_1$ 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 $n_1^{-1}LR_1$ is written as $\frac{D_1}{D_4}$.

We now consider the second term of $n_1^{-1}LR_1$. Observe that $\hat{\lambda}$ is the minimum eigenvalue solution of $|\hat{\Omega}^{-1/2}Y'P_ZY\hat{\Omega}^{-1/2} - n_1\hat{\lambda}I| = 0$, or equivalently

$$|F'Y'P_ZYF - \hat{\lambda}F'Y'M_ZYF| = 0,$$

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

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

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

A.2. Proof of Proposition 2

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

A.2.2. *Proof of (ii).* 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 $[\tilde{S} : \tilde{T}] = M_ZW$, where the ith row of W is written as

$$W_i' = [V_i'b_0(b_0'\Omega b_0)^{-1/2} : V_i'\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}], \qquad \textbf{(A.1)}$$

and the ith row V_i' of V satisfies $V_i \sim \mathrm{N}(0,\Omega)$. Since

$$\bar{S} = (b_0'\Omega b_0)^{-1/2}(Z'Z)^{-1/2}Z'(v_1 - v_2\beta_0),$$
$$\bar{T} = (Z'Z)^{-1/2}Z'(Z\Pi + V)\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2},$$

we can see that $(\bar{S}',\bar{T}')$ is uncorrelated with $(\tilde{S}',\tilde{T}')$. Also since both $(\bar{S}',\bar{T}')$ and $(\tilde{S}',\tilde{T}')$ are normally distributed, we obtain independence of $(\bar{S},\bar{T})$ and $(\tilde{S},\tilde{T})$.

A.2.3. *Proof of (iii).* 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 $W = (W_1,\dots,W_n)' = [Vb_0(b_0'\Omega b_0)^{-1/2} : V\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}]$. Since M_Z is an $n\times n$ nonrandom idempotent matrix with $\mathrm{rank}(M_Z) = n_1$, it is sufficient for the conclusion to show that given $\bar{T} = t$, the rows of W are i.i.d. $\mathrm{N}(0, I_2)$.

Thus, we can see that $\mathrm{Var}(V_i'b_0(b_0'\Omega b_0)^{-1/2}) = 1$, $\mathrm{Var}(V_i'\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}) = 1$, and $\mathrm{Cov}(V_i'b_0(b_0'\Omega b_0)^{-1/2}, V_i'\Omega^{-1}a_0(a_0'\Omega^{-1}a_0)^{-1/2}) = 0$, and the conclusion follows.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

A.3. Proof of Theorem 2

We first introduce some notation. Hereafter, let μ^2 be the concentration parameter defined in Assumption 2(a). Also let

$$\mathcal{S} \sim \mathrm{N}(0, I_k), \qquad \begin{pmatrix} \mathcal{W}_1 & \mathcal{W}_2 \\ \mathcal{W}_2 & \mathcal{W}_3 \end{pmatrix} \sim \mathrm{Wishart}(n_1, I_2), \tag{A.2}$$

be drawn independently, and define

$$\Psi(t) = \psi_1(\mathcal{S}'\mathcal{S}, \mathcal{S}'t, t't, \mathcal{W}_1, \mathcal{W}_2, \mathcal{W}_3),$$

so that the critical value function for given $\bar{T} = t$ is given by $c_{1,\eta}(t)$, the $(1-\eta)$th quantile of $\Psi(t)$. To analyze $\Psi(\bar{T})$, we standardize its arguments as

$$\begin{aligned} \mathcal{Z}_1 &= \frac{\mathcal{S}'\mathcal{S} - k}{\sqrt{k}}, \qquad \mathcal{Z}_2 = \frac{\mathcal{S}'\bar{T}}{\sqrt{k}}, \qquad \mathcal{Z}_{\bar{T}} = \frac{\bar{T}'\bar{T} - k - \mu^2}{\sqrt{k}}, \\ \mathcal{Q}_1 &= \frac{\mathcal{W}_1 - n_1}{\sqrt{n_1}}, \qquad \mathcal{Q}_2 = \frac{\mathcal{W}_2}{\sqrt{n_1}}, \qquad \mathcal{Q}_3 = \frac{\mathcal{W}_3 - n_1}{\sqrt{n_1}}, \end{aligned} \tag{A.3}$$

where μ^2 is defined in (3). For the proof of Theorem 2, we use the following lemma.

LEMMA 1. *Under Assumptions 1 and 2, it holds*

$$\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_{\bar{T}}, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3 = O_p(1),$$

$$\frac{\bar{S}'\bar{S}}{k}, \frac{\tilde{S}'\tilde{S}}{n_1}, \frac{\tilde{T}'\tilde{T}}{n_1} \xrightarrow{p} 1, \qquad \frac{\mathcal{S}'\bar{S}}{\sqrt{k}}, \frac{\bar{S}'\bar{T}}{\sqrt{k}}, \frac{\tilde{S}'\tilde{T}}{\sqrt{n_1}} = O_p(1).$$

Proof of Lemma 1. All the statements are obtained by Markov's inequality using the definitions in (A.2) and the fact that $\bar{S}, \tilde{S}, \bar{T}, \tilde{T}$ are standardized normal vectors. □

A.3.1. *Proof Under Assumption 2(a).* For the conclusion in (12), it is sufficient to show that

$$n_1 \frac{\mu^2}{k} \Psi(\bar{T}) \text{ converges to some non-degenerate distribution,} \tag{A.4}$$

$$n_1 \frac{\mu^2}{k} \{\Psi(\hat{T}) - \Psi(\bar{T})\} \xrightarrow{p} 0, \tag{A.5}$$

by utilizing $\frac{\mu^2}{\sqrt{k}} \to \infty$ in Assumption 2 (a).

For (A.4), Lemma 2 implies

$$\begin{aligned} n_1 \frac{\mu^2}{k} \Psi(\bar{T}) &= \ell \mathcal{Q}_2^2 + \frac{\sqrt{k}}{n_1} \frac{\mu^2}{k} \mathcal{Z}_1 \mathcal{Q}_1^2 - 2\sqrt{\ell} \mathcal{Z}_2 \mathcal{Q}_2 + \mathcal{Z}_2^2 + o_p(1) \\ &= \left(\mathcal{Z}_2 - \sqrt{\frac{\alpha}{1-\alpha}} \mathcal{Q}_2 \right)^2 + o_p(1), \end{aligned} \tag{A.6}$$

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 (A.4).

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

For (A.5), we need to ask what is the effect of using feasible $\hat{T}$ instead of $\bar{T}$. If we replace $\bar{T}$ with $\hat{T}$, only the terms $\{\mathcal{Z}_2, \mathcal{Z}_{\bar{T}}\}$ need to be replaced with $\{\hat{\mathcal{Z}}_2, \mathcal{Z}_{\hat{T}}\}$, where $\hat{\mathcal{Z}}_2 = \frac{\mathcal{S}'\hat{T}}{\sqrt{k}}$ and $\mathcal{Z}_{\hat{T}} = \frac{\hat{T}'\hat{T}-k-\mu^2}{\sqrt{k}}$. By repeating the same argument in the proof of Lemma 2, we can see that the (deterministic) coefficients on $\mathcal{Z}_{\hat{T}}$ will be zero. Therefore, similar to (A.6), we have

$$n_1 \frac{\mu^2}{k} \Psi(\hat{T}) = \left(\hat{\mathcal{Z}}_2 - \sqrt{\frac{\alpha}{1-\alpha}} \mathcal{Q}_2 \right)^2 + o_p(1). \tag{A.7}$$

Thus it is sufficient for (A.5) to show that

$$\hat{\mathcal{Z}}_2 = \mathcal{Z}_2 + o_p(1). \tag{A.8}$$

By using the definitions of $\hat{T}$ and $\hat{\Omega}$ and the relation

$$(Y'M_Z Y)^{-1} = F \begin{pmatrix} \tilde{S}'\tilde{S} & \tilde{S}'\tilde{T} \\ \tilde{T}'\tilde{S} & \tilde{T}'\tilde{T} \end{pmatrix}^{-1} F', \tag{A.9}$$

with $F = [b_0(b_0'\Omega b_0)^{-1/2} : \Omega^{-1} a_0 (a_0'\Omega^{-1} a_0)^{-1/2}]$, direct calculations yield

$$\begin{aligned} \hat{\mathcal{Z}}_2 &= \ell^{-1/2} \mathcal{S}'(Z'Z)^{-1/2} Z'Y(Y'M_Z Y)^{-1} a_0 (a_0'(Y'M_Z Y)^{-1} a_0)^{-1/2} \\ &= \frac{-(\mathcal{S}'\bar{S})(\tilde{S}'\tilde{T})(\tilde{S}'\tilde{S})^{-1/2} + (\mathcal{S}'\bar{T})(\tilde{S}'\tilde{S})^{1/2}}{\sqrt{\ell}\sqrt{(\tilde{S}'\tilde{S})(\tilde{T}'\tilde{T}) - (\tilde{S}'\tilde{T})^2}} \\ &= \frac{\mathcal{S}'\bar{T}}{\sqrt{k}} + o_p(1) = \mathcal{Z}_2 + o_p(1), \end{aligned}$$

where the third equality follows from Lemma 1. Therefore, we obtain (A.8) which implies (A.5). Since (A.4) and (A.5) are satisfied, the conclusion follows.

LEMMA 2. *Recall the definitions in (A.3). Under Assumptions 1 and 2, it holds*

$$\Psi(\bar{T}) = \frac{k}{n_1^2}\left(\frac{\mu^2}{k}\right)^{-1} \mathcal{Q}_2^2 + \frac{1}{n_1}\left(\frac{\mu^2}{k}\right)^{-1} \mathcal{Z}_2^2 + \frac{\sqrt{k}}{n_1^2} \mathcal{Z}_1 \mathcal{Q}_1^2 - \frac{2\sqrt{k}}{n_1^{3/2}}\left(\frac{\mu^2}{k}\right)^{-1} \mathcal{Z}_2 \mathcal{Q}_2 + o_p(n^{-1}).$$

Proof of Lemma 2. By explicitly computing the smallest eigenvalue in $\Psi(t)$, $\Psi(\bar{T})$ can be written as

$$\Psi(\bar{T}) = \frac{\mathcal{S}'\mathcal{S}}{\mathcal{W}_1} + \frac{b + \sqrt{b^2 - 4ac}}{2a}, \tag{A.10}$$

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

where the terms a, b, and c can be written as

$$a = \frac{1}{n_1^2}(\mathcal{W}_1\mathcal{W}_3 - \mathcal{W}_2^2) = 1 + \frac{\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n_1}} + \frac{\mathcal{Q}_1\mathcal{Q}_3 - \mathcal{Q}_2^2}{n_1},$$

$$b = \frac{1}{n_1^2}\{-\mathcal{W}_1(\bar{T}'\bar{T}) - (\mathcal{S}'\mathcal{S})\mathcal{W}_3 + 2(\mathcal{S}'\bar{T})\mathcal{W}_2\}$$

$$= -\ell\left\{\left(2 + \frac{\mu^2}{k}\right) + \frac{\left(1 + \frac{\mu^2}{k}\right)\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n_1}} + \frac{\mathcal{Z}_1 + \mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\mathcal{Z}_{\bar{T}}\mathcal{Q}_1 + \mathcal{Z}_1\mathcal{Q}_3}{\sqrt{n_1}\sqrt{k}} - \frac{2\mathcal{Z}_2\mathcal{Q}_2}{\sqrt{\ell n_1}}\right\},$$

$$c = \frac{1}{n_1^2}\{(\mathcal{S}'\mathcal{S})(\bar{T}'\bar{T}) - (\mathcal{S}'\bar{T})^2\}$$

$$= \ell^2\left\{\left(1 + \frac{\mu^2}{k}\right) + \frac{\left(1 + \frac{\mu^2}{k}\right)\mathcal{Z}_1 + \mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\mathcal{Z}_1\mathcal{Z}_{\bar{T}} - \mathcal{Z}_2^2}{k}\right\}.$$

By (A.3) and a Taylor expansion, the first term of (2) is written as

$$\frac{\mathcal{S}'\mathcal{S}}{\mathcal{W}_1} = \ell\left(1 + \frac{\mathcal{Z}_1}{\sqrt{k}}\right)\left(1 - \frac{\mathcal{Q}_1}{\sqrt{n_1}} + \frac{\mathcal{Q}_1^2}{n_1}\right) + o_p(n^{-1}). \quad \textbf{(A.11)}$$

Based on these expressions and by using $\frac{\mu^2}{\sqrt{k}} \to \infty$ (Assumption 2(a)), which guarantees that the term B_1 dominates B_2 (B_1 and B_2 are defined below), we can expand the second term of (A.10) as follows. □

First, by lengthy but straightforward calculations and ignoring the terms of lower orders, we have

$$b^2 - 4ac = 4B_0^2(1 + B_1 + B_2) + o_p(n^{-1}), \quad \textbf{(A.12)}$$

where $B_0 = \frac{\ell}{2}\frac{\mu^2}{k}$ and

$$B_1 = \left(\frac{\mu^2}{k}\right)^{-1}\left\{-\frac{2\mathcal{Z}_1}{\sqrt{k}} + \frac{2\mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \left(1 + \frac{\mu^2}{K}\right)\frac{2\mathcal{Q}_1}{\sqrt{n_1}} - \frac{2\mathcal{Q}_3}{\sqrt{n_1}}\right\},$$

$$B_2 = \left(\frac{\mu^2}{k}\right)^{-2}\left[\frac{\mathcal{Z}_1^2}{k} + \frac{\mathcal{Z}_{\bar{T}}^2}{k} + \left(1 + \frac{\mu^2}{k}\right)^2\frac{\mathcal{Q}_1^2}{n_1} + \frac{\mathcal{Q}_3^2}{n_1}\right.$$
$$-2\frac{\mathcal{Z}_1\mathcal{Z}_{\bar{T}}}{k} - 2\left(1 + \frac{\mu^2}{k}\right)\frac{\mathcal{Z}_1\mathcal{Q}_1}{\sqrt{k}\sqrt{n_1}} + 2\left(1 - \frac{\mu^2}{k}\right)\frac{\mathcal{Z}_1\mathcal{Q}_3}{\sqrt{k}\sqrt{n_1}}$$
$$+\left(2 + \frac{4\mu^2}{k}\right)\frac{\mathcal{Z}_{\bar{T}}\mathcal{Q}_1}{\sqrt{k}\sqrt{n_1}} - 2\frac{\mathcal{Z}_{\bar{T}}\mathcal{Q}_3}{\sqrt{k}\sqrt{n_1}} - 2\left(1 + \frac{\mu^2}{k}\right)\frac{\mathcal{Q}_1\mathcal{Q}_3}{\sqrt{n_1}\sqrt{n_1}}$$
$$\left.+4\frac{\mathcal{Z}_2^2}{k} - 4\left(2 + \frac{\mu^2}{k}\right)\frac{\mathcal{Z}_2\mathcal{Q}_2}{\sqrt{k}\sqrt{n_1}} + 4\left(1 + \frac{\mu^2}{k}\right)\frac{\mathcal{Q}_2^2}{n_1}\right].$$

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

Also express a and b as

$$a = 1 + C_1 + C_2, \qquad b = -2(A_0 + A_1 + A_2), \tag{A.13}$$

where

$$C_1 = \frac{\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n_1}}, \qquad C_2 = \frac{\mathcal{Q}_1\mathcal{Q}_3 - \mathcal{Q}_2^2}{n_1},$$

$$A_0 = \frac{\ell}{2}\left(2 + \frac{\mu^2}{k}\right), \qquad A_1 = \frac{\ell}{2}\left\{\frac{\left(1 + \frac{\mu^2}{k}\right)\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n_1}} + \frac{\mathcal{Z}_1 + \mathcal{Z}_{\bar{T}}}{\sqrt{k}}\right\},$$

$$A_2 = \frac{\ell}{2}\left(\frac{\mathcal{Z}_{\bar{T}}\mathcal{Q}_1 + \mathcal{Z}_1\mathcal{Q}_3}{\sqrt{n_1}\sqrt{k}} - \frac{2\mathcal{Z}_2\mathcal{Q}_2}{\sqrt{\ell n_1}}\right).$$

Second, by (A.4) and (A.5) combine with expansions $\sqrt{1+z} = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + o(z^2)$ and $\frac{1}{z} = 1 - (z-1) + (z-1)^2 + o(z^2)$, the second term of (A.10) can be expanded as

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

Finally, by inserting the definitions of $(A_0, A_1, A_2, B_0, B_1, B_2, C_1, C_2)$ to the above display and ignoring the terms of order $o_p(n^{-1})$ by Lemma 1, lengthy but straightforward calculations yield the conclusion of this lemma.

A.3.2. *Technical Remark on Assumption 2(a).* In this subsection, we clarify the point in the remark of Assumption 2(a): *If* $\frac{\mu^2}{\sqrt{k}} = O(1)$, *the result in (A.4) will be satisfied with a different normalization. However, under this normalization, the result in (A.5) is typically violated*. More precisely, in the case of $\frac{\mu^2}{\sqrt{k}} = O(1)$, we show that

$$\sqrt{k}\Psi(\bar{T}) \text{ converges to some non-degenerate distribution,} \tag{A.14}$$

but

$$\sqrt{k}\{\Psi(\hat{T}) - \Psi(\bar{T})\} \overset{p}{\nrightarrow} 0, \tag{A.15}$$

when $\frac{k}{n} \to \alpha \in (0,1)$.

To see this, under $\frac{\mu^2}{\sqrt{k}} = O(1)$, proceed as in the proof of Lemma 2, the second term of $\Psi(\bar{T})$ in (A.3) is expanded as

$$\frac{b + \sqrt{b^2 - 4ac}}{2a} = -(\ell + A^* + D^*)(1 - C_1) + O_p(n^{-1}), \tag{A.16}$$

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

where C_1 is defined in (A.5) and

$$A^* = \frac{\ell}{2}\left\{\frac{\mu^2}{k} + \frac{\mathcal{Q}_1 + \mathcal{Q}_3}{\sqrt{n_1}} + \frac{\mathcal{Z}_1 + \mathcal{Z}_{\bar{T}}}{\sqrt{k}}\right\},$$

$$D^* = -\frac{\ell}{2}\sqrt{\left(\frac{\mu^2}{k} - \frac{\mathcal{Z}_1 - \mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\mathcal{Q}_1 - \mathcal{Q}_3}{\sqrt{n_1}}\right)^2 + 4\left(\frac{\mathcal{Z}_2}{\sqrt{k}} - \frac{\mathcal{Q}_2}{\sqrt{n_1}}\right)^2}.$$

By (A.11) and (A.16), an expansion of $\Psi(\bar{T})$ is obtained as

$$\Psi(\bar{T}) = \frac{\ell}{2}\left(\frac{\mathcal{Z}_1 - \mathcal{Z}_{\bar{T}}}{\sqrt{k}} - \frac{\mu^2}{k} - \frac{\mathcal{Q}_1 - \mathcal{Q}_3}{\sqrt{n_1}}\right) + \frac{\ell}{2}\sqrt{\left(\frac{\mu^2}{k} - \frac{\mathcal{Z}_1 - \mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\mathcal{Q}_1 - \mathcal{Q}_3}{\sqrt{n_1}}\right)^2 + 4\left(\frac{\mathcal{Z}_2}{\sqrt{k}} - \frac{\mathcal{Q}_2}{\sqrt{n_1}}\right)^2} + O_p(k^{-1}).$$

So, by Lemma 1, we can see that $\sqrt{k}\Psi(\bar{T})$ converges to some non-degenerate distribution, but the coefficient on $\mathcal{Z}_{\bar{T}}$ is nonzero unlike in the case of $\frac{\mu^2}{\sqrt{k}} \to \infty$.

On the other hand, by using the definitions of $\hat{T}$ and $\hat{\Omega}$ and the relation in (A.9), direct calculations yield

$$\begin{aligned}\frac{\hat{T}'\hat{T}}{k} &= \ell^{-1}(a_0'(Y'M_ZY)^{-1}a_0)^{-1/2}a_0'(Y'M_ZY)^{-1}Y'P_ZY(Y'M_ZY)^{-1}a_0(a_0'(Y'M_ZY)^{-1}a_0)^{-1/2}\\ &= \frac{(\bar{S}'\bar{S})(\tilde{S}'\bar{T})^2(\tilde{S}'\tilde{S})^{-1} - 2(\bar{S}'\bar{T})(\tilde{S}'\tilde{T}) + (\bar{T}'\bar{T})(\tilde{S}'\tilde{S})}{\ell\{(\tilde{S}'\tilde{S})(\tilde{T}'\tilde{T}) - (\tilde{S}'\tilde{T})^2\}}\\ &= \left(1 + \frac{\mu^2}{k}\right) + \left(\frac{\bar{T}'\bar{T} - \mu^2 - k}{k}\right) - \left(1 + \frac{\mu^2}{k}\right)\left(\frac{\tilde{T}'\tilde{T} - n_1}{n_1}\right) + O_p(k^{-1}),\end{aligned}$$

where the third equality follows from Lemma 1. Thus, an expansion of $\mathcal{Z}_{\hat{T}} = \frac{\hat{T}'\hat{T} - \mu^2 - k}{\sqrt{k}}$ is obtained as

$$\begin{aligned}\mathcal{Z}_{\hat{T}} &= \sqrt{k}\left\{\left(\frac{\bar{T}'\bar{T} - \mu^2 - k}{k}\right) - \left(1 + \frac{\mu^2}{k}\right)\left(\frac{\tilde{T}'\tilde{T} - n_1}{n_1}\right)\right\} + o_p(1)\\ &= \mathcal{Z}_{\bar{T}} - \sqrt{\ell}\mathcal{Q}_3^* + o_p(1),\end{aligned} \qquad \textbf{(A.17)}$$

where $\mathcal{Q}_3^* = \frac{\tilde{T}'\tilde{T} - n_1}{\sqrt{n_1}} = O_p(1)$. Note that if $\frac{k}{n} \to 0$ as in Assumption 2(b), then $\ell = \frac{k}{n-k} \to 0$ and we can guarantee $\mathcal{Z}_{\hat{T}} = \mathcal{Z}_{\bar{T}} + o_p(1)$. However, under Assumption 2(a), $\mathcal{Z}_{\hat{T}}$ and $\mathcal{Z}_{\bar{T}}$ are not asymptotically equivalent due to the additional term $\mathcal{Q}_3^*$. By inserting this, an expansion of $\Psi(\hat{T})$ is obtained as

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

$$\Psi(\hat{T}) = \frac{\ell}{2}\left(-\frac{\mathcal{Z}_{\bar{T}}}{\sqrt{k}} - \frac{\mu^2}{k} - \frac{\mathcal{Q}_1}{\sqrt{n_1}} + \frac{\mathcal{Q}_3 + \mathcal{Q}_3^*}{\sqrt{n_1}} + \frac{\mathcal{Z}_1}{\sqrt{k}}\right)$$
$$+\frac{\ell}{2}\sqrt{\left(\frac{\mathcal{Z}_{\bar{T}}}{\sqrt{k}} + \frac{\mu^2}{k} + \frac{\mathcal{Q}_1}{\sqrt{n_1}} - \frac{\mathcal{Q}_3 + \mathcal{Q}_3^*}{\sqrt{n_1}} - \frac{\mathcal{Z}_1}{\sqrt{k}}\right)^2 + 4\left(\frac{\mathcal{Z}_2}{\sqrt{k}} - \frac{\mathcal{Q}_2}{\sqrt{n_1}}\right)^2} + O_p(k^{-1}),$$

and thus $\sqrt{k}\{\Psi(\hat{T}) - \Psi(\bar{T})\} \overset{p}{\nrightarrow} 0$ due to the terms involving $\mathcal{Q}_3^*$.

A.3.3. *Proof Under Assumption 2(b).* Recall $\hat{\mathcal{Z}}_2 = \frac{S'\hat{T}}{\sqrt{k}}$ and $\mathcal{Z}_{\hat{T}} = \frac{\hat{T}'\hat{T} - k - \mu^2}{\sqrt{k}}$, where $\hat{T}$ is defined in (10). By an analogous argument in the proof of Moreira (2003, Theorem 2), we can see that the conclusion under Assumption 2(b) follows by:

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

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

For (A.18), we can apply the same argument as the proof of (A.8) (since it does not use the condition on μ^2). For (A.19), by using the definitions of $\hat{T}$ and $\hat{\Omega}$ and the relation in (A.9), direct calculations yield

$$\begin{aligned}
\frac{\hat{T}'\hat{T}}{k} &= n_1(a_0'(Y'M_ZY)^{-1}a_0)^{-1/2}a_0'(Y'M_ZY)^{-1}Y'P_ZY(Y'M_ZY)^{-1}a_0(a_0'(Y'M_ZY)^{-1}a_0)^{-1/2} \\
&= \ell^{-1}\{(\tilde{S}'\tilde{S})(\tilde{T}'\tilde{T}) - (\tilde{S}'\tilde{T})^2\}^{-1}\left\{(\bar{S}'\bar{S})(\tilde{S}'\tilde{T})^2(\tilde{S}'\tilde{S})^{-1} - 2(\bar{S}'\bar{T})(\tilde{S}'\tilde{T}) + (\bar{T}'\bar{T})(\tilde{S}'\tilde{S})\right\} \\
&= \left(1 + \frac{\mu^2}{k}\right) + \frac{\bar{T}'\bar{T} - \mu^2 - k}{k} + \left(1 + \frac{\mu^2}{k}\right)\frac{\tilde{S}'\tilde{S} - n_1}{n_1} + O_p(n^{-1}),
\end{aligned}$$

where the third equality follows from Lemma 1. Therefore, (A.19) is verified as

$$\mathcal{Z}_{\hat{T}} = \mathcal{Z}_{\bar{T}} + \sqrt{\ell}\left(\frac{\mu^2}{k} + 1\right)\frac{\tilde{S}'\tilde{S} - n_1}{\sqrt{n_1}} + o_p(1) = \mathcal{Z}_{\bar{T}} + o_p(1), \tag{A.20}$$

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

A.4. Proof of Theorem 3

Under $k/n \to 0$ (Assumption 2(b)), $\left(\frac{\tilde{S}'\tilde{S} - n_1}{\sqrt{n_1}}, \frac{\tilde{S}'\tilde{T}}{\sqrt{n_1}}, \frac{\tilde{T}'\tilde{T} - n_1}{\sqrt{n_1}}\right)$ are of smaller order than $(\bar{\mathcal{Z}}_1, \bar{\mathcal{Z}}_2, \mathcal{Z}_{\bar{T}}) := \left(\frac{\bar{S}'\bar{S} - k}{\sqrt{k}}, \frac{\bar{S}'\bar{T}}{\sqrt{k}}, \frac{\bar{T}'\bar{T} - k - \mu^2}{\sqrt{k}}\right)$. A central limit theorem under the fourth moment assumption on V yields the asymptotic normality of $(\bar{\mathcal{Z}}_1, \bar{\mathcal{Z}}_2, \mathcal{Z}_{\bar{T}})$ with the limiting variance

$$\mathrm{Var}(\bar{\mathcal{Z}}_1, \bar{\mathcal{Z}}_2, \mathcal{Z}_{\bar{T}}) \to \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \tag{A.21}$$

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press

To see (A.21), let $u_i = Y_i'b_0$, $w_i = Y_i'\Omega^{-1}a_0$, $\sigma_u^2 = \mathrm{Var}(u_i)$, $\kappa_u = E[u_i^4]$, and P_{ij} be the (i,j)th element of P_Z. We have

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

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 $\sigma_w^2 = \mathrm{Var}(w_i)$, we have

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

where the second equality follows from $E[u_iw_i]=0$. The limits of the other elements can be shown in the same manner, so we obtain (A.21).

Therefore, since the limiting distribution of $(\bar{\mathcal{Z}}_1,\bar{\mathcal{Z}}_2,\mathcal{Z}_{\bar{T}})$ is identical to $(\mathcal{Z}_1,\mathcal{Z}_2,\mathcal{Z}_{\bar{T}})$ for the Gaussian case, we obtain

$$\Pr\left\{\frac{LR_1}{n_1}\geq c_{1,\eta}(\bar{T})\right\}\to\eta.$$

Finally, by repeating the same argument for the proof of Theorem 2 (under Assumption 2(b)) with the fourth moment assumption on V, we obtain the conclusion.

REFERENCES

Anatolyev, S. (2013). Instrumental variables estimation and inference in the presence of many exogenous regressors. *Econometrics Journal* 16(1), 27–72.

Anderson, T. W., Kunitomo, N., & Matsushita, Y. (2010). On the asymptotic optimality of the LIML estimator with possibly many instruments. *Journal of Econometrics* 157(2), 191–204.

Andrews, D. W. K., Moreira, M. J., & Stock, J. H. (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74(3), 715–752.

Andrews, D. W. K., & Stock, J. H. (2007a). Inference with weak instruments. In R. Blundell, W. Newey, & T. Persson (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress (pp. 122–173). Econometric Society Monographs, 3, Cambridge University Press. https://doi.org/10.1017/CBO9780511607547.007.

Andrews, D. W. K., & Stock, J. H. (2007b). Testing with many weak instruments. *Journal of Econometrics* 138(1), 24–46.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics* 11, 727–753.

Cattaneo, M., Jansson, M., & Ma, X. (2019). Two-step estimation and inference with possibly many included covariates. *Review of Economic Studies* 86, 1095–1122.

Chao, J., & Swanson, N. (2005). Consistent estimation with a large number of weak instruments. *Econometrica* 73(5), 1673–1692.

Crudu, F., Mellace, G., & Sándor, Z. (2021). Inference in instrumental variable models with heteroskedasticity and many instruments. *Econometric Theory* 37(2), 281–310.

Han, C., & Phillips, P. C. B. (2006). GMM with many moment conditions. *Econometrica* 74(1), 147–192.

Hansen, C., Hausman, J. A., & Newey, W. K. (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics* 26(4), 398–422.

Hausman, J. A. et al. (2012). Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3(2), 211–255.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70(5), 1781–1803.

Matsushita, Y., and Otsu, T. (2024). A jackknife Lagrange multiplier test with many weak instruments. *Econometric Theory* 40(2), 447–470.

Mikusheva, A., & Sun, L. (2022). Inference with many weak instruments. *Review of Economic Studies* 89(5), 2663–2686.

Mills, B., Moreira, M. J., & Vilela, L. (2014). Tests based on t-statistics for IV regression with weak instruments. *Journal of Econometrics* 182(2), 351–363.

Moreira, H., & Moreira, M. J. (2019). Optimal two-sided tests for instrumental variables regression with heteroskedastic and autocorrelated errors. *Journal of Econometrics* 213(2), 398–433.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica* 71(4), 1027–1048.

Newey, W. K., & Windmeijer, F. (2009). Generalized method of moments with many weak moment conditions. *Econometrica* 77(3), 687–719.

Phillips, P. C. B. (1980). The exact distribution of instrumental variable estimators in an equation containing $n+1$endogenous variables. *Econometrica* 48(4), 861–878.

Phillips, P. C. B (1989). Partially identified econometric models. *Econometric Theory* 5(2), 181–240.

Sawa, T. (1969). The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *Journal of the American Statistical association* 64(327), 923–937.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak Instruments. *Econometrica* 65(3), 557–586.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). Cambridge University Press.

https://doi.org/10.1017/S0266466625000088 Published online by Cambridge University Press