Inference with Weak Instruments

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

James H. Stock
Department of Economics
Harvard University

March 2005
Revised: September 2005

[1]This paper was prepared for an invited session at the 2005 Econometric Society World Congress in London. The authors thank Richard Smith for comments and the conference organizers for their organizational efforts. Andrews and Stock gratefully acknowledge the research support of the National Science Foundation via grant numbers SES-0001706 and SES-0417911 and SBR-0214131, respectively.

Abstract

This paper reviews recent developments in methods for dealing with weak instruments (IVs) in IV regression models. The focus is more on tests (and confidence intervals derived from tests) than estimators.

The paper also presents new testing results under "many weak IV asymptotics," which are relevant when the number of IVs is large and the coefficients on the IVs are relatively small. Asymptotic power envelopes for invariant tests are established. Power comparisons of the conditional likelihood ratio (CLR), Anderson-Rubin, and Lagrange multiplier tests are made. Numerical results show that the CLR test is on the asymptotic power envelope. This holds no matter what the relative magnitude of the IV strength to the number of IVs.

Keywords: Conditional likelihood ratio test, instrumental variables, many instrumental variables, power envelope, weak instruments.

JEL Classification Numbers: C12, C30.

1 Introduction

The standard approach to reporting empirical results in economics is to provide point estimates and standard errors. With this information, confidence intervals (CIs) and tests are constructed using t tests and the normal critical values. In instrumental variables (IVs) regression with weak IVs, this approach is problematic. The most widely used estimator, two-stage least squares (2SLS), has significance bias and is poorly approximated by a normal distribution when IVs are weak and the degree of endogeneity is medium to strong. In fact, when the parameter space allows for arbitrarily weak IVs the exact finite-sample level of standard CIs of the form "estimate $\pm$ std error $\times$ constant" is zero and t tests have level one.

This paper reviews recent results in the literature on weak IVs that develop an alternative approach to inference with weak IVs. With this approach, one reports point estimates accompanied by CIs or tests that have levels that are robust to the strength of the IVs. In particular, CIs are formed by inverting tests that are robust to weak IVs. That is, a CI for a parameter β, say, is the set of points β_0 for which a weak IV robust test fails to reject the null hypothesis $H_0 : \beta = \beta_0$. This is the same method that is used to generate a CI of the form "estimate $\pm$ std error $\times$ constant" except that the test employed is one whose level is robust to weak IVs, rather than a t tests based on a normal approximation.

The paper focuses on the linear IV model with a single right-hand side (rhs) endogenous variable and independent identically distributed (iid) homoskedastic normal errors. The majority of applications involve a single rhs endogenous variable. Interest usually is focussed on the coefficient on this variable. Although this basic model is relatively simple, it captures the essence of the problem. Tests whose levels are robust to weak IVs for this model can be extended to the case of non-normal errors, heteroskedastic and/or autocorrelated errors, multiple rhs endogenous variables, and nonlinear moment condition models. The paper discusses these extensions. Not all of them are completely satisfactory.

For a just-identified model, the Anderson-Rubin (AR) test, or a heteroskedasticity and/or autocorrelation robust version of it, is the preferred test because its level is robust to weak IVs and its power properties are quite good—optimal in certain respects. For over-identified models, the AR test still is robust to weak IVs, but its power properties are not as good because it effectively ignores parameter restrictions that arise naturally in the model. The literature has sought tests that are robust to weak IVs and are more powerful than the AR test in over-identified models.

Alternatives to the AR test that have been considered include an LM test and a conditional likelihood ratio (CLR) test. Both of these tests are robust to weak IVs. The power properties of the CLR test have been found to dominate those of the LM and AR tests (with iid homoskedastic normal errors). In fact, the CLR test is found to be essentially on a power envelope for two-sided invariant similar tests. Furthermore, there is no cost in terms of performance of the test when the IVs are strong—the test is asymptotically efficient under standard strong IV asymptotics.

"Conditioning" methods have been developed that can be used to convert t tests, such as the usual one based on the 2SLS estimator, into tests whose levels are robust to weak IVs. The power properties of such tests, however, are found to be distinctly inferior to those of the CLR test. Hence, the CLR test outperforms standard t tests both in terms of level and in terms of level–corrected power under weak IVs and is asymptotically efficient under strong IVs.

Although the CLR test is robust asymptotically to weak IVs and non-normality of the errors, it is not robust to heteroskedasticity and/or autocorrelation of the errors or to left-out IVs (from the reduced form equation for the rhs endogenous variables). Nevertheless, versions of the CLR test that are robust to these features have been developed. We recommend such tests and the CIs they generate for general use in IV regression with potentially weak IVs. Furthermore, generalizations of these CLR tests to moment condition models, which are typically estimated by generalized method of moments (GMM), also are available. For moment condition models, we recommend such tests because they are robust to weak IVs and can be expected to have relatively good power properties.

In addition to reviewing some of the recent literature on inference with weak IVs, this paper presents new results for testing under "many weak IV asymptotics". Such asymptotics are designed for the case in which the IVs are weak and the number of IVs, k, is relatively large compared to the sample size n. We find that in this set-up the CLR test is still completely robust asymptotically to weak IVs and is essentially on the power envelope for two-sided invariant (similar or nonsimilar) tests. This holds no matter how one specifies the relative magnitude of the strength of the IVs to k in the asymptotics. Hence, the optimal power properties of the CLR test are quite robust to k. The AR and LM tests have power that lies off the power envelope—in some cases by a considerable extent. On the other hand, the level of the CLR test is not completely robust to the magnitude of k relative to n. One does not want to take k too large relative to n. With normal errors, we show that the CLR test has correct size asymptotically provided $k^{3/2}/n \to 0$ as $n \to \infty$. With non-normal errors, Andrews and Stock (2005) show that the same is true if $k^3/n \to 0$ as $n \to \infty$.

We conclude that the "many weak IV" results for the CLR test buttress the argument for employing this test (or heteroskedasticity or autocorrelation robust versions of it) in scenarios with potentially weak IVs.

This paper focuses on hypothesis tests and CIs that are robust to weak IVs, and pays less attention to two other aspects of the weak IV problem. The first neglected topic concerns pretesting for weak instruments: if the instruments are weak, one adopts a robust strategy, but if the instruments are strong, one uses 2SLS. This approach is now common empirical practice and is an improvement over the practice of a decade ago, in which 2SLS was always used without thought about the strength of the instruments. But this approach entails standard concerns about pretests, and as a result we find fully robust tests and CIs more appealing. The second neglected aspect is point estimation. Despite a great deal of work in the finite sample and Edgeworth expansion literatures, there are few sharp results concerning point estimates. Although it is generally found that 2SLS has particularly poor finite sample behavior, each alternative estimator seems to have its own pathologies when instruments are weak. We therefore have focused on testing and CIs weak IVs, for which a solution is closer at hand than it is for estimation.

The paper does not discuss the recent literature on IV estimation of treatment effect model in which the treatment effects depend on unobserved individual specific variables, e.g., see Imbens and Angrist (1994) and Heckman and Vytlacil (2005).

The remainder of this paper is organized as follows. Section 2 introduces the model considered in much of the paper. Section 3 discusses what is meant by weak IVs and the problems with 2SLS estimators, tests, and CIs under weak IVs. Section 4 describes "weak IV asymptotics," "many IV asymptotics," and "many weak IV asymptotics." Section 5 covers formal methods for detecting weak IVs. Section 6 discusses the two approaches to tests and CIs mentioned above—t test-based CIs versus CIs obtained by inverting weak IV robust tests. Section 7 describes recent developments for tests whose levels are robust to weak IVs. This includes similar tests via conditioning, optimal power, robustness to heteroskedasticity and/or autocorrelation, power with non-normal errors, and extensions to multiple rhs endogenous variables, coefficients on exogenous variables, and moment condition models. Section 8 briefly outlines some recent developments for estimation with weak IVs. Section 9 does likewise for estimation with many weak IVs. Section 10 presents the new results for testing with many weak IVs.

We note that recent survey papers on weak IVs include Stock, Wright, and Yogo (2002), Dufour (2003), and Hahn and Hausman (2003b).

2 Model

We start by defining the model that we focus on for much of the paper. The model is an IV regression model with one endogenous right-hand side (rhs) variable, multiple exogenous variables, multiple IVs, and independent identically distributed (iid) homoskedastic normal errors. The exogenous variables and IVs are treated as fixed (i.e., nonrandom).

The reasons for considering this special model are the following. First, the case of a single rhs endogenous variable is by far the most important in empirical applications. Second, asymptotic results for non-normal errors with random or fixed exogenous variables and IVs are analogous to the finite sample results for normal errors with fixed exogenous variables and IVs. Third, results for heteroskedastic and/or autocorrelated errors can be obtained by extending the results for iid homoskedastic errors. Below we discuss these extensions.

The model consists of a structural equation and a reduced-form equation:

$$\begin{aligned} y_1 &= y_2\beta + X\gamma_1 + u, \\ y_2 &= Z\pi + X\xi + v_2, \end{aligned} \tag{2.1}$$

where $y_1, y_2 \in R^n$, $X \in R^{n\times p}$, and $Z \in R^{n\times k}$ are observed variables; $u, v_2 \in R^n$ are unobserved errors; and $\beta \in R$, $\pi \in R^k$, $\gamma_1 \in R^p$, and $\xi \in R^p$ are unknown parameters. We assume that Z is the matrix of residuals from the regression of some underlying IVs, say $\widetilde{Z} \in R^{n\times k}$, on X (i.e., $Z = M_X\widetilde{Z}$, where $M_X = I_n - P_X$ and $P_X = X(X'X)^{-1}X'$). Hence, $Z'X = 0$. The exogenous variable matrix X and the

IV matrix Z are fixed (i.e., non-stochastic) and $[X : Z]$ has full column rank $p + k$. The $n \times 2$ matrix of errors $[u : v_2]$ is iid across rows with each row having a mean zero bivariate normal distribution.

The variable y_2 is endogenous in the equation for y_1 (i.e., y_2 and u may be correlated). Endogeneity may be due to simultaneity, left-out variables, or mismeasurement of an exogenous variable. Although we refer to the equation for y_1 as a structural equation, only in the case of simultaneity is the equation for y_1 really a structural equation.

The two reduced-form equations are

$$\begin{aligned} y_1 &= Z\pi\beta + X\gamma + v_1 \\ y_2 &= Z\pi + X\xi + v_2, \text{ where} \\ \gamma &= \gamma_1 + \xi\beta \text{ and } v_1 = u + v_2\beta. \end{aligned} \tag{2.2}$$

The reduced-form errors $[v_1 : v_2]$ are iid across rows with each row having a mean zero bivariate normal distribution with 2×2 nonsingular covariance matrix Ω. The parameter space for $\theta = (\beta, \pi', \gamma', \xi')'$ is taken to be $R \times R^k \times R^p \times R^p$. Let $Y = [y_1 : y_2] \in R^{n\times 2}$ denote the matrix of endogenous variables.

In empirical applications, interest often is focussed on the parameter β on the rhs endogenous variable y_2.

3 Weak Instruments

It is well known that there are two key properties for IVs: (i) exogeneity, i.e., lack of correlation of the IVs with the structural equation error, and (ii) relevance, i.e., the ability of the IVs to explain the rhs endogenous variables. For many years, considerable attention has been paid in applications to the issue of exogeneity. Only more recently has attention been paid to relevance. Weak IVs concerns relevance.

IVs are weak if the mean component of y_2 that depends on the IVs, viz., $Z\pi$, is small relative to the variability of y_2, or equivalently, to the variability of the error v_2. This can be measured by the population partial R^2 of the equation (2.1) for y_2 (where the effect of the exogenous variables X is partialed out). In sample, it can be measured by the sample partial R^2 or, equivalently, by the F statistic for the null that $\pi = 0$ in (2.1), see Shea (1997), Godfrey (1999), and Section 5 below.

Note that IVs can be weak and the F statistic small either because π is close to zero or because the variability of Z is low relative to the variability of v_2. Also note that in practice the issue typically is how close to zero is π or $Z\pi$, not whether π or $Z\pi$ is exactly equal to zero.

There are numerous examples of weak IVs in the empirical literature. Here we mention three. The first is the classic example from labor economics of Angrist and Krueger's (1991) IV regression of wages on the endogenous variable years of education and additional covariates. Dummies for quarter of birth (with and without interactions with exogenous variables) are used as IVs for years of education. The argument is that quarter of birth is related to years of education via mandatory school laws for children aged sixteen and lower. At best, the relationship is weak, which leads to weak IVs. A notable feature of this application is that weak instrument issues arise despite the fact that Angrist and Krueger use a 5% Census sample with hundreds of thousands of observations. Evidently, weak instruments should not be thought of as merely a small-sample problem, and the difficulties associated with weak instruments can arise even if the sample size is very large.

The second example is from the macroeconomics/finance literature on the consumption CAPM. Interest concerns the elasticity of intertemporal substitution. In both linear and nonlinear versions of the model, IVs are weak, e.g., see Neeley, Roy, and Whiteman (2001), Stock and Wright (2000), and Yogo (2004). In one specification of the linear model in Yogo (2004), the endogenous variable is consumption growth and the IVs are twice lagged nominal interest rates, inflation, consumption growth, and log dividend-price ratio. Since log consumption is close to a random walk (see Hall (1978)), consumption growth is difficult to predict. This leads to the IVs being weak. For example, Yogo (2004) finds F statistics for the null hypothesis that $\pi = 0$ in the first stage regression that lie between 0.17 and 3.53 for different countries.

The third example is from the macroeconomics literature on the new Keynesian Phillips curve. Inference typically is carried out using IV methods, see Mavroeidis (2004) and Nason and Smith (2005). Twice-lagged real marginal cost (labor share in income) or change in inflation are used as IVs for current inflation growth. Given that changes in inflation are difficult to predict, the IVs are weak.

Much of the current interest in weak IVs started with two important papers by Nelson and Startz (1990, 1991) which showed the dramatically non-normal distributions of the 2SLS estimator and t-statistic when instruments are weak. The effect of weak IVs on these 2SLS statistics depends considerably on the degree of endogeneity, which is measured by the correlation, ρ_{u,v_2}, between u and v_2. If $\rho_{u,v_2} = 0$ or ρ_{u,v_2} is close to zero, then standard procedures work well (in terms of low bias and test and CI levels being close to their nominal values). When the IVs are weak and the degree of endogeneity is fairly strong, however, the 2SLS estimator has appreciable bias and the normal approximation to its distribution is poor. See Nelson and Startz (1990, 1991), Maddala and Jeong (1992), and Woglom (2001).

There is also a considerable earlier literature that is relevant to this issue, see the references in Phillips (1984). But, most of the early finite sample literature tended not to focus on the properties of procedures when the IVs are weak, because the empirical relevance of this scenario was not recognized. On other hand, for the LIML estimator, Phillips (1984, 1989) and Choi and Phillips (1992) investigate the finite sample and asymptotic properties under the limiting case of completely irrelevant IVs (lack of identification) as well as case of partial identification. Their results show quite different sampling properties under lack of identification or partial identification compared to strong identification. This suggests that problems arise when the model is identified, but the IVs are weak.

An influential paper by Bound, Baker, and Jaeger (1995) analyzes the properties of 2SLS in the context of Angrist and Krueger's (1991) regression of wages on education and exogenous variables. It shows that even when the sample size is huge, the properties of 2SLS can be poor in the face of many weak IVs. The sample size is not the key parameter. What matters most is the concentration parameter, $\lambda_{conc} = \pi' Z' Z \pi / \sigma^2_{v_2}$ (where $\sigma^2_{v_2}$ is the variance of v_2) and the degree of endogeneity, ρ_{uv_2}. The earlier finite sample literature was aware of this, e.g., see Rothenberg (1984).

Generic results for models with isolated unidentified points in the parameter space also have had an impact on the weak IV literature. Gleser and Hwang (1987) establish dramatic results concerning the levels of CIs and tests when the parameter space includes unidentified points. Dufour (1997) extends these results and applies them to the weak IV regression model. Consider any parametric model with parameters $(\beta, \pi) \in R^2$. Suppose the observations have common support. Suppose β is unidentified when $\pi = 0$. In this scenario, Gleser and Hwang's (1987) result states that a CI for β must have infinite length with positive probability—otherwise its (exact) level is zero. This result also holds (essentially by a continuity argument) if the parameter space is taken to be $\{(\beta, \pi) \in R^2 : \pi \neq 0\}$, which only includes unidentified points. The usual "estimator $\pm$ std error $\times$ constant" CI has finite length with probability one. Hence, the finite-sample confidence level of such a CI is zero under the specified conditions. Analogously, the finite sample significance levels of t-tests and Wald tests are one. These conclusions do not apply to LR and LM tests and CIs based on them because these CIs are not necessarily finite with probability one.

The idea behind this result is as follows. Suppose the parameter space is R^2. If $\pi = 0$, then a level $1 - \alpha$ CI must include $(-\infty, \infty)$ with probability greater than or equal to $1 - \alpha$ because every value of $\beta \in R$ is a true value when $\pi = 0$. Hence, the event "the CI has infinite length" has probability greater than or equal to $1 - \alpha$ when $\pi = 0$ for a CI with level $1 - \alpha$. By the common support assumption, this event must have positive probability for any (β, π) in the parameter space. In consequence, if a CI is finite with probability one for any (β, π), then its level must be less than $1 - \alpha$. If this holds for all $\alpha < 1$, then its level must be zero.

Note that the argument does not apply if the parameter space is bounded away from $\pi = 0$. But, in this case, the level of the CI typically lies between 0 and $1 - \alpha$ depending upon where the parameter space is truncated.

The conclusion from the result of Gleser and Hwang (1987) and Dufour (1997) is that CIs and tests based on t-tests and Wald tests cannot be fully robust to weak IVs. This concern is not just a theoretical nicety: numerical studies included in the papers cited in this section have demonstrated that coverage rates of conventional TSLS confidence intervals can be very poor when instruments are weak, even if the sample size is large, in designs calibrated to practical empirical applications such as the Angrist-Krueger data and the consumption CAPM.

4 Asymptotics

In this section we discuss some basic tools that are used in the weak IV literature, viz., different types of asymptotics. Asymptotic results are widely used in econometrics to provide approximations and to facilitate comparisons of estimators or tests. For the IV regression model, standard asymptotics let $n \to \infty$ and hold other features of the model fixed. We refer to these asymptotics as *strong IV asymptotics.* These asymptotics provide poor approximations in the IV regression model with weak IVs because (i) weak IVs correspond to a relatively small value of $\lambda = \pi' Z' Z \pi$, (ii) the finite sample properties of estimators and tests are sensitive to the magnitude of λ, and (iii) the asymptotic framework results in $\lambda \to \infty$.

In consequence, the literature on weak IVs has considered several alternative asymptotic frameworks. For linear models, Staiger and Stock (1997) consider asymptotics in which

$$\pi = C/\sqrt{n} \text{ for } n = 1, 2, ... \tag{4.1}$$

for some constant k-vector C. Combined with the standard assumption that $Z'Z/n \to D_Z$ as $n \to \infty$, for some $k \times k$ nonsingular matrix D_Z, this yields

$$\lambda = \pi' Z' Z \pi = C'(Z'Z/n)C \to C' D_Z C \text{ as } n \to \infty. \tag{4.2}$$

Thus, under this asymptotic set-up, λ converges to a finite constant as $n \to \infty$. Depending on the magnitude of the constant, the IVs are weaker or stronger. We refer to these asymptotics as *weak IV asymptotics.* Weak IV asymptotics are analogous to local-to-unity asymptotics that are widely used in the unit root time series literature.

An attractive feature of weak IV asymptotics is that they provide better approximations than standard asymptotics for the case where λ is small, yet still allow for the usual simplifications regarding non-normality, heteroskedasticity, and/or autocorrelation of the errors as under standard asymptotics. Under weak IV asymptotics, estimation of the 2×2 reduced-form covariance matrix, Ω, is an order of magnitude easier than estimation of the structural parameters (β, γ_1). In consequence, one typically obtains the same weak IV asymptotic results whether Ω is known or estimated, which is another useful simplification.

Under weak IV asymptotics, the "limiting experiment" is essentially the same as the finite sample model with iid homoskedastic normal errors and known reduced-form covariance matrix Ω. This has the advantages listed in the previous paragraph, but the disadvantage that the finite-sample normal model is significantly more complicated than the usual Gaussian-shift limiting experiment.

A second type of asymptotics utilized in the literature lets $k \to \infty$ as $n \to \infty$ with π fixed for all n. We call this *many IV asymptotics.* These asymptotics are appropriate when k is relatively large. But, Bekker (1994) argues that these asymptotics provide better approximations than standard asymptotics even when k is small. Many IV asymptotics have been employed by Anderson (1976), Kunitomo (1980), Morimune (1983), Bekker (1994), Donald and Newey (2001), Hahn (2002), Hahn, Hausman, and Kuersteiner (2004), and Hansen, Hausman, and Newey (2005) among others.

A third type of asymptotics, introduced by Chao and Swanson (2005), are *many weak IV asymptotics* in which $k \to \infty$ and $\pi \to 0$ as $n \to \infty$. These asymptotics are designed for the case in which one has relatively many IVs that are weak. Many weak IV asymptotics are employed by Stock and Yogo (2005a), Han and Phillips (2005), Anderson, Kunitomo, and Matsushita (2005), Hansen, Hausman, and Newey (2005), and Newey and Windmeijer (2005). Sections 9 and 10 below consider estimation and testing under many weak IV asymptotics.

5 Detecting Weak IVs

A small first stage F statistic for $H_0 : \pi = 0$ (or, equivalently, a low partial R^2) provides evidence that IVs are weak. Stock and Yogo (2005b) develop formal tests based on the F statistic for the null hypothesis: (1) the bias of 2SLS is greater than 10% of the bias based on OLS. The F test rejects the null of weak IVs at the 5% level if $F > 10.3$. They also consider the null hypothesis that (2) the null rejection rate of the nominal 5% 2SLS t test concerning β has a rejection rate 10% or greater. In this case, the F test rejects the null of weak IVs at the 5% level if $F > 24.6$. Analogous tests when the null hypothesis is specified in terms of the LIML estimator or Fuller's (1977) modification of LIML are provided in Stock and Yogo (2005b). These tests have different (smaller) critical values.

The adverse effect of weak IVs on standard methods, such as 2SLS, depends on the degree of endogeneity present as measured by ρ_{uv_2}, the correlation between the structural and reduced form errors u and v_2. But, ρ_{uv_2} is difficult to estimate precisely when the IVs are weak. In particular, ρ_{uv_2} cannot be consistently estimated under weak IVs asymptotics. (The reason is that the residuals used to estimate u depend on some estimator of β and β cannot be consistently estimated under weak IV asymptotics.) In consequence, the F tests of Stock and Yogo (2005b) are designed to be valid for any value of ρ_{uv_2}.

An alternative test for the detection of weak IVs based on reverse regressions is given by Hahn and Hausman (2002). Unfortunately, this test has very low power and is not recommended, at least for the purpose of detecting weak instruments, see Hausman, Stock, and Yogo (2005).

If one uses an F test to detect weak IVs as a pre-test procedure, then the usual pre-testing issues arise for subsequent inference, e.g., see Hall, Rudebusch, and Wilcox (1996). The approach of Chioda and Jansson (2004), which considers tests concerning β that are valid conditional on the value of the F statistic, can deal with such pre-testing issues. A drawback of this approach, however, is that it sacrifices power.

6 Approaches to Inference with Weak IVs

In most areas of econometrics, the standard method is to report a parameter estimate along with its standard error. Then, one utilizes CIs of the form "estimator $\pm$ std error $\times$ constant" and one carries out t tests using critical values from the normal distribution. This approach is suitable if the estimator has a distribution that is centered approximately at the true value and is reasonably well approximated by the normal distribution. In the case of IV estimation, this approach does not work well if the IVs are weak, as discussed Section 3.

One approach to dealing with weak IVs is to follow the "estimate/standard error" reporting method combined with a pre-test for weak IVs and/or a suitable choice of (bias-corrected) estimator, standard error estimator, and/or IVs in order to improve the normal approximation. For example, a sophisticated version of the latter is given in Donald and Newey (2001). See Hahn and Hausman (2003b), Hahn, Hausman, and Kuersteiner (2004), Hansen, Hausman, and Newey (2005), Newey and Windmeijer (2005), and Section 8 below for results and references to the literature. This approach is justified if the parameter space is bounded away from the unidentified set of parameter values where $\pi = 0$. But, the approach is not fully robust to weak IVs.

An alternative approach is to report an estimate and CI, where the CI is fully robust to weak IVs asymptotically. By this we mean that the CI has asymptotically correct coverage probability under standard asymptotics for $\pi \neq 0$ and under weak IV asymptotics (which includes the standard asymptotic case with $\pi = 0$). For testing, this approach uses tests that are fully robust to weak IVs asymptotically. For the weak IV problem, this approach was first advocated by Dufour (1997) and Staiger and Stock (1997).

Fully-robust CIs can be obtained by inverting fully-robust tests. For example, a CI for β of (approximate) level $100(1-\alpha)\%$ consists of all parameter values β_0 for which the null hypothesis $H_0 : \beta = \beta_0$ is not rejected at (approximate) level 5%. Note that standard CIs of the form "estimator $\pm$ std error $\times$ constant" are obtained by inverting t tests. Thus, the only difference between the CIs used in the first and second approaches is that the second approach employs tests that are fully robust to weak IVs rather than t tests. Papers that discuss the mechanics of the inversion of robust tests to form CIs include Zivot, Startz, and Nelson (1998), Dufour and Jasiak (2001), and Dufour and Taamouti (2005). Papers that report empirical results using CIs obtained by inverting fully-robust tests include Yogo (2004) and Nason and Smith (2005).

We advocate the second approach. In consequence, we view robust tests to be very important and we focus more attention in this paper on testing than on estimation.

7 Developments in Testing for the Linear Model

In this section, we consider testing in the model specified in (2.1)-(2.2). In applications, interest often is focused on the parameter β on the rhs endogenous variable y_2. Hence, our interest is in the null and alternative hypotheses:

$$H_0 : \beta = \beta_0 \text{ and } H_1 : \beta \neq \beta_0. \tag{7.1}$$

The parameter π, which determines the strength of the IVs, is a nuisance parameter that appears under the null and alternative hypotheses. The parameters γ, ξ, and Ω are also nuisance parameters, but are of lesser importance because tests concerning β typically are invariant to γ and ξ and the behavior of standard tests, such as t tests, are much less sensitive to Ω than to π.

We desire tests that are robust to weak IVs under normal or non-normal errors. In addition, we desire tests that exhibit robustness under weak and strong IV asymptotics to (i) left-out IVs, (ii) nonlinearity of the reduced form equation for y_2, (iii) heteroskedasticity—in cross-section contexts, (iv) heteroskedasticity and/or autocorrelation—in some time series contexts, and (v) many IVs—in some contexts.

Some, e.g., Dufour (1997), desire tests whose finite-sample null rejection rate is exactly the desired significance level under iid homoskedastic normal errors for any π. But, we view this as putting excessive weight on the iid homoskedastic normal assumptions, which are not likely to hold in practice.

Here we evaluate tests based on their significance levels and power functions. An alternative approach, based on decision theory, is studied by Chamberlain (2005) for the linear IV model. In the latter approach, nuisance parameters, such as π, are integrated out using a prior.

7.1 Anderson-Rubin and LM Tests

The first test employed specifically to deal with weak IVs is the Anderson-Rubin (1949) (AR) test, see Dufour (1997) and Staiger and Stock (1997). The AR test imposes the null $\beta = \beta_0$ and uses the F test for the artificial null hypothesis $H_0^* : \kappa = 0$ in the model

$$y_1 - y_2\beta_0 = Z\kappa + X\gamma + u. \tag{7.2}$$

The AR test statistic is

$$\begin{aligned} AR(\beta_0) &= \frac{(y_1 - y_2\beta_0)' P_Z (y_1 - y_2\beta_0)/k}{\widehat{\sigma}_u^2(\beta_0)}, \text{ where} \\ \widehat{\sigma}_u^2(\beta_0) &= \frac{(y_1 - y_2\beta_0)' M_{[Z:X]} (y_1 - y_2\beta_0)}{n-k-p}. \end{aligned} \tag{7.3}$$

Under the null hypothesis $H_0 : \beta = \beta_0$, we have

$$AR(\beta_0) = \frac{u' P_Z u/k}{u' M_{[Z:X]} u/(n-k-p)}. \tag{7.4}$$

The null distribution of the AR statistic does not depend on π regardless of the distribution of the errors u and the AR test is fully robust to weak IVs. Under H_0, $AR(\beta_0) \to_d \chi_k^2/k$ under strong and weak IV asymptotics assuming iid homoskedastic errors u with two moments finite. Under the additional assumption of normal errors, $AR(\beta_0) \sim F_{k,n-k-p}$. Hence, an F critical value is typically employed with the AR test. As pointed out by Dufour (2003), the AR test does not rely on any assumptions concerning the reduced-form equation for y_2. But, the AR test is not robust to heteroskedasticity and/or autocorrelation of the structural equation error u.

Under the alternative hypothesis,

$$AR(\beta_0) = \frac{(u + Z\pi(\beta - \beta_0))' P_Z (u + Z\pi(\beta - \beta_0))}{u' M_{[Z:X]} u/(n-k-p)}. \tag{7.5}$$

In consequence, power depends on the magnitude of $\pi' Z' Z\pi(\beta - \beta_0)^2$. The power of AR test is very good when $k = 1$. Moreira (2001) shows that it is UMP unbiased when the errors are iid homoskedastic normal, Ω is known, and $k = 1$.

On the other hand, when $k > 1$, the power of AR test is not so good. It is a k degrees of freedom test when only one parameter is under test. When the true parameter is β, model is can be written as

$$y_1 - y_2\beta_0 = Z\pi(\beta - \beta_0) + X\gamma + v_1. \tag{7.6}$$

The AR test tests whether Z enters this equation. The AR test sacrifices power because it ignores the restriction that $\kappa = \pi(\beta - \beta_0)$. Obviously, low power is an undesirable property for tests, and it leads to excessively long CIs based on such tests.

The literature has sought more powerful tests than the AR test that are robust to weak IVs. To start, one might consider the LR test of $H_0 : \beta = \beta_0$. The LR statistic combined with the conventional χ_1^2 critical value is not robust to weak IVs when $k > 1$. But, Wang and Zivot (1998) show that a larger critical value, obtained by bounding the asymptotic distribution of the LR statistic over π values, is robust to weak IVs. However, the resulting test is nonsimilar asymptotically under weak IV asymptotics (i.e., its asymptotic null rejection rate depends on π). In consequence, the test is biased and sacrifices power. For instance, the test is not asymptotically efficient under strong IV asymptotics.

Kleibergen (2002) and Moreira (2001) independently introduce an LM test whose null rejection rate is robust to weak IVs. They use different ideas to arrive at the LM test. Kleibergen's idea is as follows: The AR statistic projects $y_1 - y_2\beta_0$ onto the k-dimensional space spanned by Z. Instead, one can estimate π under the null hypothesis via the ML estimator, $\widehat{\pi}(\beta_0)$, and project onto the one-dimensional space spanned by $Z\widehat{\pi}(\beta_0)$. It turns out that $\widehat{\pi}(\beta_0)$ is asymptotically independent of $y_1 - y_2\beta_0$ under the null under strong and weak IV asymptotics. Hence, a suitably scaled version of the projection residuals is asymptotically χ_1^2 under the null hypothesis under both strong and weak IV asymptotics. The resulting test statistic is an LM statistic and it is fully robust to weak IVs. (For an alternative derivation of this LM statistic, see Poskitt and Skeels (2005).)

The power of the LM test often is better than that of the AR test when $k > 1$, but not always. A drawback of the LM test is that it exhibits quirky power properties including non-monotonicity in $|\beta - \beta_0|$.

We note that when $k = 1$ the LM, LR, and AR tests are equivalent because the LM and LR statistics equal k times the AR statistic.

7.2 Similar Tests

Moreira (2001, 2003) gives a characterization of (exactly) similar tests of $H_0 : \beta = \beta_0$ for the model of (2.1)-(2.2) with normal errors and known Ω. Similar tests are necessarily fully robust to weak IVs because their null distribution does not depend on π. When Ω is unknown, a consistent estimator of Ω, say $\widehat{\Omega}$, can be plugged in for Ω and an exactly similar test becomes a test that is asymptotically similar under weak IV asymptotics (and typically under strong IV asymptotics as well).

By definition, a level α test $\phi(Y)$ (which equals one when the test rejects H_0 and zero otherwise) is similar if

$$E_\pi \phi(Y) = \alpha \text{ for all } \pi, \tag{7.7}$$

where $E_\pi(\cdot)$ denotes expectation under (β_0, π).

Moreira's argument leading to the characterization of similar tests is as follows. Under the null hypothesis, the unknown parameter is π. Using standard methods, e.g., Lehmann (1986), the null-restricted ML estimator of π, $\widehat{\pi}(\beta_0)$, can be shown to be a complete sufficient statistic for π because the parametric model is in the exponential family of distributions. By definition, a sufficient statistic $\widehat{\pi}(\beta_0)$ is complete for π if, for any real function $h(\cdot)$, $E_\pi h(\widehat{\pi}(\beta_0)) = c$ for all π implies that $h(\cdot) = c$ a.s. for some constant c. Applying this definition of complete sufficiency with $h(\cdot) = E_\pi(\phi(Y)|\widehat{\pi}(\beta_0) = \cdot)$ and using the law of iterated expectations gives:

$$E_\pi h(\widehat{\pi}(\beta_0)) = E_\pi E_\pi(\phi(Y)|\widehat{\pi}(\beta_0)) = E_\pi \phi(Y) = \alpha \text{ for all } \pi \tag{7.8}$$

implies that $h(\cdot) = E_\pi(\phi(Y)|\widehat{\pi}(\beta_0) = \cdot) = \alpha$ a.s. That is, $\phi(Y)$ is similar with level α implies that

$$E_\pi(\phi(Y)|\widehat{\pi}(\beta_0) = x) = \alpha \text{ for all } x. \tag{7.9}$$

The converse is immediate by the law of iterated expectations.

Given this result, similar tests can be created from non-similar tests by employing a conditional critical value function $\kappa_{\phi,\alpha}(x)$ defined by

$$\kappa_{\phi,\alpha}(x) = E(\phi(Y)|\widehat{\pi}(\beta_0) = x) = \alpha \tag{7.10}$$

(where the expectation is taken under the null). Then, a similar test based on ϕ rejects H_0 if

$$\phi(Y) > \kappa_{\phi,\alpha}(\widehat{\pi}(\beta_0)). \tag{7.11}$$

This method generates "conditional tests" that are similar. For example, one can generate a conditional Wald test given an estimator such as 2SLS or LIML.

In addition, one can consider the conditional LR (CLR) test. Moreira (2003) focuses on this test. It is a more sophisticated version of Wang and Zivot's (1998) LR test in which a critical value function replaces a constant to achieve the desired level α. The CLR test has higher power than the Wang-Zivot LR test.

The CLR test based on a plug-in value of Ω has asymptotic null rejection rate α under both weak and strong IV asymptotics and, hence, is fully robust to weak IVs.

Note that the critical value function $\kappa_{\phi,\alpha}(x)$ of (7.10) can be computed by numerical integration or simulation and then tabulated, see Moreira (2003) and Andrews, Moreira, and Stock (2004b, c).

Moreira, Porter, and Suarez (2004) introduce a residual-based bootstrap for the CLR test which is shown to be first-order correct under strong IV asymptotics whether $\pi = 0$ or $\pi \neq 0$. Its behavior under weak IV asymptotics is not discussed. This bootstrap does not deliver higher-order improvements.

7.3 Optimal Tests

The "conditioning" method leads to a surfeit of tests that are fully robust to weak IVs because any test can be made fully robust. Given this, Andrews, Moreira, and Stock (2004a) (AMS) address the question of optimal tests that are robust to weak IVs. They consider the class of similar tests for the model of (2.1)-(2.2) with iid homoskedastic normal errors and known Ω. If π is known, then it is plausible that an optimal two-sided test is just the t test of $H_0 : \beta = \beta_0$ in the model

$$y_1 - y_2\beta_0 = (Z\pi)(\beta - \beta_0) + X\gamma + u. \tag{7.12}$$

Indeed, Moreira (2001) shows that the two-sided t test for the case when π and Ω are known is UMP unbiased. But π is not known in practice, so no optimal two-sided test exists. The problem is that the class of tests considered is too large. In consequence, AMS restricts attention to similar tests that satisfy a rotational invariance property.

The data matrix Y has a multivariate normal distribution, which is a member of the exponential family of distributions. In consequence, low dimensional sufficient statistics are available. For tests concerning β, there is no loss (in terms of attainable power functions) in considering tests that are based on the sufficient statistic $Z'Y$ for $(\beta, \pi')'$, see AMS. This eliminates the nuisance parameters (γ, ξ) from the problem. The nuisance parameter π remains. As in Moreira (2003), we consider a one-to-one transformation of $Z'Y$ given by $[S : T]$, where

$$\begin{aligned} S &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\Omega b_0)^{-1/2}, \\ T &= (Z'Z)^{-1/2}Z'Y\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2}, \\ b_0 &= (1, -\beta_0)' \text{ and } a_0 = (\beta_0, 1)'. \end{aligned} \tag{7.13}$$

The invariant tests considered in AMS depend on S and T only through the maximal invariant statistic Q defined by

$$Q = [S{:}T]'[S{:}T] = \begin{bmatrix} S'S & S'T \\ T'S & T'T \end{bmatrix} = \begin{bmatrix} Q_S & Q_{ST} \\ Q_{ST} & Q_T \end{bmatrix}. \tag{7.14}$$

(See AMS for the definition of the groups of transformations on the data matrix $[S{:}T]$ and the parameters (β, π) that yields the maximal invariant to be Q. Note that $Y'P_ZY$ is an equivalent statistic to Q.)

For example, the AR, LM, and LR test statistics can be written as

$$\begin{aligned} AR &= Q_S/k, \ LM = Q_{ST}^2/Q_T, \text{ and} \\ LR &= \frac{1}{2}\left(Q_S - Q_T + \sqrt{(Q_S - Q_T)^2 + 4Q_{ST}^2}\right). \end{aligned} \tag{7.15}$$

The only tests that we are aware of that are not functions of Q are tests that involve leaving out some IVs and t tests based on Chamberlain and Imbens (2004) many IV estimator.

The matrix Q has a noncentral Wishart distribution with means matrix of rank one and identity variance matrix. The distribution of Q (and hence of invariant similar tests) only depends on the unknown parameters β and

$$\lambda = \pi' Z' Z \pi \in R, \tag{7.16}$$

where λ indicates the strength of the IVs (and is proportional to the concentration parameter λ_{conc}). The utilization of invariance has reduced the k-vector nuisance parameter π to a scalar nuisance parameter λ. The distribution of Q also depends on the number IVs, k, and the parameter of interest, β.

AMS derives a power envelope for two-sided invariant similar (IS) tests and compares new and existing tests to the power envelope. The power envelope is determined by the tests that are point optimal for each (β, λ) combination. There are several ways to impose two-sidedness. First, AMS considers average power against two points (β^*, λ^*) and (β_2^*, λ_2^*), where (β_2^*, λ_2^*) is selected given (β^*, λ^*) to be the unique point such that β_2^* is on the opposite side of the null value β_0 from β^* and such that the test that maximizes power against these two points is asymptotically efficient (AE) under strong IV asymptotics. The power envelope is then a function of (β^*, λ^*). Second, AMS considers tests that satisfy a sign-invariance condition that the test depends on Q_{ST} only through $|Q_{ST}|$. This is satisfied by most tests in the literature including the AR, LM, and CLR tests. Third, AMS considers locally unbiased tests. AMS shows that the first and second approaches yield exactly the same power envelope. Furthermore, numerical calculations show that the third approach yields a power envelope that is hard to distinguish from the first and second.

AMS develops a class of new tests based on maximizing weighted average power (WAP) given different weight functions on (β, λ). These tests and the AR, LM, and CLR tests are compared numerically to the two-sided power envelope for IS tests. The power envelope only depends on β, λ, k, $\rho = corr(v_1, v_2)$ and is smooth in these parameters. Hence, it is possible to make fairly exhaustive comparisons.

Figure 1 illustrates some of these comparisons for the case of $k = 5$ instruments. The figure plots the power of the AR, LM, and CLR tests as a function of β, along with the asymptotically efficient power envelope for two-sided invariant similar tests. The first three panels show the effect of varying $\rho = corr(v_1, v_2)$ (negative values of ρ correspond to reversing the sign of $\beta - \beta_0$) and for $\beta_0 = 0$ (taken without loss of generality). Panels (a) - (c) consider rather weak instruments, $\lambda = 10$, which corresponds to a first-stage F-statistic having a mean of $\lambda/k + 1 = 3$ under the null. Panel (d) presents the power functions for a case of relatively strong instruments, $\lambda = 80$.

Figure 1 and other results in AMS show that the CLR test is essentially on the power envelope for all β, λ, k, and ρ. In contrast, the AR test typically is below the power envelope—the more so, the larger is k. In addition, the LM test has a quirky non-monotone power function that is sometimes on the power envelope and sometimes far from it. See AMS for an explanation of this behavior. Results in AMS indicate that some point-optimal IS two-sided (POIS2) tests are very close to the power envelope, like the CLR test. This is also true of some WAP tests based on nondegenerate weight functions. But, these tests do not have as simple closed-form expressions as the LR test statistic.

One might ask the question of whether the restriction to similar tests is overly restrictive. Andrews, Moreira, and Stock (2004b) provides some results for the power envelope for invariant non-similar tests using the least favorable distribution approach of Lehmann (1986). In the cases in which the nonsimilar power envelope could be computed, it is found to be very close to that for similar tests. Hence, the imposition of similarity does not appear to be overly restrictive.

The finite sample power envelope for known Ω that is determined in AMS is shown to be the same as the weak IV asymptotic power envelope for unknown Ω. Hence, the feasible CLR test is (essentially) on the weak IV asymptotic power envelope for two-sided IS tests. In addition, AMS show that the CLR test (and the LM test) are asymptotically efficient under strong IV asymptotics.

Based on these results, we recommend the CLR test among tests that are designed for iid homoskedastic errors.

7.4 Conditional LR Test

There is a one-to-one transformation from the restricted ML estimator $\widehat{\pi}(\beta_0)$ to the statistic Q_T. Hence, the CLR test can be written such that it rejects the null hypothesis when

$$LR > \kappa_\alpha^{CLR}(Q_T), \tag{7.17}$$

where $\kappa_\alpha^{CLR}(Q_T)$ is defined to satisfy $P_{\beta_0}(LR > \kappa_\alpha^{CLR}(q_T)|Q_T = q_T) = \alpha$.

We note that the LR statistic combines the AR and LM statistics based on the magnitude of $Q_T - Q_S$. If Q_T is much larger than Q_S, then LR is essentially LM. If Q_T is much smaller than Q_S, then LR is essentially k·AR. This can be seen more clearly by re-writing LR as follows:

$$LR = \frac{1}{2}\left(Q_S - Q_T + |Q_T - Q_S|\sqrt{1 + \frac{4Q_T}{(Q_T - Q_S)^2}LM}\right). \tag{7.18}$$

Now, as $Q_T \to \infty$ and $Q_S/Q_T \to 0$, we have (i) $Q_T - Q_S \to \infty$, (ii) $Q_T/(Q_T - Q_S)^2 \to 0$, (iii) the $\sqrt{\cdot}$ term is approximately $1 + 2[Q_T/(Q_T - Q_S)^2]LM$ by a mean-value expansion, (iv) $|Q_T - Q_S| = Q_T - Q_S$, (v) $|Q_T - Q_S|[Q_T/(Q_T - Q_S)^2]LM$ goes to LM, and (vi) LR goes to LM. On the other hand, as $Q_T \to 0$, (i) the $\sqrt{\cdot}$ term goes to 1, (ii) $Q_S - Q_T + |Q_T - Q_S|$ goes to $2Q_S$, and (iii) LR goes to $Q_S = k \cdot AR$.

We briefly illustrate the use of the CLR test to construct CIs using some results from Yogo (2004). Yogo reports CIs for the elasticity of intertemporal substitution obtained from regressions of stock returns on consumption growth using quarterly data from 1970 to 1998 with four IVs: twice-lagged nominal interest rate, inflation, consumption growth, and log of the dividend/price ratio. CIs are obtained by inverting AR, LM, and CLR tests. Yogo (2004) reports results for several countries. For most countries, the CIs are $(-\infty, \infty)$ using all three methods because the IVs are very weak in this model specification. For Canada, the CIs are AR: $[.02, 4.03]$, LM: $[.05, .35]$, and CLR: $[.04, .41]$. In this case, the AR CI is much wider than LM and CLR CIs. For France, the CIs are AR: $[-.28, .20]$, LM: $(-\infty, \infty)$, and CLR: $[-.16, .11]$. In this case, the LM CI is uninformative and the CLR CI is noticeably shorter than the AR CI. These results illustrate that the power advantages of the CLR test translate into shorter more informative CIs (at least in these applications).

7.5 Conditional t Tests

The usual test employed in practice is a t test based on the 2SLS estimator. Analogously, the usual CI employed is a t test-generated CI of the form "estimator $\pm$ std error $\times$ constant." As noted above, a t test using the normal approximation is not robust to weak IVs. But, one can construct a conditional t test that is robust by using a conditional critical value function like that for the CLR test. In consequence, one can answer the question of how good is a t test in terms of power once it has been corrected to get its size correct.

Andrews, Moreira, and Stock (2006) compare the power properties of conditional t tests based on several estimators, including 2SLS, LIML, and Fuller's (1977) modification of LIML, to the CLR test and the power envelope for two-sided IS tests. In short, the results indicate that t tests have very poor power properties when the IVs are weak. Power is often very low on one side of the null hypothesis. The CLR test has much better power properties. We conclude that in the presence of weak IVs, t tests not only lack robustness to weak IVs in terms of size, but their power properties are also poor after size-correction by conditioning.

7.6 Robustness to Heteroskedasticity and/or Autocorrelation

Although the CLR test has good power properties, it is not robust to heteroskedasticity or autocorrelation. However, the LR test statistic can be replaced by a heteroskedasticity-robust version, HR-LR, or a heteroskedasticity and autocorrelation-robust version, HAR-LR, that is robust, see Andrews, Moreira, and Stock(2004a).

Given the HR-LR or HAR-LR statistic, the same conditional critical value function as for the CLR test is used to yield HR-CLR and HAR-CLR tests. These tests are robust to heteroskedasticity and/or autocorrelation under weak and strong IV asymptotics. They are also robust to left-out IVs and nonlinear reduced-form for y_2, which can be viewed to be causes of heteroskedasticity. Furthermore, the asymptotic properties of the HR-CLR and HAR-CLR tests under homoskedasticity are same as those of the CLR test. So, HR-CLR and HAR-CLR tests are optimal IS two-sided tests under homoskedastic iid errors.

Kleibergen (2005a, b) also gives heteroskedasticity and/or autocorrelation robust tests that reduce to the CLR test under homoskedasticity. Kleibergen's tests apply in the more general context of nonlinear moment conditions. Note that Kleibergen's tests are not the same as those in AMS under heteroskedasticity and/or autocorrelation even asymptotically under weak IV asymptotics. But, they are the equivalent asymptotically under weak and strong IV asymptotics with homoskedastic iid errors.

Other tests that are robust to weak IVs as well as to heteroskedasticity and/or autocorrelation are the generalized empirical likelihood (GEL) tests of Guggenberger and Smith (2005a, b), Otsu (2005), and Caner (2003). The GEL tests are analogues of the AR and LM tests discussed above. They have the same asymptotic properties under weak and strong IV asymptotics as heteroskedasticity and/or autocorrelation robust versions of the AR and LM tests. The GEL tests are not as powerful asymptotically as the HR-CLR or HAR-CLR tests under iid homoskedastic errors. This is probably also true under a variety of forms of heteroskedasticity and/or autocorrelation.

Note that it should be possible to construct a GEL version of the HR-CLR and HAR-CLR tests by combining the GEL versions of the AR and LM statistics to form a GEL-LR statistic using the formula for LR given in (7.15) written as a function of the AR and LM statistics and employing the critical value function for the CLR test.

7.7 Power with Non-normal Errors

The power results of Section 7.3 are for normal errors. The question arises whether tests exist with good power for normal errors and higher power for non-normal errors. In certain contexts, the answer is yes. In particular, rank-based versions of the AR and CLR tests have this property, see Andrews and Marmer (2003) and Andrews and Soares (2004). With iid homoskedastic (possibly non-normal) errors and given certain conditions on the exogenous variables and IVs, the rank-based AR test based on normal scores has exact significance level α and has power that asymptotically dominates the power of the AR test. Specifically, its power is higher for thick-tailed error distributions. Hence, under the given conditions and $k = 1$ (in which case, the AR, LM, and CLR tests are the same), the rank-based AR test is quite attractive. It is robust to left-out IVs and nonlinear reduced form for y_2. On the other hand, it is not robust to heteroskedasticity or autocorrelation of the structural errors u. It also relies on stronger assumptions concerning the exogenous variables and IVs than the AR test. Hence, there is a trade-off between the rank and non-rank tests.

When $k > 1$, the rank-based CLR test has power advantages over the CLR test for thick tailed errors. But, it is not robust to heteroskedasticity or autocorrelation. Hence, there is a trade-off between the rank-based CLR test and the HR-CLR and HAR-CLR tests in terms of power against non-normal errors and these robustness properties.

We conclude that in certain circumstances rank-based tests are preferable to the HR-CLR or HAR-CLR tests. But, in most circumstances, the latter are preferred.

7.8 Multiple Right-hand Side Endogenous Variables

Next, suppose y_2 and β are vectors. The AR, LM, and CLR tests of $H_0 : \beta = \beta_0$ all generalize to this case, see Anderson and Rubin (1949), Kleibergen (2002), and Moreira (2003). All of these tests are robust to weak IVs. One would expect the relative power properties of these tests to be similar to those in the case of a scalar β. However, the optimal power properties of the CLR test established in AMS do not carry over in a straightforward manner because the Wishart distribution of the (data) matrix Q that appears in the vector β case has a means matrix of rank two or greater, rather than rank one. Nevertheless, based on the scalar β power comparisons, we recommend the CLR (HR-CLR, or HAR-CLR) test over the AR and LM tests. The CLR test does have the drawback that the conditional critical value function is higher dimensional and, hence, more cumbersome in the vector β case.

To obtain a CI or test for an individual coefficient, say β_1, in the vector $\beta = (\beta_1, \beta_2')' \in R^m$, there are several approaches to inference when the IVs are weak—none is completely satisfactory. First, one can construct a CI via the projection method. The idea is to construct an approximate $100(1-\alpha)\%$ confidence region in R^m for β using a test that is robust to weak IVs. Then, the approximate $100(1-\alpha)\%$ CI for β_1 is the set of β_1 values for which $(\beta_1, \beta_2')'$ is in the confidence region for some β_2. In turn, an approximate level α test of $H_0 : \beta_1 = \beta_{10}$ rejects the null hypothesis if β_{10} is not in the CI for β_1. For the application of this method using the AR test, see Dufour and Jasiak (2001) and Dufour and Taamouti (2005). A drawback of this method is that it is conservative. Hence, it yields CIs that are longer than desirable.

An alternative method, discussed in Moreira (2005), relies on exclusion restrictions in the reduced-form equations for the endogenous variables whose coefficients are β_2 in the structural equation. Given suitable restrictions, tests are available that are asymptotically similar under weak and strong IV asymptotics. Such tests have the drawback of relying on exclusion restrictions and of sacrificing power—they are not asymptotically efficient under strong IV asymptotics.

A third approach is available that is partially robust to weak IVs. Suppose β_1 is asymptotically weakly identified and β_2 is asymptotically strongly identified. See Stock and Wright (2000), Kleibergen (2004), or Guggenberger and Smith (2005a) for precise definitions of what this means. Then, asymptotically non-conservative tests for β_1 are available by concentrating out β_2. Simulation results in Guggenberger and Smith (2005a) indicate that this method works fairly well in terms of size even if β_2 not very strongly identified.

7.9 Inference on Other Coefficients

Suppose one is interested in a test or CI concerning a coefficient, say γ_a, on an exogenous variable, say X_a, in the structural equation for y_1, where $X\gamma_1 = X_a\gamma_a + X_b\gamma_b$. In this case, t tests based on the 2SLS or LIML estimator of γ_a do not necessarily perform poorly under weak IVs asymptotics. What is key for good performance asymptotically for 2SLS is that y_2 is explained asymptotically by more than just the part of X_a that is orthogonal to X_b. For example, one can show that consistency and asymptotic normality of the 2SLS estimator holds in the model of (2.1)-(2.2) with iid homoskedastic errors u with two moments finite if

$$\begin{aligned} &\liminf_{n\to\infty} (\pi' Z' Z\pi + \xi_b^{*\prime} X_b' X_b \xi_b^*)/n > 0 \text{ and } \lim_{n\to\infty} X_a^{*\prime} X_a^*/n = M > 0, \text{ where} \\ &y_2 = Z\pi + X_a^*\xi_a + X_b\xi_b^*,\ X_a^* = M_{X_b}X_a,\ \text{and } \xi_b^* = \xi_b + (X_b'X_b)^{-1}X_b'X_a\xi_a. \end{aligned} \tag{7.19}$$

Thus, consistency holds even if $\pi = C/\sqrt{n}$ provided $\liminf_{n\to\infty} \xi_b^{*\prime} X_b' X_b \xi_b^*/n > 0$ (and $\lim_{n\to\infty} X_a^{*\prime} X_a^*/n = M > 0$). This sort of result is closely related to results under partial identification of Phillips (1989).

There are several approaches to inference that are robust to weak IVs and small values of $\xi_b^{*\prime} X_b^{\prime} X_b \xi_b^*$, but none is completely satisfactory. One one can use projection as in Section 7.8, but this has the same potential drawback as described above. If X_a (or equivalently X_a^*) does not enter the reduced-form equation for y_2, then $\xi_a = 0$ and least squares estimation of y_1 on X_a^* yields a consistent estimator of γ_a and least squares t tests are invariant to π and ξ_b^*. But, such an exclusion restriction may not be plausible in a given application.

7.10 Tests in Nonlinear Moment Condition Models

Weak IVs appear not only in linear models, but also in nonlinear models specified by moment conditions. Such models typically are estimated by generalized method of moments (GMM), although generalized empirical likelihood (GEL) estimation also is possible. There have been significant contributions to the literature recently concerning inference in such models when the IVs are weak.

Stock and Wright (2000) extend weak IV asymptotics to nonlinear models by taking part of the population moment function to be local to zero for all values of the parameter as $n \to \infty$. This implies that part of the matrix of derivatives of the sample moments is local to zero as $n \to \infty$. These results allow one to assess both the null rejection probabilities of tests under weak IV asymptotics and their power properties. Stock and Wright (2000) consider extensions of the AR test and corresponding CIs to the nonlinear moment condition model.

Kleibergen (2002, 2005a, b) extends the weak IV-robust LM test for the IV regression model to the moment condition model. This allows him to construct, by analogy, a CLR test for the nonlinear moment condition model, call it GMM-CLR. The GMM-CLR test is robust to weak IVs and allows for heteroskedasticity and/or autocorrelation. Its power properties have not been investigated, but one would think that in many cases they would reflect the power advantages of the CLR test in linear models.

Generalized empirical likelihood (GEL) versions of the weak IV robust AR and LM tests have been constructed by Guggenberger and Smith (2005a, b), Otsu (2005), and Caner (2003). These tests are robust to heteroskedasticity and/or autocorrelation. Their properties under the null and alternatives using the weak IV asymptotics of Stock and Wright (2000) are the same as those of the AR and LM moment condition tests. Hence, one would expect that these GEL tests are not as powerful as the GMM-CLR test, at least if the amount of heteroskedasticity and autocorrelation is not too large. We note that it should be possible to construct a GEL-based test that is an analogue of the CLR test.

Given current knowledge, we recommend the GMM-CLR test of Kleibergen (2005a, b) for dealing with weak IVs in moment condition models. However, knowledge in this area is not completely developed and it remains an open question whether a test that dominates GMM-CLR can be developed.

8 Estimation with Weak IVs

This section gives a brief discussion of estimation with weak IVs. Under weak IV asymptotics, what is relevant is estimation in the normal linear model with known Ω. There is a substantial older literature comparing estimators in this model. See Rothenberg (1984) and Phillips (1984) for references. But, the relevance of some of this literature is diminished for two reasons. First, many comparisons are based on higher-order expansions under standard SI asymptotics. For example, see Nagar (1959) and Rothenberg (1984) for further references. Hahn, Hausman, and Kuersteiner (2004) (HHK) and Chao and Swanson (2003) find such higher-order expansions are not accurate under weak IV parameter configurations. Second, Monte Carlo comparisons in the older literature tend to be for relatively strong IV parameter configurations, see HHK.

Chamberlain (2005) has provided a theoretical optimality result for the LIMLk estimator (i.e., LIML for the case of known Ω) in the model of (2.1)-(2.2). He shows that LIMLk is minimax for a particular bounded loss function. He also shows that LIMLk minimizes average risk over certain ellipses in π-space independently of the radius of these ellipses. These are nice theoretical properties, but the lack of finite integer moments of LIML and LIMLk indicates that the loss function employed may not be desirable if one is concerned with large estimation errors.

Via simulation, HHK confirm that LIML displays high dispersion under weak IV parameter configurations relative to other estimators even when dispersion is measured by the interquartile range rather than by the variance (which is infinite for LIML and LIMLk). LIML does exhibits low median bias, which is consistent with LIMLk being exactly median unbiased.

Fuller's (1977) modification of LIML with $a = 1$ or 4 has finite moments and its median-bias is relatively small in HHK's simulations with weak IV parameter configurations. The jackknife 2SLS estimator also fares well in these simulations.

There has been recent research on Bayes estimators for the model of (2.1)-(2.2) that is relevant for weak IV contexts. Kleibergen and van Dijk (1998) consider a diffuse prior that is designed to handle lack of identification due to weak IVs. Kleibergen and Zivot (2003) consider priors that yield posteriors that are of the same form as the densities of 2SLS and LIML. Chao and Phillips (1998, 2002) construct the Jeffreys prior Bayes estimator for the limited information model. Zellner (1998) introduces a Bayesian method of moments estimator (BMOM), which is in the family of double k class estimators. Unlike the other Bayes estimators mentioned above, BMOM is not equivariant to shifts in β. Gao and Lahiri (1999) carry out a Monte Carlo comparisons of several of these Bayes estimators under weak IV parameter configurations.

By standard results, Bayes estimators based on proper priors and their limits yield a complete class of estimators for the model with iid homoskedastic normal errors and known Ω. Hence, these estimators also form an asymptotically complete class under weak IV asymptotics when Ω is unknown and is replaced by a consistent estimator.

Other related papers include Forchini and Hillier (2003), who consider conditional estimation given the eigenvalues of the “first stage F” matrix, and Hahn and Hausman (2003a) and Kiviet and Niemczyk (2005), who consider the relative attributes of OLS and 2SLS.

We conclude this section by noting that there does not seem to be any dominant estimation method for the linear model with weak IVs. At this time, we recommend Fuller’s (1977) modified LIML estimator with $a = 1$ or 4 as a good choice in terms of overall properties.

9 Estimation with Many Weak IVs

To date the literature concerning models with many weak IVs has focussed on estimation. Chao and Swanson (2005) consider asymptotics in which $k \to \infty$ and $\pi \to 0$ as $n \to \infty$. The idea of their paper is that a sufficient number of weak IVs may provide enough additional information regarding β to yield consistent estimation. Indeed, they establish that consistent estimation of β is possible even if $\pi = C/\sqrt{n}$ (as in weak IV asymptotics) provided $k \to \infty$. For consistency of LIML or the jackknife IV estimator (JIVE), one needs $\lambda/k^{1/2} \to \infty$ as $n \to \infty$, where $\lambda = \pi Z' Z \pi$ indexes the strength of the IVs. For consistency of 2SLS, on the other hand, one needs a faster growth rate of λ: $\lambda/k \to \infty$ as $n \to \infty$. This advantage of LIML over 2SLS is consistent with the higher-order bias properties of LIML and 2SLS under many (non-weak) IV asymptotics, see HHK.

Han and Phillips (2005) consider fixed weight-matrix GMM estimators for nonlinear moment condition models. Their results show that many different types of asymptotic behavior of such estimators is possible depending on the rates of growth of the strength of the IVs relative to k. They provide conditions for convergence in probability to a constant, which is not necessarily true value. They provide results based on high-level conditions that indicate that the asymptotic distributions of fixed weight-matrix GMM estimators can be normal or non-normal.

Hansen, Hausman, and Newey (2005) and Newey and Windmeijer (2005) analyze LIML and Fuller’s (1977) modified LIML estimators and generalized empirical likelihood (GEL) estimators, respectively, when $\lambda/k \to c$ as $n \to \infty$ for some constant $c > 0$. They show that these estimators are asymptotically normal, but with a variance matrix has an additional term compared to usual fixed k case. They provide a new asymptotic variance estimator that is asymptotically correct under many weak IV asymptotics when $\lambda/k \to c$. Although this variance estimator helps to robustify inference to many weak IVs, tests and CIs based on this asymptotic approximation are not fully robust to weak IVs or large k.

Anderson, Kunitomo, and Matsushita (2005) consider a (restricted) class of estimators in the linear IV model and show that the LIML estimator satisfies some asymptotic optimality properties within this class under many weak IV asymptotics when $\lambda/k \to c$ as $n \to \infty$ for some constant $c > 0$. (The same is true for Fuller’s (1977) modified LIML estimator.)

Chamberlain and Imbens (2004) develop a random-effects quasi-ML estimator based on a random coefficients structure on the relation between the rhs endogenous variable and the IVs. This structure reduces the number of parameters to be estimated. (They do not consider the asymptotic distribution of the estimator under many weak IV asymptotics.)

Chao and Swanson (2003) introduce bias-corrected linear IV estimators based on sequential asymptotics in which $n \to \infty$ and then $\lambda_{conc}/k \to \infty$.

10 Testing with Many Weak Instruments

In this section, we present new results concerning tests in the asymptotic framework of many weak IVs. We are interested in the relative performance of the AR, LM, and CLR tests and in their performance relative to an asymptotic power envelope.

10.1 Asymptotic Distribution of Q for Large λ and k

We consider the model of (2.1)-(2.2) with iid homoskedastic normal errors. The hypotheses of interest are given in (7.1). The statistics S and T are defined in (7.13). As in AMS, we focus on invariant tests, which are functions of the maximal invariant Q defined in (7.14). Our goal is to obtain an asymptotic two-sided power envelope for invariant tests concerning β when $(\lambda, k) \to \infty$. To achieve this, we need to determine the asymptotic distribution of Q as $(\lambda, k) \to \infty$. For clarity, we sometimes denote Q by $Q_{\lambda,k}$.

The means of S and T depend on the following quantities:

$$\begin{aligned} \mu_\pi &= (Z'Z)^{1/2}\pi \in R^k, \ c_\beta = (\beta - \beta_0) \cdot (b_0'\Omega b_0)^{-1/2} \in R, \text{ and} \\ d_\beta &= a'\Omega^{-1}a_0 \cdot (a_0'\Omega^{-1}a_0)^{-1/2} \in R, \text{ where } a = (\beta, 1)'. \end{aligned} \tag{10.1}$$

The distributions of S and T are $S \sim N(c_\beta \mu_\pi, I_k)$, $T \sim N(d_\beta \mu_\pi, I_k)$,and S and T are independent, see Lemma 2 of AMS.

Under the model specification given above, the following assumption holds.

Assumption 1. $Q = Q_{\lambda,k} = [S{:}T]'[S{:}T]$, where $S \sim N(c_\beta \mu_\pi, I_k)$, $T \sim N(d_\beta \mu_\pi, I_k)$, S and T are independent, and $(\mu_\pi, c_\beta, d_\beta)$ are defined in (10.1).

By definition, under Assumption 1, Q has a noncentral Wishart distribution with mean matrix $M = \mu_\pi (c_\beta, d_\beta)'$ (of rank one) and identity covariance matrix. The distribution of Q only depends on μ_π through $\lambda = \mu_\pi' \mu_\pi$. The density of Q is given in Lemma 3 of AMS.

Next, we specify the rate at which λ and k diverge to infinity. Our results allow for a wide range of possibilities. All limits are as $(\lambda, k) \to \infty$. Let $\chi_1^2(\delta)$ denote a noncentral chi-square distribution with one degree of freedom and noncentrality parameter δ.

Assumption 2. $\lambda/k^\tau \to r_\tau$ for some constants $\tau \in (0, \infty)$ and $r_\tau \in [0, \infty)$.

The asymptotic distribution of Q depends on the following quantities:

$$\begin{aligned} V_{3,\tau} &= \begin{cases} Diag\{2,1,2\} & \text{if } 0<\tau\leq 1/2 \\ Diag\{2,1,0\} & \text{if } 1/2<\tau<1 \\ Diag\{2,1+d_{\beta_0}^2 r_1,0\} & \text{if } \tau=1 \\ Diag\{2,d_{\beta_0}^2 r_\tau,0\} & \text{if } \tau>1 \quad \text{and} \end{cases} \\ \gamma_B &= (b_0'\Omega b_0)^{-1/2} d_{\beta_0} B \end{aligned} \tag{10.2}$$

for a scalar constant B.

The asymptotic distribution of Q is given in the following theorem.

Theorem 1 *Suppose Assumptions* 1 *and* 2 *hold.*
(a) *If* $0<\tau<1/2$ *and* β *is fixed,*

$$\begin{gathered} \begin{pmatrix} (S'S-k)/k^{1/2} \\ S'T/k^{1/2} \\ (T'T-k)/k^{1/2} \end{pmatrix} \to_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, V_{3,\tau} \right), \\ (AR-1)k^{1/2} \to_d \overline{Q}_{S,\infty} \sim N(0,2),\ LM \to_d \overline{Q}_{ST,\infty}^2 \sim \chi_1^2(0),\ \textit{and} \\ LR/k^{1/2} \to_d \frac{1}{2}\left(\overline{Q}_{S,\infty} - \overline{Q}_{T,\infty} + \sqrt{(\overline{Q}_{T,\infty} - \overline{Q}_{S,\infty})^2 + 4\overline{Q}_{ST,\infty}^2} \right). \end{gathered}$$

(b) *If* $\tau=1/2$ *and* β *is fixed,*

$$\begin{aligned} &\begin{pmatrix} (S'S-k)/k^{1/2} \\ S'T/k^{1/2} \\ (T'T-k)/k^{1/2} \end{pmatrix} \to_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim N\left(\begin{pmatrix} c_\beta^2 r_{1/2} \\ c_\beta d_\beta r_{1/2} \\ d_\beta^2 r_{1/2} \end{pmatrix}, V_{3,1/2} \right), \\ &\qquad (AR-1)k^{1/2} \to_d \overline{Q}_{S,\infty} \sim N(c_\beta^2 r_{1/2}, 2), \\ &\qquad LM \to_d \overline{Q}_{ST,\infty}^2 \sim \chi_1^2(c_\beta^2 d_\beta^2 r_{1/2}^2),\ \textit{and} \\ &LR/k^{1/2} \to_d \frac{1}{2}\left(\overline{Q}_{S,\infty} - \overline{Q}_{T,\infty} + \sqrt{(\overline{Q}_{T,\infty} - \overline{Q}_{S,\infty})^2 + 4\overline{Q}_{ST,\infty}^2} \right). \end{aligned}$$

(c) *If* $1/2<\tau\leq 1$ *and* $\beta=\beta_0+Bk^{1/2-\tau}$ *for a scalar constant* B,

$$\begin{aligned} &\begin{pmatrix} (S'S-k)/k^{1/2} \\ S'T/k^{1/2} \\ (T'T-k)/k^{\tau} \end{pmatrix} \to_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \gamma_B r_\tau \\ d_{\beta_0}^2 r_\tau \end{pmatrix}, V_{3,\tau} \right), \\ &\qquad (AR-1)k^{1/2} \to_d \overline{Q}_{S,\infty} \sim N(0,2), \\ &\qquad LM \to_d \overline{Q}_{ST,\infty}^2 \sim \chi_1^2(\gamma_B^2 r_\tau^2) \ \textit{when}\ 1/2<\tau<1, \\ &LM \to_d \overline{Q}_{ST,\infty}^2/(1+d_{\beta_0}^2 r_1) \sim \chi_1^2(\gamma_B^2 r_\tau^2/(1+d_{\beta_0}^2 r_1)) \ \textit{when}\ \tau=1, \\ &LR = (1/(d_{\beta_0}^2 r_\tau))k^{1-\tau} LM(1+o_p(1)) \ \textit{when}\ 1/2<\tau<1,\ \textit{and} \\ &LR = ((1+d_{\beta_0}^2 r_1)/(d_{\beta_0}^2 r_1))LM + o_p(1) \ \textit{when}\ \tau=1. \end{aligned}$$

(d) *If* $\tau > 1$, $r_\tau > 0$, *and* $\beta = \beta_0 + Bk^{-\tau/2}$,

$$\begin{pmatrix} (S'S-k)/k^{1/2} \\ S'T/k^{\tau/2} \\ (T'T-k)/k^{\tau} \end{pmatrix} \rightarrow_d \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \gamma_B r_\tau \\ d_{\beta_0}^2 r_\tau \end{pmatrix}, V_{3,\tau}\right),$$

$$(AR-1)k^{1/2} \rightarrow_d \overline{Q}_{S,\infty} \sim N(0,2),$$

$$LM \rightarrow_d \overline{Q}_{ST,\infty}^2/(d_{\beta_0}^2 r_\tau) \sim \chi_1^2(\gamma_B^2 r_\tau / d_{\beta_0}^2)) \text{ provided } d_{\beta_0} \neq 0, \text{ and}$$

$$LR = LM + o_p(1).$$

Comments. 1. An interesting feature of Theorem 1 is that the statistics $S'S$, $S'T$, and $T'T$ are asymptotically independent.

2. Part (a) of the Theorem shows that when the concentration parameter λ grows at a rate slower than $k^{1/2}$ the statistic Q has an asymptotic distribution that does not depend on the parameter β. Hence, in this case, no test has non-trivial asymptotic power.

3. The result of part (a) also holds when λ is fixed and $k \to \infty$.

4. Part (b) of the Theorem is the most interesting case. When the concentration parameter λ grows at the rate $k^{1/2}$, all three normalized statistics $S'S$, $S'T$, and $T'T$ have asymptotic distributions that depend on β. In this case, the growth in λ is not sufficiently fast that consistent estimation of β is possible (otherwise, tests with asymptotic power one against fixed alternatives would be available).

5. Parts (c) and (d) show that when λ grows at a rate faster than $k^{1/2}$ the AR statistic has trivial asymptotic power against local alternatives for which the LM and LR statistics have non-trivial power. Furthermore, in this case the LM and LR statistics are asymptotically equivalent.

6. The cases considered in Chao and Swanson (2005) and Han and Phillips (2005) correspond to $\tau > 1/2$. Those considered in Stock and Yogo (2005a), Anderson, Kunitomo, and Matsushita (2005), Hansen, Hausman, and Newey (2005), and Newey and Windmeijer (2005) correspond to the case where $\tau = 1$.

10.2 Two-sided Asymptotic Power Envelopes

In the next several subsections, we determine asymptotic power envelopes for two-sided tests. We start with the most interesting case in which $\lambda/k^{1/2} \to r_{1/2}$, i.e., the case $\tau = 1/2$. Subsequently, we consider the case $\tau > 1/2$. In contrast to the results in AMS, we do not restrict attention to tests that are asymptotically similar. Rather, we allow for both asymptotically similar and non-similar tests and show that the power envelope is determined by tests that are asymptotically similar.

There are several ways of constructing a two-sided power envelope depending on how one imposes the two-sidedness condition. The approach we take here is based on determining the highest possible average power against a point $(\beta, \lambda) = (\beta^*, \lambda^*)$ and another point, say (β_2^*, λ_2^*), for which β_2^* lies on the other side of the null value β_0 than β^*. (The power envelope then is a function of $(\beta, \lambda) = (\beta^*, \lambda^*)$.) Given (β^*, λ^*), we select (β_2^*, λ_2^*) in the same way as in AMS. In particular, the point (β_2^*, λ_2^*) has the property that the test that maximizes average power against these two points is *asymptotically efficient* under strong IV asymptotics when the number of IVs k is fixed as $\lambda \to \infty$. Furthermore, the power of the test that maximizes average power against these two points is the same for each of the two points. This choice also has the desirable properties that (a) the marginal distributions of Q_S, Q_{ST}, and Q_T under (β_2^*, λ_2^*) are the same as under (β^*, λ^*), (b) the joint distribution of (Q_S, Q_{ST}, Q_T) under (β_2^*, λ_2^*) equals that of $(Q_S, -Q_{ST}, Q_T)$ under (β^*, λ^*), which corresponds to β_2^* being on the other side of the null from β^*, and (c) the distribution of $[-S : T]$ under (β_2^*, λ_2^*) equals that of $[S : T]$ under (β^*, λ^*).

Given (β^*, λ^*), the point (β_2^*, λ_2^*) that has these properties is shown in AMS to satisfy $(\lambda_2^*)^{1/2} c_{\beta_2^*} = -(\lambda^*)^{1/2} c_{\beta^*}$ $(\neq 0)$ and $(\lambda_2^*)^{1/2} d_{\beta_2^*} = (\lambda^*)^{1/2} d_{\beta^*}$ and is given by

$$\begin{aligned} \beta_2^* &= \beta_0 - \frac{d_{\beta_0}(\beta^* - \beta_0)}{d_{\beta_0} + 2g(\beta^* - \beta_0)} \text{ and } \lambda_2^* = \lambda^* \frac{(d_{\beta_0} + 2g(\beta^* - \beta_0))^2}{d_{\beta_0}^2}, \text{ where} \\ g &= e_1' \Omega^{-1} a_0 \cdot (a_0' \Omega^{-1} a_0)^{-1/2} \text{ and } e_1 = (1, 0)'. \end{aligned} \tag{10.3}$$

(provided $\beta^* \neq \beta_{AR}$, where β_{AR} denotes the point β at which $d_\beta = 0$, see AMS).

The average power of a test $\phi(Q)$ against the two points (β^*, λ^*) and (β_2^*, λ_2^*) is given by

$$K(\phi; \beta^*, \lambda^*) = \frac{1}{2}\left[E_{\beta^*, \lambda^*}\phi(Q) + E_{\beta_2^*, \lambda_2^*}\phi(Q)\right] = E^*_{\beta^*, \lambda^*}\phi(Q), \tag{10.4}$$

where $E_{\beta,\lambda}$ denotes expectation with respect to the density $f_{Q_1,Q_T}(q_1, q_T; \beta, \lambda)$, which is the joint density of (Q_1, Q_T) at (q_1, q_T) when (β, λ) are the true parameters, and $E^*_{\beta^*,\lambda^*}$ denotes expectation with respect to the density

$$f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*) = \frac{1}{2}\left[f_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*) + f_{Q_1,Q_T}(q_1, q_T; \beta_2^*, \lambda_2^*)\right]. \tag{10.5}$$

Hence, the average power (AP) of $\phi(Q)$ against (β^*, λ^*) and (β_2^*, λ_2^*) can be written as the power against the single density $f^*_{Q_1,Q_T}(q_1, q_T; \beta^*, \lambda^*)$.

10.3 Asymptotic Power Envelope for $\tau = 1/2$

We now consider the case where $\tau = 1/2$. We restrict ourselves to invariant tests that depend on the data only through

$$\begin{aligned} \overline{Q}_{\lambda,k} &= (Q_{\lambda,k} - kI_2)/k^{1/2}, \text{ where} \\ vech(\overline{Q}_{\lambda,k}) &= (\overline{Q}_{S,k}, \overline{Q}_{ST,k}, \overline{Q}_{T,k})'. \end{aligned} \tag{10.6}$$

A test $\phi_k(\overline{Q}_{\lambda,k})$ is $\{0,1\}$-valued and rejects the null hypothesis when $\phi_k = 1$. We say that a sequence of tests $\{\phi_k : k \geq 1\}$ is a *convergent sequence of asymptotically level α tests* for $\tau = 1/2$ if there exists a $\{0,1\}$-valued function ϕ such that under Assumption 1, under any sequence $\lambda/k^{1/2} \to r_{1/2}$ for any $r_{1/2}$ in some non-empty subset of $[0, \infty)$, and for any β in some set that includes $\{\beta_0, \beta^*, \beta_2^*\}$, we have

$$\begin{aligned}
&\phi_k(\overline{Q}_{\lambda,k}) \rightarrow_d \phi(\overline{Q}_\infty), \text{ where} \\
&vech(\overline{Q}_\infty) = \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim N\left(\begin{pmatrix} c_\beta^2 r_{1/2} \\ c_\beta d_\beta r_{1/2} \\ d_\beta^2 r_{1/2} \end{pmatrix}, V_{3,1/2} \right), \text{ and} \\
&P_{\beta_0, r_{1/2}}(\phi(\overline{Q}_\infty) = 1) \leq \alpha,
\end{aligned} \tag{10.7}$$

where $P_{\beta, r_{1/2}}(\cdot)$ denotes probability when the true parameters are $(\beta, r_{1/2})$. By Theorem 1, examples of convergent sequences of asymptotically level α tests include sequences of CLR, LM, and AR tests. Standard Wald and LR tests are not asymptotically level α.

We now determine the average power envelope for the asymptotic testing problem. Let $\overline{Q}_\infty$ be distributed as in (10.7). The unknown parameters are β and $r_{1/2}$. We are interested in average power against two points $(\beta^*, r_{1/2}^*)$ and $(\beta_2^*, r_{2,1/2}^*)$, where β_2^* is as defined in (10.3) and $r_{2,1/2}^*$ is the asymptotic analogue of λ_2^* and is defined in (10.3) with λ^* replaced by $r_{1/2}^*$. As in (10.4), the average power of a test based on $\overline{Q}_\infty$ equals its power against the single alternative whose density is the average of the densities for the points $(\beta^*, r_{1/2}^*)$ and $(\beta_2^*, r_{2,1/2}^*)$.

The null hypothesis is composite and the alternative is simple, so we use the "least favorable" approach of Lehmann (1986, Sec. 3.8) to find a best test of level α. The idea is as follows. Given a distribution over the parameters in the null hypothesis, say $G(\cdot)$, one obtains a single distribution by integrating the null distribution with respect to $G(\cdot)$. For any $G(\cdot)$, one can construct the level α LR test for the simple null versus the simple alternative and it is best according to the Neyman-Pearson Lemma. If one can find a distribution, $G_{LF}(\cdot)$, called a least favorable distribution, for which the simple versus simple LR test is of level α not just for the simple null but for the underlying composite null as well, then this LR test is the best level α test for the composite null versus the simple alternative. The reason is that this test is best against the simple alternative subject to the constraint that the average null rejection rate weighted by $G_{LF}(\cdot)$ is less than or equal to α, which is a weaker constraint than the constraint that the pointwise rejection rate is less than or equal to α for all points in the composite null.

The key step in the implementation of the least favorable approach is the determination of a least favorable distribution. In the present case, the parameter that appears under the null hypothesis is $r_{1/2}$ and it only affects the distribution of $\overline{Q}_{T,\infty}$ because $c_{\beta_0} = 0$ implies that $\overline{Q}_{S,\infty}$ and $\overline{Q}_{ST,\infty}$ have mean zero under H_0 (see (10.7)). The distribution of $\overline{Q}_{T,\infty}$ under $(\beta^*, r_{1/2}^*)$ and $(\beta_2^*, r_{2,1/2}^*)$ is the same because $d_{\beta^*}^2 r_{1/2}^* = d_{\beta_2^*}^2 r_{2,1/2}^*$, see (10.3), and $\overline{Q}_{T,\infty} \sim N(d_\beta^2 r_{1/2}, 2)$ under $(\beta, r_{1/2})$ by (10.7). Hence, the distribution of $\overline{Q}_{T,\infty}$ under the simple alternative whose density is the average of the densities for the points $(\beta^*, r_{1/2}^*)$ and $(\beta_2^*, r_{2,1/2}^*)$ is just $N(d_{\beta^*}^2 r_{1/2}^*, 2)$.

Under the null hypothesis, $\overline{Q}_{T,\infty} \sim N(d_{\beta_0}^2 r_{1/2}, 2)$. Hence, if we take the least favorable distribution to be a point mass at $r_{1/2}^{LF}$, where $d_{\beta_0}^2 r_{1/2}^{LF} = d_{\beta^*}^2 r_{1/2}^*$ or equivalently

$$r_{1/2}^{LF} = d_{\beta^*}^2 r_{1/2}^* / d_{\beta_0}^2, \tag{10.8}$$

then $\overline{Q}_{T,\infty}$ has the same distribution under the simple null hypothesis as under the simple alternative hypothesis. Given the independence of $(\overline{Q}_{S,\infty}, \overline{Q}_{ST,\infty})$ and $\overline{Q}_{T,\infty}$, the simple versus simple LR test statistic does not depend on $\overline{Q}_{T,\infty}$. This implies that the pointmass distribution at $r_{1/2}^{LF}$ is indeed least favorable because the simple versus simple LR statistic depends only on $(\overline{Q}_{S,\infty}, \overline{Q}_{ST,\infty})$, the null distribution of latter does not depend on $r_{1/2}$, and hence the pointwise null rejection rate of the simple versus simple level α LR test is α for each point in the null hypothesis.

The above discussion establishes that the best test for testing the composite null against the simple alternative based on the average density determined by $(\beta^*, r_{1/2}^*)$ and $(\beta_2^*, r_{2,1/2}^*)$ is given by the likelihood ratio for $(\overline{Q}_{S,\infty}, \overline{Q}_{ST,\infty})$. This likelihood ratio statistic is given by

$$\begin{aligned}
&LR^*(\overline{Q}_\infty; \beta^*, r_{1/2}^*) = LR^*(\overline{Q}_{S,\infty}, \overline{Q}_{ST,\infty}; \beta^*, r_{1/2}^*) \\
&= \exp(-\frac{1}{4}(\overline{Q}_{S,\infty} - c_{\beta^*}^2 r_{1/2}^*)^2) \times \\
&\quad \frac{(\exp(-\frac{1}{2}(\overline{Q}_{ST,\infty} - c_{\beta^*} d_{\beta^*} r_{1/2}^*)^2) + \exp(-\frac{1}{2}(\overline{Q}_{ST,\infty} + c_{\beta^*} d_{\beta^*} r_{1/2}^*)^2))}{2\exp(-\frac{1}{4}\overline{Q}_{S,\infty}^2 - \frac{1}{2}\overline{Q}_{ST,\infty}^2))} \\
&= \exp(-c_{\beta^*}^4 r_{1/2}^{*2}/4 - c_{\beta^*}^2 d_{\beta^*}^2 r_{1/2}^{*2}/2) \exp(c_{\beta^*}^2 r_{1/2}^* \overline{Q}_{S,\infty}/2) \times \qquad (10.9) \\
&\quad \left(\exp(c_{\beta^*} d_{\beta^*} r_{1/2}^* \overline{Q}_{ST,\infty}) + \exp(-c_{\beta^*} d_{\beta^*} r_{1/2}^* \overline{Q}_{ST,\infty})\right)/2 \\
&= \exp(-c_{\beta^*}^4 r_{1/2}^{*2}/4 - c_{\beta^*}^2 d_{\beta^*}^2 r_{1/2}^{*2}/2) \exp(c_{\beta^*}^2 r_{1/2}^* \overline{Q}_{S,\infty}/2) \cosh(c_{\beta^*} d_{\beta^*} r_{1/2}^* \overline{Q}_{ST,\infty}).
\end{aligned}$$

The critical value, $\kappa_\alpha^*(\beta^*, r_{1/2}^*)$, for the LR^* test is defined by

$$\begin{aligned}
&P(LR^*(N_1, N_2; \beta^*, r_{1/2}^*) > \kappa_\alpha^*(\beta^*, r_{1/2}^*)) = \alpha, \text{ where} \\
&(N_1, N_2)' \sim N(0, Diag\{2, 1\}).
\end{aligned} \tag{10.10}$$

By varying $(\beta^*, r_{1/2}^*)$, the power of the $LR^*(\overline{Q}_\infty; \beta^*, r_{1/2}^*)$ test traces out the average power envelope for level α similar tests for the asymptotic testing problem.

The results above lead to the following Theorem, which provides an upper bound on asymptotic average power.

Theorem 2 *The average power over (β^*, λ^*) and (β_2^*, λ_2^*) of any convergent sequence of invariant and asymptotically level α tests for $\tau = 1/2$, $\{\phi_k(\overline{Q}_{\lambda,k}) : k \geq 1\}$, satisfies*

$$\begin{aligned}
\lim_{k\to\infty} (1/2)[P_{\beta^*,\lambda^*}(\phi_k(\overline{Q}_{\lambda,k}) = 1) + P_{\beta_2^*,\lambda_2^*}(\phi_k(\overline{Q}_{\lambda,k}) = 1)] \\
= P_{\beta^*, r_{1/2}^*}^*(\phi(\overline{Q}_\infty) = 1) \\
\leq P_{\beta^*, r_{1/2}^*}^*(LR^*(\overline{Q}_\infty; \beta^*, r_{1/2}^*) > \kappa_\alpha(\beta^*, r_{1/2}^*)),
\end{aligned}$$

where $P_{\beta,\lambda}(\cdot)$ denotes probability when $Q_{\lambda,k}$ has the distribution specified in Assumption 1 with parameters (β, λ), $P_{\beta, r_{1/2}}(\cdot)$ denotes probability when $\overline{Q}_\infty$ has the distribution in (10.7), and $P_{\beta^*, r_{1/2}^*}^*(\cdot) = (1/2)[P_{\beta^*, r_{1/2}^*}(\cdot) + P_{\beta_2^*, r_{2,1/2}^*}(\cdot)]$.

The upper bound on average asymptotic power given in Theorem 2 is attained by a point optimal invariant two-sided (POI2) test that rejects H_0 if

$$LR^*(\overline{Q}_{S,k}, \overline{Q}_{ST,k}; \beta^*, r^*_{1/2}) > \kappa^*_\alpha(\beta^*, r^*_{1/2}). \tag{10.11}$$

This test is asymptotically similar. Hence, the asymptotic power envelope for similar and non-similar invariant tests is the same. The asymptotic distribution of the test statistic in (10.11) is given in the following Corollary to Theorem 1(b) (which holds by the continuous mapping theorem).

Corollary 1 *Suppose Assumptions* 1 *and* 2 *hold with* $\tau = 1/2$ *and* β *is fixed. Then,*

$$LR^*(\overline{Q}_{\lambda,k}; \beta^*, r^*_{1/2}) \to_d LR^*(\overline{Q}_\infty; \beta^*, r^*_{1/2}),$$

where $\overline{Q}_\infty$ *is distributed as in* (10.7).

Comments. 1. Corollary 1 implies that the POI2 test of (10.11) is a convergent sequence of invariant and asymptotically level α tests that attains the upper bound on asymptotic average power given in Theorem 2 at $(\beta^*, r^*_{1/2})$.

2. Corollary 1 shows that the upper bound in Theorem 2 is attainable and, hence, that the upper bound is the asymptotic power envelope when $\tau = 1/2$.

10.4 Numerical Results for Many Weak Instruments

We now turn to a brief summary of numerical properties of the many weak instrument power envelope and the AR, LM, and CLR tests under the $\tau = 1/2$ sequence. The limiting power envelope and power functions for various values of $r_{1/2}$ are presented in Figure 2 for $\rho = .95$ and in Figure 3 for $\rho = .5$. These figures, and unreported additional numerical work, support three main conclusions. First, the power function of the CLR test is effectively on the limiting asymptotically efficient power envelope, so the CLR test is, in effect, UMP among asymptotically efficient invariant tests under the $\tau = 1/2$ sequence. Second, the power function of the LM test sometimes falls well below the power envelope and is not monotonic in the many-instrument limit. Third, the performance of the AR test, relative to the power envelope and the CLR and LM tests, depends heavily on the strength of the instruments. For very weak instruments, the AR power function is below the power envelope but the AR test still exhibits nontrivial power. As the strength of the instruments increases (that is, as $r_{1/2}$ increases), the power of the AR test, relative to the other tests, is increasingly poor, and the power function is nearly flat in the case of panel (d) in both figures.

These results apply to the limit of the sequence (λ, k) as $\lambda/k^{1/2} \to r_{1/2}$. It is of interest to examine numerically the speed of convergence of the finite-k power envelopes and power functions to this limit. This is done in Figures 4 and 5, which present the power envelope and the CLR power function for various values of k and for the $k \to \infty$ limit. Evidently the speed of convergence, and the quality of the $k \to \infty$ approximation, depends on the strength of the instruments. For very weak instruments (panels (a) and (b)), the rate of convergence is fairly fast and the limiting functions are close to the finite-k approximations. For stronger instruments, the limiting approximation is less good and is achieved less quickly. An important point to note in Figures 4 and 5 is that for each parameter value and for each value of k, the CLR power function is effectively on the power envelope. Whether or not the $k \to \infty$ approximation is a good one for finite k, the CLR power function is in effect on the power envelope for asymptotically efficient two-sided invariant similar tests.

Figures 6 and 7 present a final set of numerical results, in which we consider performance of the CLR and AR tests along the sequence $\tau = 0$; this corresponds to the addition of irrelevant instruments as k increases. In the limit that $k \to \infty$, these tests have trivial power, but this result does not tell us directly how costly it is to err on the side of adding an irrelevant instrument. Perhaps surprisingly, for these and other cases not reported, adding a few irrelevant instruments is not very costly in terms of power for the CLR test; less surprisingly, adding a great number of irrelevant instruments drives the power to zero.

10.5 Asymptotic Power Envelope When $\tau > 1/2$

We now consider the asymptotic power envelope for the case where $\tau > 1/2$. In this case, the alternatives that we consider are local to the null hypothesis and significant simplifications occur. By Theorem 1(c) and (d), the asymptotic distribution of the normalized $Q_{\lambda,k}$ matrix only depends on the unknown localization parameter B through the distribution of $\overline{Q}_{ST,\infty}$ (because $\overline{Q}_{S,\infty} \sim N(0,2)$ and $\overline{Q}_{T,\infty} = d^2_{\beta_0} r_\tau$). The asymptotic testing problem concerns the hypotheses $H_0' : B = 0$ versus $H_1' : B \neq 0$ with no nuisance parameter (because $\overline{Q}_{T,\infty} = d^2_{\beta_0} r_\tau$ implies that r_τ is known asymptotically). An asymptotic sufficient statistic for B is $\overline{Q}_{ST,\infty}$, which has distribution

$$\overline{Q}_{ST,\infty} \sim \begin{cases} N(\gamma_B r_\tau, 1) & \text{when } 1/2 < \tau < 1 \\ N(\gamma_B r_1, 1 + d^2_{\beta_0} r_1) & \text{when } \tau = 1 \\ N(\gamma_B r_\tau^{1/2}, d^2_{\beta_0} r_\tau) & \text{when } \tau > 1. \end{cases} \tag{10.12}$$

Using the same sort of argument as in Section 10.3, by the Neyman-Pearson Lemma, the level α test based on $\overline{Q}_{ST,\infty}$ that maximizes average power against B^* and $-B^*$ is constructed using the following likelihood ratio statistic:

$$\begin{aligned} LR^{**} &= \frac{\exp(-\frac{1}{2}(\overline{Q}_{ST,\infty} - \gamma_{B^*} r_\tau)^2) + \exp(-\frac{1}{2}(\overline{Q}_{ST,\infty} + \gamma_{B^*} r_\tau)^2)}{2\exp(-\frac{1}{2}\overline{Q}^2_{ST,\infty})} \\ &= \exp(-\gamma^2_{B^*} r^2_\tau/2)\left(\exp(\gamma_{B^*} r_\tau \overline{Q}_{ST,\infty}) + \exp(-\gamma_{B^*} r_\tau \overline{Q}_{ST,\infty})\right)/2 \\ &= \exp(-\gamma^2_{B^*} r^2_\tau/2)\cosh(\gamma_{B^*} r_\tau \overline{Q}_{ST,\infty}) \end{aligned} \tag{10.13}$$

for the case $1/2 < \tau < 1$. The test that rejects when LR^{**} is large is equivalent to the test that rejects when $\overline{Q}^2_{ST,\infty}$ is large (because $\cosh(x)$ is increasing in $|x|$). Hence, the level α test that maximizes average power against B^* and $-B^*$ does not depend on B^*. The test rejects H_0' if

$$\overline{Q}^2_{ST,\infty} > \chi^2_1(\alpha), \tag{10.14}$$

where $\chi_1^2(\alpha)$ denotes the $1-\alpha$ quantile of a chi-squared distribution with one degree of freedom. Similar calculations for the cases $\tau = 1$ and $\tau > 1$ yield the same test as delivering maximal average power for any B^*.

Returning now to the finite sample testing problem, we restrict attention to invariant tests $\phi_k(\overline{Q}_{\lambda,k})$ that depend on the normalized data matrix

$$\overline{Q}_{\lambda,k} = \begin{pmatrix} (S'S-k)/k^{1/2} \\ S'T/k^{1/2} \\ (T'T-k)/k^{\tau} \end{pmatrix} \text{ or } \overline{Q}_{\lambda,k} = \begin{pmatrix} (S'S-k)/k^{1/2} \\ S'T/k^{\tau/2} \\ (T'T-k)/k^{\tau} \end{pmatrix} \tag{10.15}$$

for $1/2 < \tau \leq 1$ or $\tau > 1$, respectively. We say that a sequence of tests $\{\phi_k(\overline{Q}_{\lambda,k}) : k \geq 1\}$ is a *convergent sequence of asymptotically level α tests* for $\tau > 1/2$ if there exists a $\{0,1\}$-valued function ϕ such that under Assumption 1, under any sequence $\lambda/k^{\tau} \to r_{\tau}$ for any r_{τ} in some non-empty subset of $[0,\infty)$, for $\beta = \beta_0 + Bk^{1/2-\tau}$ when $1/2 < \tau \leq 1$ and $\beta = \beta_0 + Bk^{-\tau/2}$ when $\tau > 1$, and for any B in some set that includes $\{0, B^*, -B^*\}$, we have

$$\begin{aligned} \phi_k(\overline{Q}_{\lambda,k}) &\to_d \phi(\overline{Q}_{\infty}), \text{ where} \\ vech(\overline{Q}_{\infty}) &= \begin{pmatrix} \overline{Q}_{S,\infty} \\ \overline{Q}_{ST,\infty} \\ \overline{Q}_{T,\infty} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ \gamma_B \overline{r}_{\tau} \\ d_{\beta_0}^2 r_{\tau} \end{pmatrix}, V_{3,\tau}\right), \\ \overline{r}_{\tau} &= \begin{cases} r_{\tau} & \text{when } 1/2 < \tau \leq 1 \\ r_{\tau}^{1/2} & \text{when } \tau > 1, \end{cases} \\ P_{\beta_0}(\phi(\overline{Q}_{\infty}) &= 1) \leq \alpha, \end{aligned} \tag{10.16}$$

and $P_{\beta_0}(\cdot)$ denotes probability when the true parameter is β_0. By Theorem 1, the CLR, LM, and AR tests are examples of convergent sequences of asymptotically level α tests for $\tau > 1/2$.

The next result provides an upper bound on average power for convergent sequences of invariant asymptotically level α tests.

Theorem 3 *The average power over B^* and $-B^*$ of any convergent sequence of invariant asymptotically level α tests for $\tau > 1/2$, $\{\phi_k(\overline{Q}_{\lambda,k}) : k \geq 1\}$, satisfies*

$$\begin{aligned} &\lim_{k\to\infty} (1/2)[P_{\beta_k^*,\lambda}(\phi_k(\overline{Q}_{\lambda,k}) = 1) + P_{\beta_{2,k}^*,\lambda}(\phi_k(\overline{Q}_{\lambda,k}) = 1)] \\ &\qquad = P_{B^*}^*(\phi(\overline{Q}_{\infty}) = 1) \leq P_{B^*}^*(\overline{Q}_{ST,\infty}^2 > \chi_1^2(\alpha)), \end{aligned}$$

where $P_{\beta,\lambda}(\cdot)$ denotes probability when $Q_{\lambda,k}$ has the distribution specified in Assumption 1 with parameters (β,λ), $\lambda/k^{\tau} \to r_{\tau}$, $\beta_k^* = \beta_0 + B^*k^{1/2-\tau}$ and $\beta_{2,k}^* = \beta_0 - B^*k^{1/2-\tau}$ when $1/2 < \tau \leq 1$, $\beta_k^* = \beta_0 + B^*k^{-\tau/2}$ and $\beta_{2,k}^* = \beta_0 - B^*k^{-\tau/2}$ when $\tau > 1$, $P_B(\cdot)$ denotes probability when $\overline{Q}_{ST,\infty} \sim N(\gamma_B \overline{r}_{\tau}, 1)$, and $P_B^*(\cdot) = (1/2)[P_B(\cdot) + P_{-B}(\cdot)]$.

Comment. 1. The upper bound on average asymptotic power given in Theorem 3 is attained for all B^* and $-B^*$ by the LM and CLR tests by Theorem 1(c) and (d) (because the LM and LR test statistics are scalar multiples of $\overline{Q}^2_{ST,\infty}$ asymptotically). In consequence, we say that these tests are *asymptotically efficient* when $\tau > 1/2$.

10.6 Asymptotic Power Envelope for Unknown Ω

The asymptotic power envelopes provided in the preceding sections presume that the covariance matrix Ω of the reduced form errors in (2.2) is known. Tests based on $Q_{\lambda,k}$ have asymptotic level α only if the true Ω is used in their construction. In this section, we show that under fairly weak conditions on the growth rate of the number of IVs, k, these asymptotic power envelopes also apply when Ω is unknown. Obviously, an upper bound on asymptotic average power for known Ω is also an upper bound when Ω is unknown. Hence, to show that the upper bound is the power envelope, it suffices to show that it is attainable at each point by some sequence of tests.

We estimate Ω ($\in R^{2\times 2}$) via

$$\widehat{\Omega}_n = (n-k-p)^{-1}\widehat{V}'\widehat{V}, \text{ where } \widehat{V} = Y - P_Z Y - P_X Y. \tag{10.17}$$

We define analogues of S, T, and $Q_{\lambda,k}$ with Ω replaced by $\widehat{\Omega}_n$:

$$\begin{aligned}
\widehat{S}_n &= (Z'Z)^{-1/2}Z'Yb_0 \cdot (b_0'\widehat{\Omega}_n b_0)^{-1/2}, \\
\widehat{T}_n &= (Z'Z)^{-1/2}Z'Y\widehat{\Omega}_n^{-1}a_0 \cdot (a_0'\widehat{\Omega}_n^{-1}a_0)^{-1/2}, \\
\widehat{Q}_{\lambda,k,n} &= [\widehat{S}_n : \widehat{T}_n]'[\widehat{S}_n : \widehat{T}_n], \text{ and } \widehat{\overline{Q}}_{\lambda,k,n} = (\widehat{Q}_{\lambda,k,n} - kI_2)/k^{1/2}.
\end{aligned} \tag{10.18}$$

The LR, LM, AR, and POIS2 test statistics for the case of unknown Ω are defined in the same way as when Ω is known, but with $\widehat{Q}_{\lambda,k,n}$ in place of $Q_{\lambda,k}$. Denote these test statistics by $\widehat{LR}_n$, $\widehat{LM}_n$, $\widehat{AR}_n$, and $\widehat{LR}^*_n = LR^*(\widehat{\overline{Q}}_{\lambda,k,n};\beta^*,\lambda^*)$, respectively.

Consistency of $\widehat{\Omega}_n$ with rate $k^{1/2}$ is established in the following Lemma.

Lemma 1 *Suppose $\{V_i : i \geq 1\}$ are iid with mean zero, variance Ω, and finite fourth moment, and $k^{3/2}/n \to 0$. Then, $k^{1/2}(\widehat{\Omega}_n - \Omega) \to_p 0$ as $n \to \infty$.*

Lemma 1 can be used to show that $\widehat{Q}_{\lambda,k,n}$ and $Q_{\lambda,k}$ have the same asymptotic distributions.

Theorem 4 *Theorem 1 holds under the given assumptions with $(\widehat{S}_n, \widehat{T}_n)$ in place of (S,T) provided $(k,n) \to \infty$ such that $k^{3/2}/n \to 0$.*

Comments. 1. Theorem 4 and the continuous mapping theorem combine to show that when $\tau = 1/2$ and β is fixed, then

$$\widehat{LR}^*_n = LR^*(\widehat{\overline{Q}}_{\lambda,k,n};\beta^*,r^*_{1/2}) \to_d LR^*(\overline{Q}_\infty;\beta^*,r^*_{1/2}). \tag{10.19}$$

In consequence, the $\widehat{LR}_n^*$ and LR^* statistics are asymptotically equivalent under the null and fixed alternatives. Thus, the upper bound on asymptotic average power for $\tau = 1/2$ given in Theorem 2 is attained by the tests based on $\widehat{LR}_n^*$, which are asymptotically level α and similar, by varying $(\beta^*, r_{1/2}^*)$. In turn, this implies that the upper bound in Theorem 2 is the asymptotic power envelope whether or not Ω is known.

2. Similarly, Theorem 4 implies that when $\tau > 1/2$, $\widehat{LM}_n \to_d \overline{Q}_{ST,\infty}^2$ under the null hypothesis and local alternatives. Hence, the upper bound on average power given in Theorem 3 is attained for all B^* and $-B^*$ by the tests based on $\widehat{LM}_n$ and $\widehat{LR}_n$. The upper bound in Theorem 3 is the asymptotic power envelope whether or not Ω is known and the tests based on $\widehat{LM}_n$ and $\widehat{LR}_n$ are asymptotically efficient.

10.7 Asymptotics with Non-normal Errors

Andrews and Stock (2005) investigate the asymptotic properties of the statistic $\widehat{Q}_{\lambda,k,n}$ and the test statistics $\widehat{LR}_n$, $\widehat{LM}_n$, and $\widehat{AR}_n$ when the errors are not necessarily normally distributed. For the case of $\tau \in [0, 2]$, they show that one obtains the same limit distributions (given in Theorem 1) when the errors are non-normal and Ω is estimated as when the errors are normal and Ω is known provided $k^3/n \to 0$ as $n \to \infty$.

To conclude, the many weak IV results given in this section show that the significance level of the CLR test is completely robust to weak IVs. The test, however, is not completely robust to many IVs. One cannot employ too many IVs relative to the sample size. For normal errors, the CLR test has correct asymptotic significance level provided $k^{3/2}/n \to 0$ as $n \to \infty$ regardless of the strength of the IVs. For non-normal errors, the restriction is greater: $k^3/n \to 0$ as $n \to \infty$. The power results show that the CLR test is essentially on the two-sided power envelope for invariant tests for any value of τ when the errors are iid homoskedastic normal.

These level and power results established for many IV asymptotics, combined with the properties of the CLR test under weak IV asymptotics, lead us to recommend the CLR test (or heteroskedasticity and/or autocorrelation robust versions of it) for general use in scenarios where the IVs may be weak.

11 Appendix of Proofs

In this Appendix, we prove the results stated in Section 10.1.

Proof of Theorem 1. First, we determine the means, variances, and covariances of the components of Q. Let $S = (S_1, ..., S_k)'$, $\mu_S = ES = c_\beta \mu_\pi = (\mu_{S1}, ..., \mu_{Sk})'$, $S^* = S - \mu_S = (S_1^*, ..., S_k^*)' \sim N(0, I_k)$. Define T_j, μ_T, T^*, and T_j^* for $j = 1, ..., k$ analogously. We have

$$\begin{aligned}
ES'S &= \sum_{j=1}^k E(S_j^* + \mu_{Sj})^2 = \sum_{j=1}^k (1 + \mu_{Sj}^2) = k + c_\beta^2 \lambda, \\
E(S'S)^2 &= E(\sum_{j=1}^k S_j^2)^2 = \sum_{j=1}^k ES_j^4 + \sum^k \sum^k_{j \neq \ell} ES_j^2 ES_\ell^2 \\
&= \sum_{j=1}^k E(S_j^* + \mu_{Sj})^4 + \sum_{j=1}^k \sum_{\ell=1}^k (1 + \mu_{Sj}^2)(1 + \mu_{S\ell}^2) - \sum_{j=1}^k (1 + \mu_{Sj}^2)^2 \\
&= \sum_{j=1}^k (3 + 6\mu_{Sj}^2 + \mu_{Sj}^4) + \left(\sum_{j=1}^k (1 + \mu_{Sj}^2) \right)^2 - \sum_{j=1}^k (1 + \mu_{Sj}^2)^2 \\
&= 2k + 4c_\beta^2 \lambda + \left(ES'S \right)^2, \text{ and} \\
Var(S'S) &= 2(k + 2c_\beta^2 \lambda).
\end{aligned} \tag{11.1}$$

Analogously,

$$ET'T = k + d_\beta^2 \lambda \text{ and } Var(T'T) = 2(k + 2c_\beta^2 \lambda). \tag{11.2}$$

Next, we have

$$\begin{aligned}
ES'T &= ES'ET = c_\beta d_\beta \lambda, \\
E(S'T)^2 &= E(\sum_{j=1}^k S_j T_j)^2 = \sum_{j=1}^k ES_j^2 ET_j^2 + \sum^k \sum^k_{j \neq \ell} ES_j ET_j ES_\ell ET_\ell \\
&= \sum_{j=1}^k (1 + \mu_{Sj}^2)(1 + \mu_{Tj}^2) + \left(\sum_{j=1}^k ES_j ET_j \right)^2 - \sum_{j=1}^k (ES_j ET_j)^2 \\
&= \sum_{j=1}^k (1 + \mu_{Sj}^2 + \mu_{Tj}^2 + \mu_{Sj}^2 \mu_{Tj}^2) + (ES'T)^2 - \sum_{j=1}^k \mu_{Sj}^2 \mu_{Tj}^2 \\
&= k + (c_\beta^2 + d_\beta^2)\lambda + (ES'T)^2, \text{ and} \\
Var(S'T) &= k + (c_\beta^2 + d_\beta^2)\lambda.
\end{aligned} \tag{11.3}$$

Finally, we have

$$E(S'SS'T) = \sum_{j=1}^k \sum_{\ell=1}^k ES_j^2 ES_\ell ET_\ell$$

$$
\begin{aligned}
&= \sum_{j=1}^{k} ES_j^3 ET_j + \sum_{j=1}^{k}\sum_{\ell=1}^{k} ES_j^2 ES_\ell ET_\ell - \sum_{j=1}^{k} ES_j^2 ES_j ET_j \\
&= \sum_{j=1}^{k} E(S_j^* + \mu_{Sj})^3 \mu_{Tj} + ES'S \cdot ES'T - \sum_{j=1}^{k} ES_j^2 ES_j ET_j \\
&= \sum_{j=1}^{k} (3\mu_{Sj}\mu_{Tj} + \mu_{Sj}^3\mu_{Tj}) + ES'S \cdot ES'T - \sum_{j=1}^{k} (1+\mu_{Sj}^2)\mu_{Sj}\mu_{Tj}, \text{ and} \\
Cov(S'S, S'T) &= 2c_\beta d_\beta \lambda. \qquad (11.4)
\end{aligned}
$$

Using the Cramer-Wold device and the Liapounov CLT, we show that

$$
\begin{pmatrix} (S'S - (k + c_\beta^2\lambda))/(2(k+2c_\beta^2\lambda))^{1/2} \\ (S'T - c_\beta d_\beta \lambda)/(k + (c_\beta^2 + d_\beta^2)\lambda)^{1/2} \\ (T'T - (k + d_\beta^2\lambda))/(2(k+2d_\beta^2\lambda))^{1/2} \end{pmatrix} \to_d N(0, I_3) \qquad (11.5)
$$

under the conditions of the Theorem. The proof is as follows. Wlog we can take $\mu_\pi = (\lambda/k)^{1/2} 1_k$, where $1_k = (1, 1, ..., 1)' \in R^k$ (because the distribution of Q only depends on μ_π through λ). Then, $\mu_{Sj} = c_\beta(\lambda/k)^{1/2}$ and $\mu_{Tj} = d_\beta(\lambda/k)^{1/2}$ for all j. We have

$$
S'S - (k + c_\beta^2\lambda) = \sum_{j=1}^{k} ((S_j^* + \mu_{Sj})^2 - 1 - \mu_{Sj}^2) = \sum_{j=1}^{k} [(S_j^{*2} - 1) + 2\mu_{S1} S_j^*]. \qquad (11.6)
$$

By the Liapounov CLT, if $\{X_{kj} : j \leq k, k \geq 1\}$ is a triangular array of mean zero row-wise iid random variables with $\sum_{j=1}^{k} Var(X_{kj}) = 1$ and $kE|X_{kj}|^{2+2\delta} = o(1)$ for some $\delta > 0$, then $\sum_{j=1}^{k} X_{kj} \to_d N(0, 1)$, e.g., see Chow and Teicher (1978, Cor. 9.1.1, p. 293). If $X_{kj} = X_{kj1} + X_{kj2}$, where $\{X_{kjs} : j \leq k, k \geq 1\}$ are iid across j for $s = 1, 2$, then by Minkowski's inequality it suffices to show that $kE|X_{kjs}|^{2+2\delta} = o(1)$ for $s = 1, 2$. We apply this CLT with

$$
X_{kj1} = (S_j^{*2} - 1)/(2(k + 2c_\beta^2\lambda))^{1/2} \text{ and } X_{kj2} = 2\mu_{S1} S_j^*/(2(k + 2c_\beta^2\lambda))^{1/2}. \qquad (11.7)
$$

We have

$$
\begin{aligned}
kE|X_{kj1}|^{2+2\delta} &= \frac{kE|S_j^{*2} - 1|^{2+2\delta}}{(2(k + 2c_\beta^2\lambda))^{1+\delta}} = o(1) \text{ as } k \to \infty \text{ and} \\
kE|X_{kj2}|^{2+2\delta} &= \frac{kE|2\mu_{S1} S_j^*|^{2+2\delta}}{(2(k + 2c_\beta^2\lambda))^{1+\delta}} = \frac{k(\lambda/k)^{1+\delta}|2c_\beta|^{2+2\delta} E|S_j^*|^{2+2\delta}}{(2(k + 2c_\beta^2\lambda))^{1+\delta}} \qquad (11.8)
\end{aligned}
$$

using the fact that $S_j^* \sim N(0, 1)$ has all moments finite. The term $kE|X_{kj2}|^{2+2\delta}$ is $o(1)$ if $\tau < 1/2$ because then c_β is fixed. It is $o(1)$ when $1/2 < \tau \leq 1$ because then $\lambda/k = O(1)$ and $c_\beta = O(1)$. It is $o(1)$ when $\tau > 1$ because then $c_\beta \propto k^{-\tau/2}$ and $(\lambda/k)^{1+\delta}|c_\beta|^{2+2\delta} = O((\lambda/k)^{1+\delta} k^{-\tau(1+\delta)}) = O((\lambda/k^\tau)^{1+\delta} k^{-(1+\delta)}) = o(1)$. Hence, the first element in (11.5) is asymptotically $N(0,1)$. Analogous arguments apply to the second and third elements in (11.5).

To obtain the joint result in (11.5), we consider an arbitrary linear combination, say $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ with $||\alpha|| = 1$, of the three terms in (11.5) and apply the above CLT. We have

$$\frac{Cov(S'S, S'T)}{Var^{1/2}(S'S)Var^{1/2}(S'T)} = \frac{2c_\beta d_\beta \lambda}{(2(k+2c_\beta^2\lambda))^{1/2}(k+(c_\beta^2+d_\beta^2)\lambda)^{1/2}}. \tag{11.9}$$

The rhs converges to zero because (i) $\lambda/k \to 0$ when $\tau < 1$ and (ii) $c_\beta = O(k^{-\tau/2}) = o(1)$ when $\tau \geq 1$. The same result holds with $S'S$ replaced by $T'T$. Hence, the asymptotic covariances between the three terms in (11.5) are all zero. In consequence, although the variance of the inner product of α with the three terms in (11.5) is not one, it converges to one as $k \to \infty$ (which is sufficient for the CLT by rescaling). We establish the Liapounov condition $kE|X_{kj}|^{2+2\delta} = o(1)$ for the linear combination by the same method as above. This concludes the proof of (11.5).

Now, using (11.5), the first result in parts (a), (b), (c), and (d) of the Theorem hold because (a) when $\tau < 1/2$, $\lambda/k^{1/2} \to 0$, (b) when $\tau = 1/2$, $\lambda/k \to 0$ and $\lambda/k^{1/2} \to r_{1/2}$, (c) when $1/2 < \tau \leq 1$, (11.5) implies that $(T'T-k)/k^\tau \to_p d_{\beta_0}^2 r_\tau$, $c_\beta d_\beta \lambda/k^{1/2} = (b_0'\Omega b_0)^{-1/2} B k^{1/2-\tau} d_\beta \lambda/k^{1/2} \to \gamma_B r_\tau$, and $c_\beta^2\lambda/k^{1/2} = \gamma_B^2 k^{1-2\tau}\lambda/k^{1/2} = o(1)$; when $1/2 < \tau < 1$, $(k + (c_\beta^2 + d_\beta^2)\lambda)/k \to 1$; when $\tau = 1$, $(k + (c_\beta^2 + d_\beta^2)\lambda)/k \to 1 + d_{\beta_0}^2 r_1$, and (d) when $\tau > 1$ and $r_\tau > 0$, (11.5) implies that $(T'T-k)/k^\tau \to_p d_{\beta_0}^2 r_\tau$, $(k+(c_\beta^2+d_\beta^2)\lambda)/\lambda \to d_{\beta_0}^2$, $c_\beta d_\beta \lambda^{1/2} = (b_0'\Omega b_0)^{-1/2} B k^{-\tau/2} d_\beta \lambda^{1/2} = (b_0'\Omega b_0)^{-1/2} B d_\beta (\lambda/k^\tau)^{1/2} \to \gamma_B r_\tau^{1/2}$, and $c_\beta^2 \lambda^{1/2} \to 0$.

Next, we establish the results for the AR, LM, and LR statistics. The results of parts (a)-(d) for AR hold because $(AR-1)k^{1/2} = (S'S-k)/k^{1/2}$. For parts (a)-(d) of the Theorem, we have

$$\begin{aligned}
\text{(a) \& (b) } LM &= \frac{(S'T)^2}{T'T} = \frac{(S'T/k^{1/2})^2}{k^{-1/2}\left(\frac{T'T-k}{k^{1/2}}\right)+1} = (S'T/k^{1/2})^2(1+o_p(1)),\\
\text{(c) } LM &= \frac{(S'T/k^{1/2})^2}{k^{\tau-1}\left(\frac{T'T-k}{k^\tau}\right)+1} = \begin{cases} (S'T/k^{1/2})^2(1+o_p(1)) & \text{if } 1/2<\tau<1 \\ \frac{(S'T/k^{1/2})^2}{d_{\beta_0}^2 r_1 + 1 + o_p(1)} & \text{if } \tau = 1 \end{cases}\\
\text{(d) } LM &= \frac{(S'T/k^{\tau/2})^2}{\left(\frac{T'T-k}{k^\tau}\right)+k^{1-\tau}} = \frac{(S'T/k^{\tau/2})^2}{d_{\beta_0}^2 r_\tau + o_p(1)} \to_d (Q_{ST,\infty}^2/(d_{\beta_0} r_\tau^{1/2}))^2.
\end{aligned} \tag{11.10}$$

Combining these expressions with the asymptotic results for Q given in the Theorem gives the stated asymptotic results for LM.

For the LR statistic, we use (7.18) and write

$$LR = \frac{1}{2}\left((Q_S-k)-(Q_T-k)+\sqrt{((Q_S-k)-(Q_T-k))^2+4Q_{ST}^2}\right). \tag{11.11}$$

The results of parts (a) and (b) follow from (11.11) by dividing through by $k^{1/2}$ and applying the results of parts (a) and (b) for the asymptotic distribution of Q.

Next, for the case where $\tau > 1/2$, by parts(c) and (d) for the asymptotic distribution of Q, we have

$$(T'T - k)k^{-\tau} \to_p d_{\beta_0}^2 r_\tau \text{ and } (S'S - k)k^{-\tau} \to_p 0, \text{ and so} \tag{11.12}$$
$$\frac{Q_T}{(Q_T - Q_S)^2} = \frac{(Q_T - k)k^{-2\tau} + k^{1-2\tau}}{((Q_T - k)k^{-\tau} - (Q_S - k)k^{-\tau})^2} = \frac{o_p(1)}{(d_{\beta_0}^2 r_\tau + o_p(1))^2} = o_p(1).$$

By a mean-value expansion $\sqrt{1+x} = 1 + (1/2)x(1 + o(1))$ as $x \to 0$. Hence,

$$\begin{aligned} LR &= \frac{1}{2}\left(Q_S - Q_T + |Q_T - Q_S|\sqrt{1 + \frac{4Q_T}{(Q_T - Q_S)^2}LM}\right) \\ &= \frac{1}{2}\left(Q_S - Q_T + |Q_T - Q_S|\left(1 + \frac{2Q_T(1 + o_p(1))}{(Q_T - Q_S)^2}LM\right)\right) \\ &= \frac{Q_T(1 + o_p(1))}{Q_T - Q_S}LM, \end{aligned} \tag{11.13}$$

where the third equality uses $|Q_T - Q_S| = Q_T - Q_S$ with probability that goes to one by the calculation in the denominator of (11.12). By results of parts (c) and (d) for Q, we have

$$\begin{aligned} \frac{Q_T}{Q_T - Q_S} &= \frac{(Q_T - k)k^{-\tau} + k^{1-\tau}}{(Q_T - k)k^{-\tau} - (Q_S - k)k^{-\tau}} = \frac{d_{\beta_0}^2 r_\tau + o_p(1) + k^{1-\tau}}{d_{\beta_0}^2 r_\tau + o_p(1)} \\ &= \begin{cases} 1 + o_p(1) & \text{if } \tau > 1 \\ \frac{1 + d_{\beta_0}^2 r_1}{d_{\beta_0}^2 r_1} + o_p(1) & \text{if } \tau = 1 \\ \frac{1 + o_p(1)}{d_{\beta_0}^2 r_\tau} k^{1-\tau} & \text{if } 1/2 < \tau < 1. \end{cases} \end{aligned} \tag{11.14}$$

Equations (11.13) and (11.14) combine to give the results for LR stated in parts (c) and (d) of the Theorem. □

Proof of Lemma 1. Let $\widehat{\Omega}_{n,rs}$ and Ω_{rs} denote the (r,s) elements of $\widehat{\Omega}_n$ and Ω, respectively, for $r, s = 1, 2$. Let $n_k = n - k - p$. We have

$$\begin{aligned} k^{1/2}(\widehat{\Omega}_{n,rs} - \Omega_{rs}) = & \frac{k^{1/2}}{n_k}(v_r'v_s - n\Omega_{rs}) - \frac{k^{1/2}}{n_k}v_r'P_Z v_s - \frac{k^{1/2}}{n_k}v_r'P_X v_s \\ & + k^{1/2}(1 - \frac{n}{n_k})\Omega_{rs}. \end{aligned} \tag{11.15}$$

Next, we have

$$0 \le \frac{k^{1/2}}{n_k}Ev_r'P_Z v_s = \frac{k^{1/2}}{n_k}tr(P_Z E v_r v_s') = \frac{k^{1/2}}{n_k}tr(P_Z)\Omega_{rs} = \frac{k^{3/2}}{n_k}\Omega_{rs} \to 0 \tag{11.16}$$

provided $k^{3/2}/n \to 0$. L^1-convergence implies convergence in probability. Hence, $(k^{1/2}/n_k)v_r'P_Z v_s \to_p 0$. Analogously, $(k^{1/2}/n_k)v_r'P_X v_s \to_p 0$. In addition, $k^{1/2}(1 - n/n_k) \to 0$.

Lastly, by Markov's inequality, for any $\varepsilon > 0$,

$$P\left(\frac{k^{1/2}}{n_k}\left|v_r' v_s - n\Omega_{rs}\right| > \varepsilon\right) \leq \frac{kE(\sum_{i=1}^n (v_{i,r}v_{i,s} - \Omega_{rs}))^2}{n_k^2 \varepsilon^2} = \frac{knE(v_{i,r}v_{i,s} - \Omega_{rs})^2}{n_k^2 \varepsilon^2} \to 0 \tag{11.17}$$

provided $k/n \to 0$. The above results combine to prove the Lemma. □

Proof of Theorem 4. It suffices to show that

$$\begin{aligned} &k^{-1/2}(\widehat{Q}_{\lambda,k,n} - kI_2) - k^{-1/2}(Q_{\lambda,k} - kI_2) \to_p 0 \text{ or} \\ &k^{-1/2}(\widehat{Q}_{\lambda,k,n} - Q_{\lambda,k}) \to_p 0. \end{aligned} \tag{11.18}$$

Using (7.13), we have

$$\begin{aligned} \widehat{S}_n' \widehat{S}_n - S'S &= b_0' Y' Z (Z'Z)^{-1} Z' Y b_0 \cdot [(b_0' \widehat{\Omega}_n b_0)^{-1} - (b_0' \Omega b_0)^{-1}] \text{ and} \\ k^{-1/2}(\widehat{S}_n' \widehat{S}_n - S'S) &= (S'S/k)\left(\frac{k^{1/2}(b_0'\Omega b_0 - b_0'\widehat{\Omega}_n b_0)}{b_0'\widehat{\Omega}_n b_0}\right) = o_p(1), \end{aligned} \tag{11.19}$$

where the last equality uses Theorem 1 and Lemma 1. Similar arguments, but with more steps because Ω enters T in two places, yield $k^{-1/2}(\widehat{S}_n' \widehat{T}_n - S'T) = o_p(1)$ and $k^{-1/2}(\widehat{T}_n' \widehat{T}_n - T'T) = o_p(1)$. This completes the proof. □

References

Anderson, T. W. (1976): "Estimation of Linear Functional Relationships: Approximate Distributions and Connections with Simultaneous Equations in Econometrics," *Journal of the Royal Statistical Society,* Ser. B, 38, 1-36.

Anderson, T. W., N. Kunitomo, and Y. Matsushita (2005): "A New Light from Old Wisdoms: Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models," Graduate School of Economics, University of Tokyo.

Anderson, T. W. and H. Rubin (1949): "Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations," *Annals of Mathematical Statistics,* 21, 570-582.

Andrews, D. W. K. and V. Marmer (2004): "Exactly Distribution-Free Inference in Instrumental Variables Regression with Possibly Weak Instruments," Cowles Foundation Discussion Paper No. 1501, Yale University. Available at http://cowles.econ.yale.edu.

Andrews, D. W. K., M. J. Moreira, and J. H. Stock (2004a): "Optimal Invariant Similar Tests for Instrumental Variables Regression with Weak Instruments," Cowles Foundation Discussion Paper No. 1476, Yale University. Available at http://cowles.econ.yale.edu.

——— (2004b): "Supplement to 'Optimal Invariant Similar Tests for Instrumental Variables Regression'," Department of Economics, Harvard University. Available at http://ksghome.harvard.edu/∼.JStock. Academic.Ksg/ams/websupp/index.htm.

——— (2006): "Performance of Conditional Wald Tests in IV Regression with Weak Instruments," *Journal of Econometrics,* forthcoming.

Andrews, D. W. K. and G. Soares (2004): "Rank Tests for Instrumental Variables Regression," Cowles Foundation, Yale University.

Andrews, D. W. K. and J. H. Stock (2005): "Inference with Weak Instruments," Cowles Foundation Discussion Paper No. 1530, Yale University. Available at http://cowles.econ.yale.edu.

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

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica,* 62, 657-681.

Bound, J. D. A. Jaeger, and R. M. Baker (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association,* 90, 443-450.

Caner, M. (2003): "Exponential Tilting with Weak Instruments: Estimation and Testing," Department of Economics, University of Pittsburgh.

Chamberlain, G. (2005): "Decision Theory Applied to an Instrumental Variables Model," Department of Economics, Harvard University.

Chamberlain, G. and G. Imbens (2004): "Random Effects Estimators with Many Instrumental Variables, *Econometrica,* 72, 295-306.

Chao, J. C. and P. C. B. Phillips (1998): "Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model Using the Jeffreys Prior," *Journal of Econometrics,* 87, 49-86.

——— (2002): "Jeffreys Prior Analysis of the Simultaneous Equations Model in the Case with $n+1$ Endogenous Variables," *Journal of Econometrics,* 111, 251-283.

Chao, J. C. and N. R. Swanson (2003): "Alternative Approximations of the Bias and MSE of the IV Estimator under Weak Identification with an Application to Bias Correction," Cowles Foundation Discussion Paper No. 1418, Yale University.

——— (2005): "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica,* 73, 1673-1692.

Chioda, L. and M. Jansson (2004): "Optimal Inference for Instrumental Variables Regression," Department of Economics, University of California, Berkeley.

Choi, I. and P. C. B. Phillips (1992): "Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations," *Journal of Econometrics,* 51, 113-150.

Chow Y.-C. and H. Teicher (1978): *Probability Theory.* New York: Springer.

Donald, S. G. and W. K. Newey (2001): "Choosing the Number of Instruments," *Econometrica,* 69, 1161-1191.

Dufour, J.-M. (1997): "Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica,* 65, 1365-1387.

Dufour, J.-M. (2003): "Identification, Weak Instruments, and Statistical Inference in Econometrics," *Canadian Journal of Economics,* 36, 767-808.

Dufour, J.-M. and J. Jasiak (2001): "Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors," *International Economic Review*, 42, 815-843.

Dufour, J.-M. and M. Taamouti (2005): "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica,* 73, 1351-1365.

Forchini, G. and G. Hillier (2003): "Condtional Inference for Possibly Unidentified Structural Equations," *Econometric Theory,* 19, 707-743.

Fuller, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica,* 45, 939-954.

Gao, C. and K. Lahiri (1999): "A Comparison of Some Recent Bayesian and Classical Procedures for Simulaneous Equation Models with Weak Instruments," Department of Economics, SUNY Albany.

Gleser, L. J. and J. T. Hwang (1987): "The Nonexistence of 100(1-α)% Confidence Sets of Finite Expected Diameter in Errors-in-Variables and Related Models," *Annals of Statistics,* 15, 1351-1362.

Godfrey, L. (1999): "Instrument Relevance in Multivariate Linear Models," *Review of Economics and Statistics,* 81, 550-552.

Guggenberger, P. and R. J. Smith (2005a): "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory,* 21, forthcoming.

——— (2005b): "Generalized Empirical Likelihood Tests in Time Series Models with Potential Identification Failure," Department of Economics, UCLA.

Hahn, J. (2002): "Optimal Inference with Many Instruments," *Econometric Theory,* 18, 140-168.

Hahn, J. and J. Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables," *Econometrica,* 70, 163-189.

——— (2003a): "IV Estimation with Valid and Invalid Instruments," Department of Economics, M.I.T.

——— (2003b): "Weak Instruments: Diagnosis and Cures in Empirical Economics," *American Economic Review*, 93, 118-125.

Hahn, J., J. Hausman, and G. Kuersteiner (2004): "Estimation with Weak Instruments: Accuracy of Higher-order Bias and MSE Approximations," *Econometrics Journal,* 7, 272-306.

Hall, A. R., G. D. Rudebusch, and D. W. Wilcox (1996): "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review,* 37, 283-298.

Hall, R. E. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy,* 86, 971-987.

Han, C. and P. C. B. Phillips (2005): "GMM with Many Moment Conditions," *Econometrica,* 73, forthcoming.

Hansen, C., J. Hausman, and W. K. Newey (2005): "Estimation with Many Instrumental Variables," Department of Economics, M.I.T.

Hausman, J., J. H. Stock, and M. Yogo (2005): "Asymptotic Properties of the Hahn-Hausman Test for Weak Instruments," *Economics Letters,* forthcoming.

Heckman, J. J. and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Analysis," *Econometrica,* 73, 669-738.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica,* 62, 467-476.

Kiviet, J. F. and J. Niemczyk (2005): "The Asymptotic and Finite Sample Distributions of OLS and IV in Simultaneous Equations," Discussion Paper 2005/01, Department of Quantitative Economics, University of Amsterdam.

Kleibergen, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica,* 70, 1781-1803.

——— (2004): "Testing Subsets of Structural Parameters in the Instrumental Variables Regression Model," *Review of Economics and Statistics,* 86, 418-423.

Kleibergen, F. (2005a): "Generalizing Weak Instrument Robust IV Statistics Towards Multiple Parameters, Unrestricted Covariance Matrices and Identification Statistics," Departent of Economics, Brown University.

——— (2005b): "Testing Parameters in GMM without Assuming That They Are Identified," *Econometrica,* 73, 1103-1123.

Kleibergen, F. and H. K. van Dijk (1998): "Bayesian Simultaneous Equation Analysis Using Reduced Rank Structures," *Econometric Theory,* 14, 701-743.

Kleibergen, F. and E. Zivot (2003): "Bayesian and Classical Approaches to Instrumental Variable Regression," *Journal of Econometrics,* 114, 29-72.

Kunitomo, N. (1980): "Asymptotic Expansions of the Distribution of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association,* 75, 693-700.

Lehmann, E. L. (1986): *Testing Statistical Hypotheses.* New York: Wiley.

Maddala, G. S. and J. Jeong (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica,* 60, 181-183.

Mavroeidis, S. (2004): "Weak Identification of Forward-looking Models in Monetary Models," *Oxford Bulletin of Economics and Statistics,* 66, Supplement, 609-635.

Moreira, M. J. (2001): "Tests with Correct Size When Instruments Can Be Arbitrarily Weak," Department of Economics, University of California, Berkeley.

——— (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica,* 71, 1027-1048.

——— (2005): "Towards a General Theory of Hypothesis Testing for Structural Models, Department of Economics, Harvard University.

Moreira, M. J., J. R. Porter, and G. Suarez (2004): "Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak," Department of Economics, Harvard University.

Morimune, K. (1983): "Approximate Distributions of the k-class Estimators when the Degree of Overidentifiability Is Large Compared with the Sample Size," *Econometrica,* 51,821-841.

Nagar, A. L. (1959): "The Bias and Moment Matrix of the General k-class Estimators of the Parameters in Simultaneous Equations," *Econometrica,* 27, 575-595.

Nason J. M. and G. W. Smith (2005): "Identifying the New Keynesian Phillips Curve," Working Paper 2005-1, Federal Reserve Bank of Atlanta.

Neeley, C. J., A. Roy, and C. H. Whiteman (2001): "Risk Aversion Versus Intertemporal Substitution: A Case Study of Identification Failure in the Intertemporal Consumption Capital Asset Pricing Model," *Journal of Business and Economic Statistics,* 19, 395-403.

Nelson, C. R. and R. Startz (1990a): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica,* 58, 967-976.

——— (1990b): "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One," *Journal of Business,* 63, S125-S140.

Newey, W. K. and F. Windmeijer (2005): "Many Weak Moment Asymptotics for Generalized Empirical Likelihood Estimators," Department of Economics, M.I.T.

Otsu, T. (2005): "Generalized Empirical Likelihood under Weak Identification," *Econometric Theory,* 21, forthcoming.

Phillips, P. C. B. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics,* Vol. 2. Amsterdam: North Holland.

——— (1989): "Partially Identified Econometric Models," *Econometric Theory,* 5, 181-240.

Poskitt, D. S. and C. L. Skeels (2005): "Small Concentration Asymptotics and Instrumental Variables Inference," Department of Economics, University of Melbourne.

Rothenberg, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics,* Vol. 2. Amsterdam: North Holland.

Shea, J. (1997): "Instrument Relevance in Multivariate Linear Models: A Simple Measure," *Review of Economics and Statistics,* 79, 348-352.

Staiger, D. and J. H. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica,* 65, 557-586.

Stock, J. H. and J. H. Wright (2000): "GMM with Weak Instruments," *Econometrica,* 68, 1055-1096.

Stock, J. H. and M. Yogo (2005a): "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," in *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg,* ed. by D. W. K. Andrews and J. H. Stock. Cambridge, UK: Cambridge University Press.

——— (2005b): "Testing for Weak Instruments in Linear IV Regression," in *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg,* ed. by D. W. K. Andrews and J. H. Stock. Cambridge, UK: Cambridge University Press.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics,* 20, 518-529.

Wang, J. and E. Zivot (1998): "Inference on Structural Parameters in Instrumental Variables Regression with Weak Instruments," *Econometrica,* 66, 1389-1404.

Woglom, G. (2001): "More Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica,* 69, 1381-1389.

Yogo, M. (2004): "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economic Studies,* 86, 797-810.

Zellner, A. (1998): "The Finite Sample Properties of Simultaneous Equations' Estimates and Estimators: Bayesian and Non-Bayesian Approaches," *Journal of Econometrics,* 83, 185-212.

Zivot, E., R. Startz, and C. R. Nelson (1998): "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review,* 39, 1119-1144.

Figure 1. Asymptotically efficient two-sided power envelopes for invariant similar tests and power functions for the two-sided CLR, LM, and AR tests, $k = 5$, $\rho = .95, .5, .2$

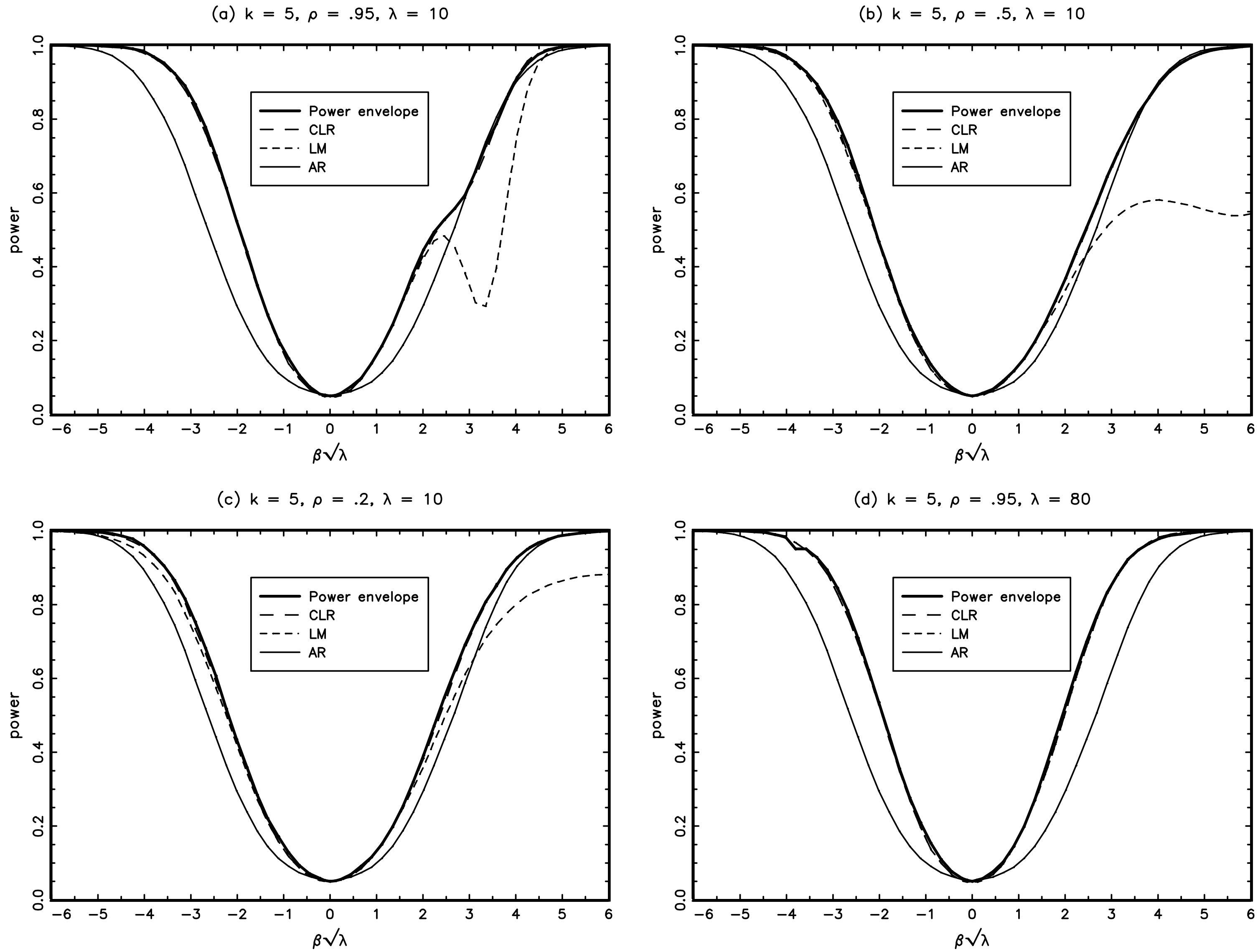

Figure 2. Many-instrument $\tau = ½$ limiting power envelope and power functions of the CLR, LM, and AR tests, $\rho = .95$

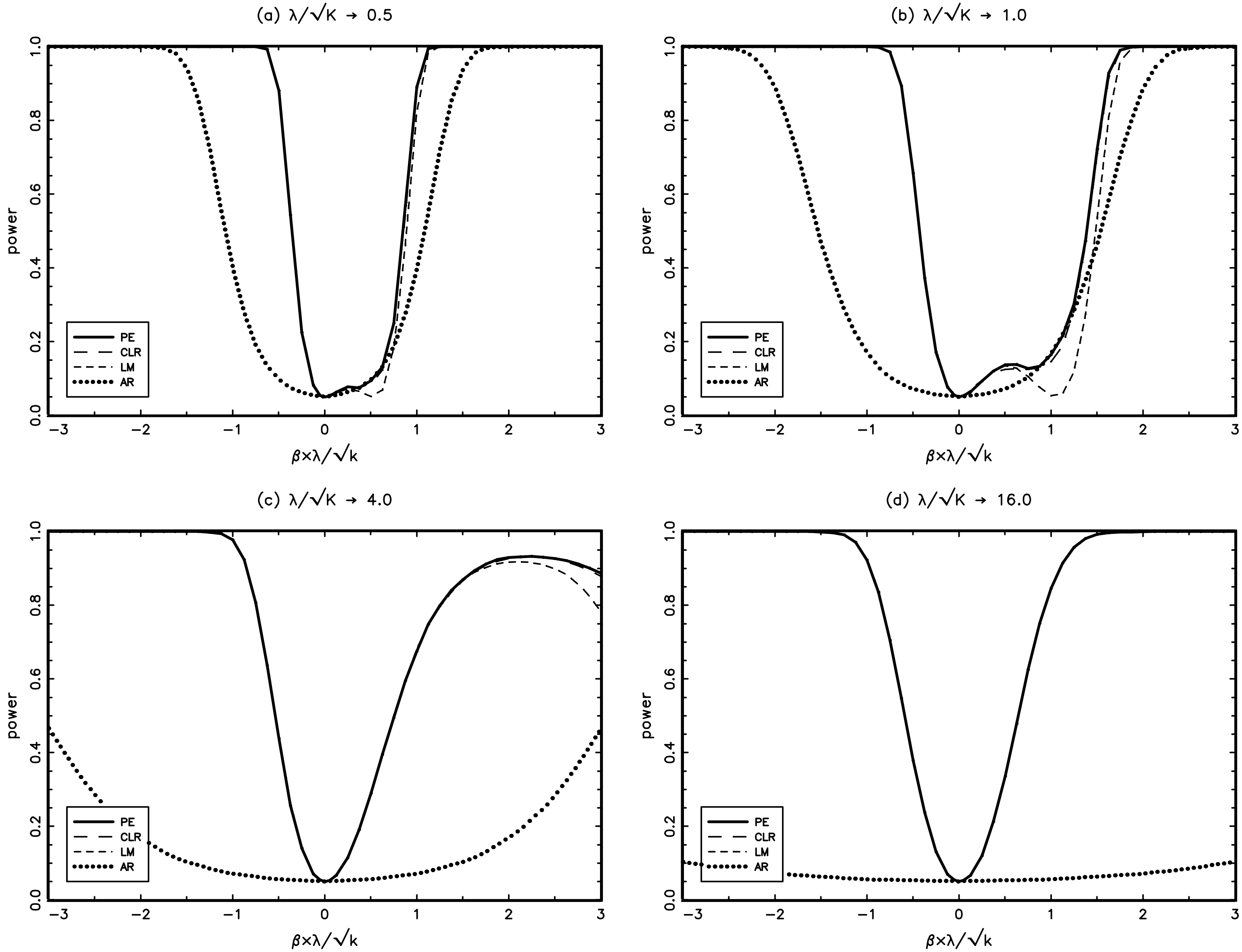

Figure 3. Many-instrument $\tau = ½$ limiting power envelope and power functions of the CLR, LM, and AR tests, $\rho = .5$

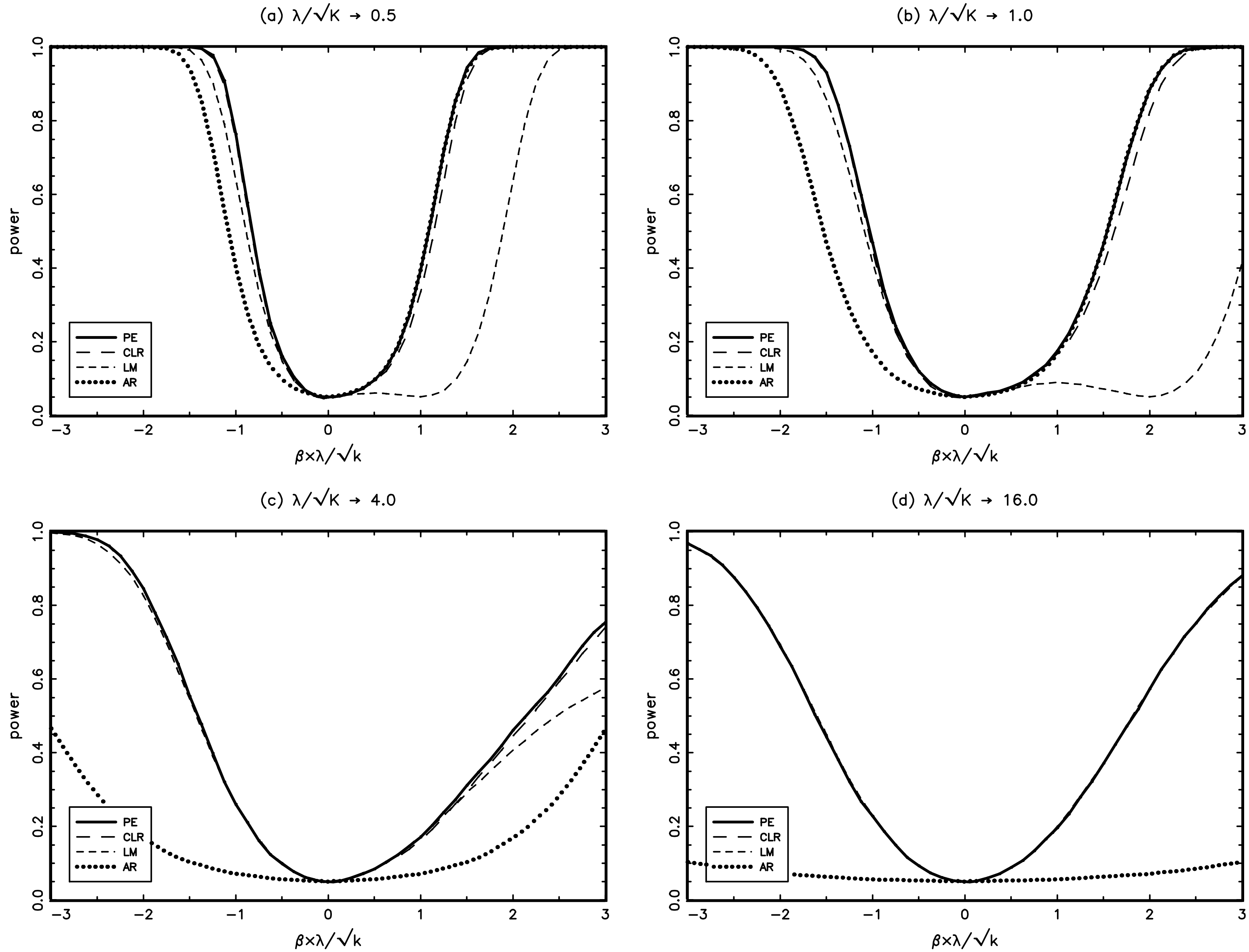

Figure 4. Convergence of the power envelope (solid line) and the CLR power function (dashed line) to the $\tau = ½$, $K \rightarrow \infty$ limit, $K = 10$ (bottom pair), 40, 160, and ∞ (top pair), $\rho = .95$

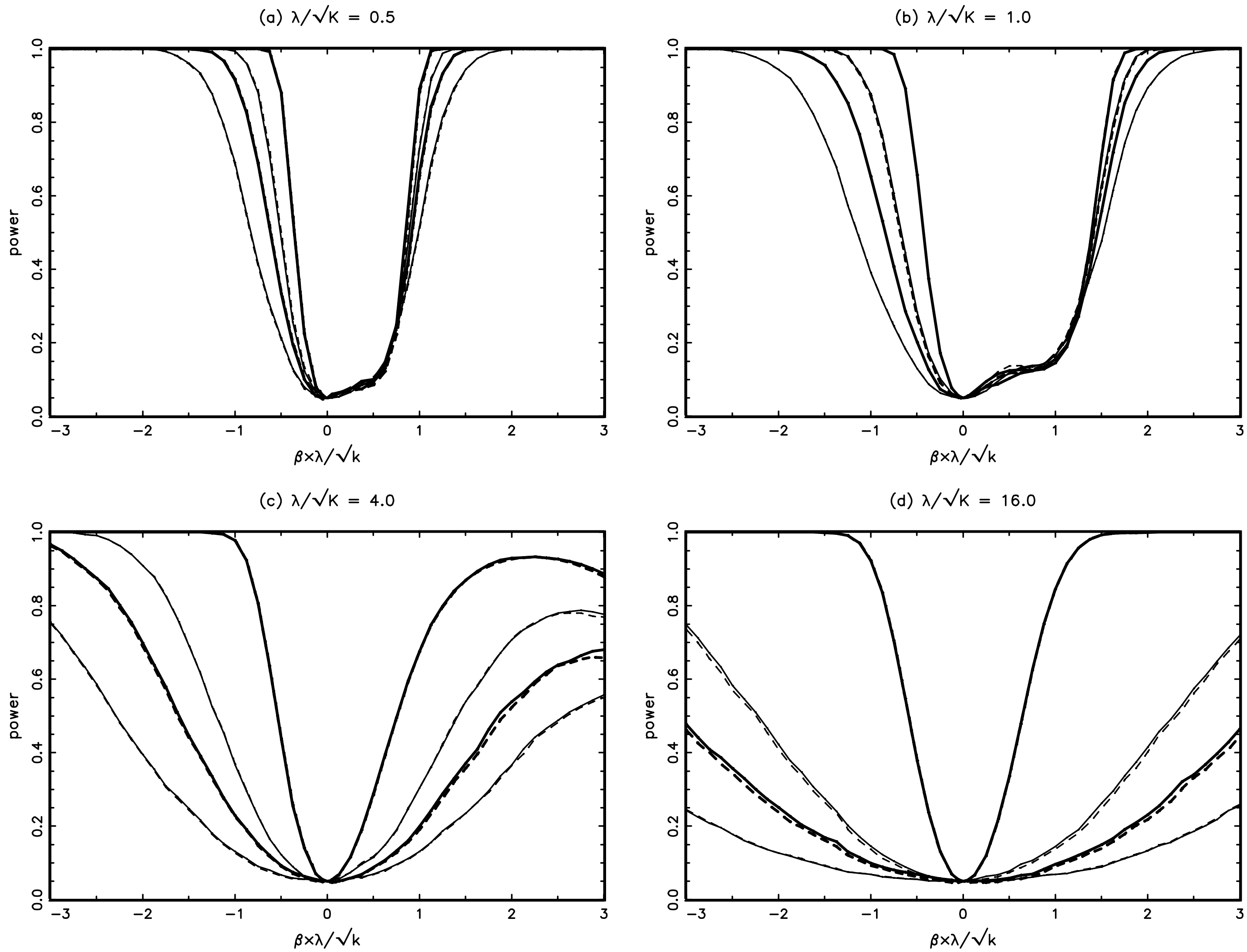

Figure 5. Convergence of the power envelope (solid line) and the CLR power function (dashed line) to the $\tau = ½$, $K \to \infty$ limit: $K = 10$ (bottom pair), 40, 160, and ∞ (top pair), $\rho = .5$

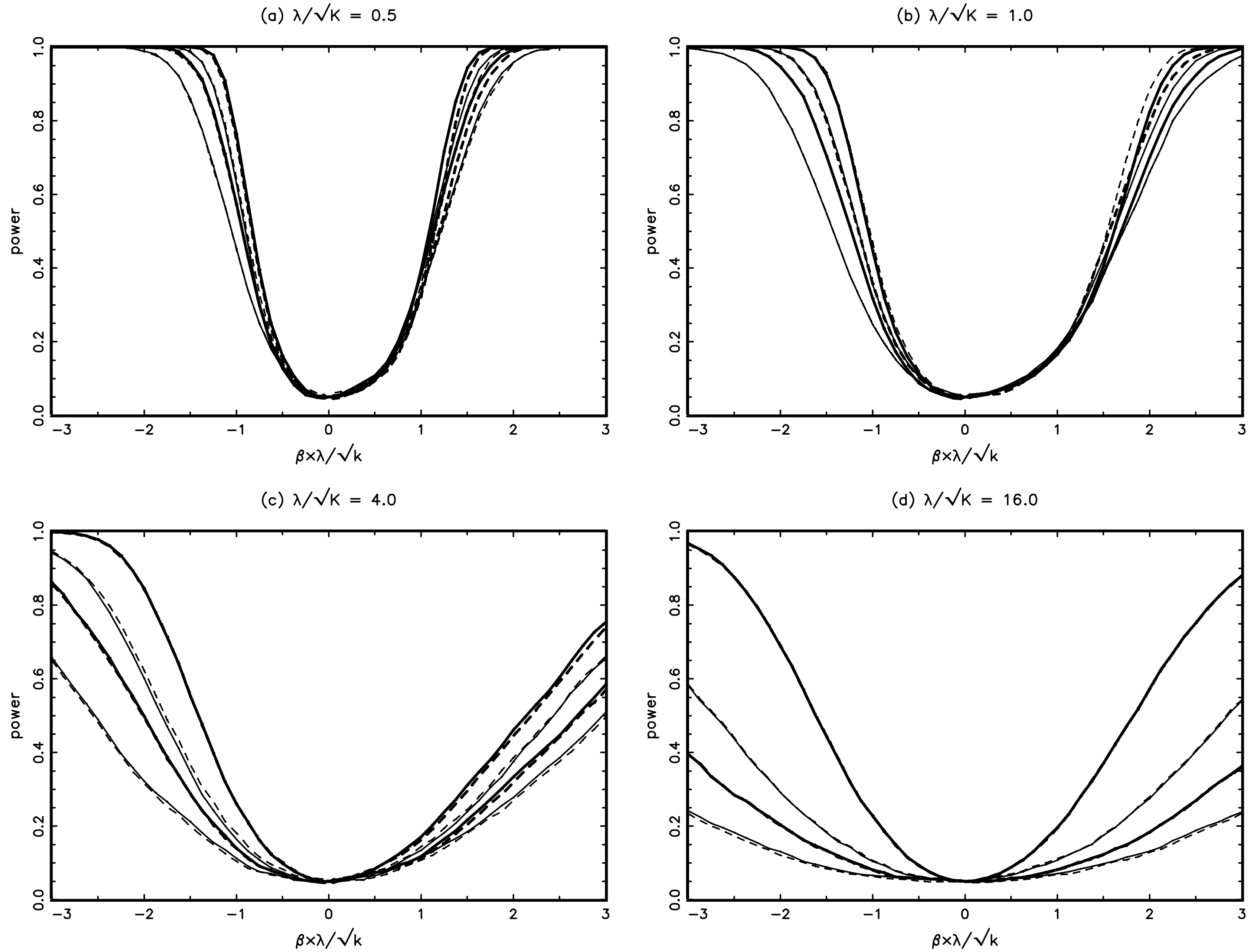

Figure 6. Effect of irrelevant instruments ($\tau = 0$) on the CLR and AR tests, $\lambda = 80$ and $\rho = .95$

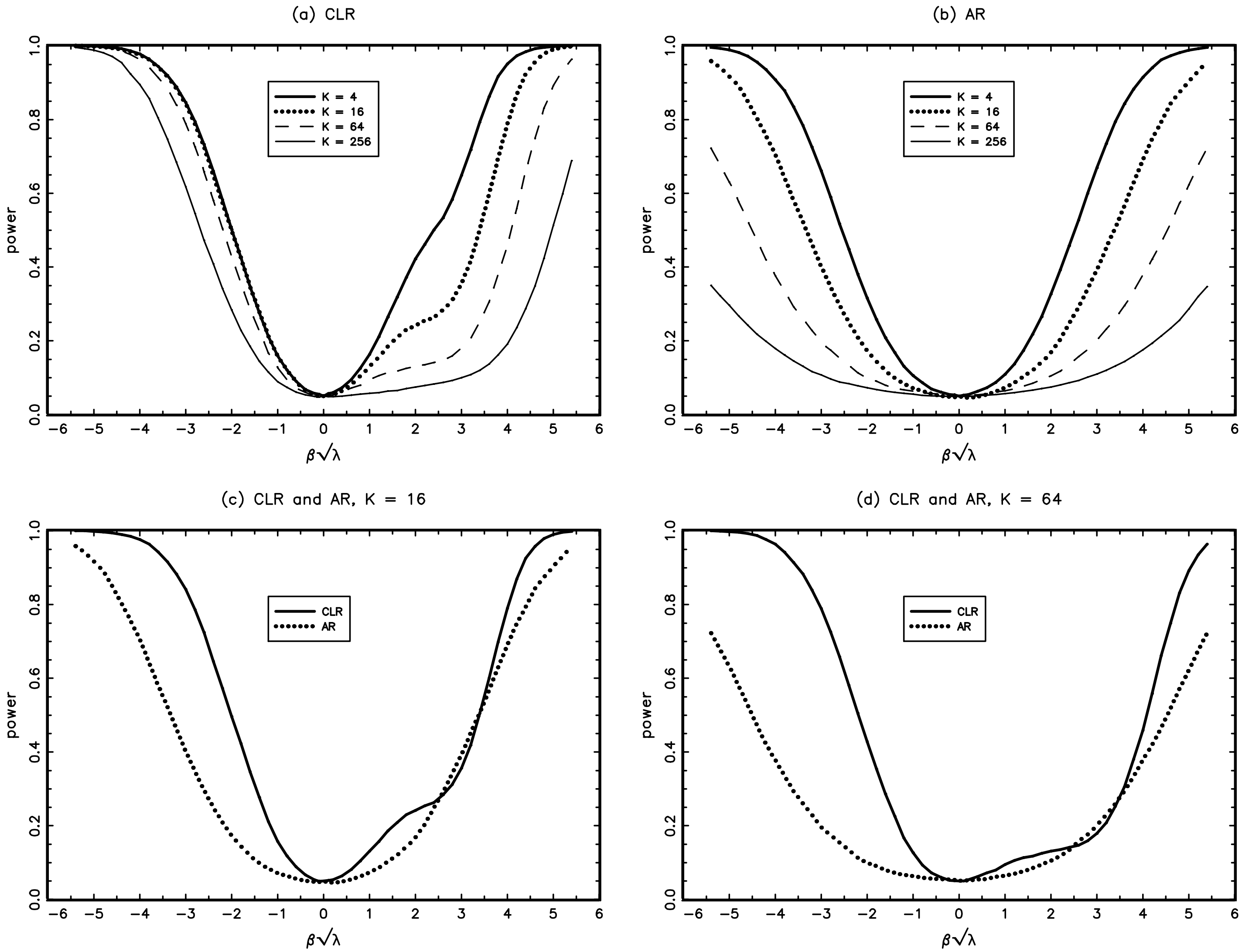

Figure 7. Effect of irrelevant instruments ($\tau = 0$) on the CLR and AR tests, $\lambda = 80$ and $\rho = .5$