Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://www.tandfonline.com/loi/ubes20

A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments

James H Stock, Jonathan H Wright & Motohiro Yogo

A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments

James H. Stock
Harvard University and the National Bureau of Economic Research, Cambridge, MA 02138

Jonathan H. Wright
Federal Reserve Board, Washington, DC 20551

Motohiro Yogo
Harvard University, Cambridge, MA 02138

Weak instruments arise when the instruments in linear instrumental variables (IV) regression are weakly correlated with the included endogenous variables. In generalized method of moments (GMM), more generally, weak instruments correspond to weak identification of some or all of the unknown parameters. Weak identification leads to GMM statistics with nonnormal distributions, even in large samples, so that conventional IV or GMM inferences are misleading. Fortunately, various procedures are now available for detecting and handling weak instruments in the linear IV model and, to a lesser degree, in nonlinear GMM.

KEY WORDS: Instrument relevance; Instrumental variables; Similar tests.

1. INTRODUCTION

A subtle but important contribution of Hansen and Singleton's (1982) and Hansen's (1982) original work on generalized method of moments (GMM) estimation was to recast the requirements for instrument exogeneity. In the linear simultaneous equations framework then prevalent, instruments are exogenous if they are excluded from the equation of interest; in GMM, instruments are exogenous if they satisfy a conditional mean restriction that, in Hansen and Singleton's (1982) application, was implied directly by a tightly specified economic model. Although these two requirements are the same mathematically, they have conceptually different starting points. The shift from debatable ["incredible," according to Sims (1980)] exclusion restrictions to first-order conditions derived from economic theory has been productive, and careful consideration of instrument exogeneity is now a standard part of solid empirical analysis using GMM.

But instrument exogeneity is only one of the two criteria necessary for an instrument to be valid. Recently, the other criterion—instrument relevance—has received increased attention by theoretical and applied researchers. It now appears that some, perhaps many, applications of GMM and instrumental variables (IV) regression have what is known as "weak instruments" or "weak identification," that is, instruments that are only weakly correlated with the included endogenous variables. Unfortunately, weak instruments pose considerable challenges to inference using GMM and IV methods.

This survey of weak instruments and weak identification has five themes:

1. If instruments are weak, then the sampling distributions of GMM and IV statistics are in general nonnormal, and standard GMM and IV point estimates, hypothesis tests, and confidence intervals are unreliable.

2. Empirical researchers often confront weak instruments. Finding exogenous instruments is hard work, and the features that make an instrument plausibly exogenous, such as occurring sufficiently far in the past to satisfy a first-order condition or the as-if random coincidence that lies behind a quasi-experiment, can also work to make the instrument weak.

3. It is not useful to think of weak instruments as a "small-sample" problem: Bound, Jaeger, and Baker (1995) provided an empirical example of weak instruments despite having 329,000 observations.

4. There are methods more robust to weak instruments than conventional GMM.

5. What to do about weak identification is a more difficult issue in nonlinear GMM than in linear IV regression, and much theoretical work remains.

This survey emphasizes the linear IV regression model with homoscedastic, serially uncorrelated errors, mainly because much more is known about weak instruments in this case. Section 2 provides a primer on weak instruments in linear IV regression, and Section 3 discusses some empirical applications that confront weak instruments. Sections 4–6 discuss recent econometric methods for handling weak instruments in the linear model with homoscedastic errors: detection of weak instruments (Sec. 4); methods that are fully robust to weak instruments, at least in large samples (Sec. 5); and partially robust methods that are somewhat simpler to use (Sec. 6). Section 7 turns to weak identification in GMM for nonlinear models and/or heteroscedastic or serially correlated errors. Section 8 concludes.

Journal of Business & Economic Statistics
October 2002, Vol. 20, No. 4
DOI 10.1198/073500102288618658

Although many of the key ideas of weak instruments have been understood for decades, most of the literature on solutions to the problem of weak instruments is quite recent, and this literature is expanding rapidly. We both fear and hope that much of the practical advice in this survey will soon be outdated.

2. A PRIMER ON WEAK INSTRUMENTS IN THE LINEAR REGRESSION MODEL

Many of the problems posed by weak instruments in the linear IV regression model are best explained in the context of the classical version of that model with fixed exogenous variables and iid normal errors. We therefore begin by using this model to show how weak instruments lead to the two-stage least squares (TSLS) estimator with a nonnormal sampling distribution, regardless of sample size. In general, exact distributions of IV statistics are not a practical basis for inference, and the section concludes with a synopsis of asymptotic methods designed to retain the insights gained from the finite-sample distribution theory.

2.1 The Linear Gaussian Instrumental Variables Regression Model With a Single Regressor

The linear IV regression model with a single endogenous regressor and no included exogenous variables is

$$y = Y\beta + u \tag{1}$$

and

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

where y and Y are $T \times 1$ vectors of observations on endogenous variables, Z is a $T \times K$ matrix of instruments, and u and v are $T \times 1$ vectors of disturbance terms. The instruments are assumed to be nonrandom (fixed). The errors $[u_t\ v_t]'$ $(t = 1, \ldots, T)$ are assumed to be iid $N(0, \Sigma)$, where the elements of Σ are σ_u^2, σ_{uv}, and σ_v^2, and let $\rho = \sigma_{uv}/(\sigma_u\sigma_v)$. Equation (1) is the structural equation, and β is the scalar parameter of interest. The reduced-form equation (2) relates the endogenous regressor to the instruments.

2.1.1. The Concentration Parameter. The concentration parameter, μ^2, is a unitless measure of the strength of the instruments and is defined as

$$\mu^2 = \Pi' Z' Z\Pi/\sigma_v^2. \tag{3}$$

A useful interpretation of μ^2 is in terms of F, the F statistic for testing the hypothesis $\Pi = 0$ in (2) (i.e., the "first-stage F statistic"). Let $\widetilde{F}$ be the infeasible counterpart of F, computed using the true value of σ_v^2. Then $K\widetilde{F}$ is distributed as a noncentral chi-squared distribution with degrees of freedom K and noncentrality parameter μ^2, and $E(\widetilde{F}) = \mu^2/K + 1$. If the sample size is large, then F and $\widetilde{F}$ are close, so $E(F) \cong \mu^2/K + 1$. Thus larger values of μ^2/K shift out the distribution of the first-stage F statistic, and $F - 1$ can be thought of as an estimator of μ^2/K.

2.1.2. An Expression for the Two-Stage Least Squares Estimator. The TSLS estimator is $\hat\beta^{\mathrm{TSLS}} = (Y'P_Z y)/(Y'P_Z Y)$, where $P_Z = Z(Z'Z)^{-1}Z'$. Rothenberg (1984) presented a useful expression for the TSLS estimator that obtains by substituting $Y'P_Z u = \Pi' Z' u + v' P_Z u$ and $Y'P_Z Y = \Pi' Z' Z\Pi + 2\Pi' Z' v + v' P_Z v$ into the expression for $\hat\beta^{\mathrm{TSLS}} - \beta$ and collecting terms,

$$\mu(\hat\beta^{\mathrm{TSLS}} - \beta) = (\sigma_u/\sigma_v)\frac{z_u + S_{uv}/\mu}{1 + 2z_v/\mu + S_{vv}/\mu^2}, \tag{4}$$

where $z_u = (\Pi' Z' u)/(\sigma_u\sqrt{\Pi' Z' Z\Pi})$, $z_v = (\Pi' Z' v)/(\sigma_v \times \sqrt{\Pi' Z' Z\Pi})$, $S_{uv} = (v' P_Z u)/(\sigma_v\sigma_u)$, and $S_{vv} = (v' P_Z v)/\sigma_v^2$. Under the assumptions of fixed instruments and normal errors, z_u and z_v are standard normal random variables with correlation ρ, and S_{uv} and S_{vv} are quadratic forms of normal random variables with respect to the idempotent matrix P_Z.

Because the distributions of z_u, z_v, S_{uv}, and S_{vv} do not depend on the sample size T, the sample size enters the distribution of $\hat\beta^{\mathrm{TSLS}}$ only through the concentration parameter. If μ^2 is small, then the terms z_v, S_{uv}, and S_{vv} in (4) lead to a nonnormal distribution. In contrast, the leading term z_u dominates if μ^2 is large, yielding the usual normal approximation to the distribution of $\hat\beta^{\mathrm{TSLS}}$. Formally, μ^2 plays the role in (4) usually played by the number of observations: As μ^2 becomes large, the distribution of $\mu(\hat\beta^{\mathrm{TSLS}} - \beta)$ is increasingly well approximated by the $N(0, \sigma_u^2/\sigma_v^2)$ distribution. For the normal approximation to the distribution of the TSLS estimator to be accurate, the concentration parameter must be large.

2.1.3. Bias of the Two-Stage Least Squares Estimator in the Unidentified Case. When $\mu^2 = 0$ (equivalently, when $\Pi = 0$), the instruments are not just weak, but irrelevant. In this case, the mean of the TSLS estimator is the probability limit of the ordinary least squares (OLS) estimator, $\mathrm{plim}(\hat\beta^{\mathrm{OLS}})$. Specifically, when $K \geq 3$ so that its mean exists, $E(\hat\beta^{\mathrm{TSLS}} - \beta) = \mathrm{plim}(\hat\beta^{\mathrm{OLS}} - \beta) = \sigma_{uY}/\sigma_Y^2$. To derive this result, note that when $\Pi = 0$, $\hat\beta^{\mathrm{TSLS}} - \beta = (v' P_Z u)/(v' P_Z v), \sigma_{uv} = \sigma_{uY}$, and $\sigma_v^2 = \sigma_Y^2$. Because $u = E(u|v) + \eta = (\sigma_{uv}/\sigma_v^2)v + \eta$ with η and v independent, $E(v' P_Z \eta|v) = 0$ and the result follows.

When the instruments are relevant but weak, the TSLS estimator is biased toward $\mathrm{plim}(\hat\beta^{\mathrm{OLS}})$. Specifically, define the "relative bias" of TSLS to be the bias of TSLS relative to the inconsistency of OLS, that is, $E(\hat\beta^{\mathrm{TSLS}} - \beta)/\mathrm{plim}(\hat\beta^{\mathrm{OLS}} - \beta)$. When μ^2 is moderately large, the TSLS relative bias is approximately inversely proportional to $\mu^2/(K - 2)$, a result that holds even if the errors are not normally distributed (Buse 1992).

2.1.4. Numerical Examples. Figures 1(a) and 1(b) show the pdf's of the TSLS estimator and its t statistic for various values of the concentration parameter when the true value of β is 0. The other parameter values mirror those of Nelson and Startz (1990a, b): $K = 1$, $\sigma_u = \sigma_v = 1$, and $\rho = .99$, so $\mathrm{plim}(\hat\beta^{\mathrm{OLS}}) = .99$. For small values of μ^2/K, such as Nelson and Startz's value of .25, the distributions are strikingly nonnormal, even bimodal. As μ^2/K increases, the distributions approach the usual normal limit.

The dramatic Nelson–Startz results drew econometricians' attention to the problem of weak instruments. Their results build on a large literature on the exact distribution of IV

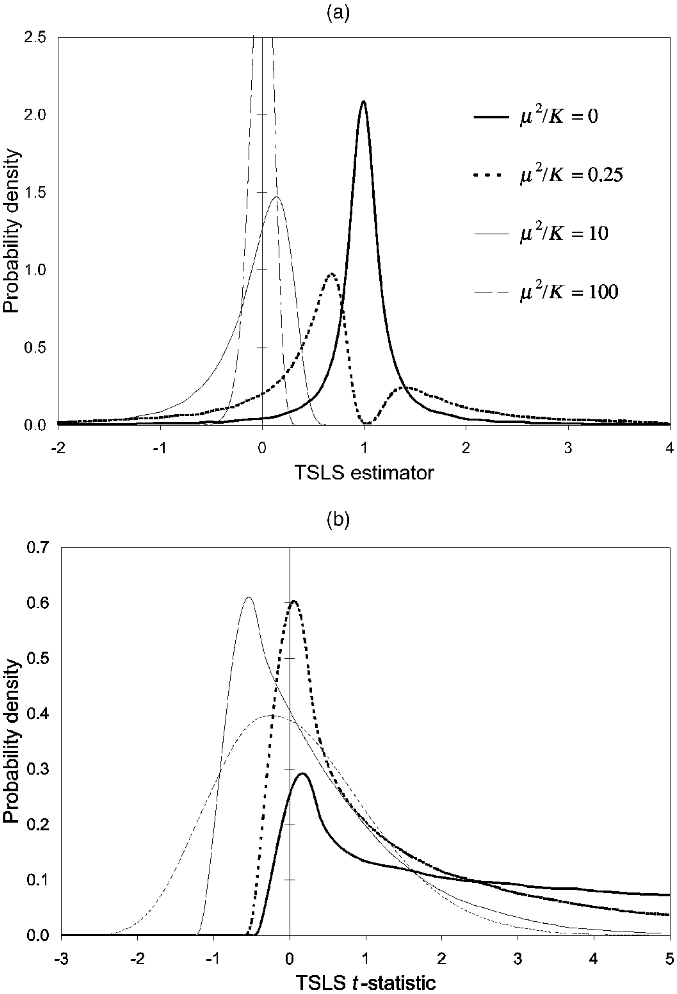

Figure 1. pdf of TSLS Estimator (a) and t Statistic (b) for $\mu^2/K = 0, .25, 10, 100$; One Instrument ($K = 1$); and $\rho = .99$, Computed by Monte Carlo Simulation.

estimators under the assumptions of fixed instruments and iid normal errors (e.g., Sawa 1969; Richardson 1968). However, the results in this literature, comprehensively reviewed by Phillips (1984), are offputting and pose substantial computational challenges. Moreover, the assumptions of fixed instruments and normal errors are generally too restrictive to be appropriate in empirical application. To overcome these limitations, researchers have used asymptotic approximations, to which we now turn.

2.2 Asymptotic Approximations

Conventional asymptotic approximations to finite-sample distributions are calculated for a fixed model in the limit that $T \to \infty$, but sometimes this approach does not provide the most useful approximating distribution. This is the case for the weak instruments problem; as is evident in Figure 1, the usual fixed-model asymptotic normal approximations can be quite poor when the concentration parameter is small, even if the number of observations is large. For this reason, alternative asymptotic methods are used to analyze IV statistics in the presence of weak instruments. Three such methods are Edgeworth expansions, many-instrument asymptotics, and weak-instrument asymptotics. These methods aim to improve the quality of the approximations when the sample is large but μ^2/K is not.

2.2.1. Edgeworth Expansions. An Edgeworth expansion is a representation of the distribution of the statistic of interest in powers of $1/\sqrt{T}$. As Rothenberg (1984) pointed out in the fixed-instrument, normal-error model, an Edgeworth expansion in $1/\sqrt{T}$ with a fixed model is formally equivalent to an Edgeworth expansion in $1/\mu$. In this sense, Edgeworth expansions improve on the conventional normal approximation when μ is small enough for the term in $1/\mu^2$ to matter, but not so small that the terms in $1/\mu^3$ and higher matter. Rothenberg (1984) suggested that the Edgeworth approximation is "excellent" for $\mu^2 > 50$ and "adequate" for μ^2 as small as 10, as long as the number of instruments is small (less than μ).

2.2.2. Many-Instrument Asymptotics. Although the problems of many instruments and weak instruments might at first seem different, they are in fact related. With many strong instruments, the adjusted R^2 of the first-stage regression would be nearly 1, so a small first-stage adjusted R^2 indicates that the instruments, taken as a set, are weak. Bekker (1994) formalized this notion by developing asymptotic approximations for a sequence of models with fixed instruments and normal errors, in which the number of instruments, K, is proportional to the sample size and μ^2/K converges to a constant, finite limit; similar approaches were taken by Anderson (1976), Kunitomo (1980), and Morimune (1983). Many-instrument asymptotic distributions are generally normal, and simulation evidence suggests that these approximations are good for both moderate and large values of K, although they cannot capture the nonnormality evident in the Nelson–Startz example of Figure 1. Distributions derived using this approach generally depend on the distribution of the errors (see Bekker and van der Ploeg 1999), so some procedures that are justified using many-instrument asymptotics require adjustments for nonnormal errors. However, rate and consistency results are more robust to nonnormality (see Chao and Swanson 2002).

2.2.3. Weak-Instrument Asymptotics. Like many-instrument asymptotics, weak-instrument asymptotics (Staiger and Stock 1997) involves a sequence of models chosen to keep μ^2/K constant as $T \to \infty$. However, unlike many-instrument asymptotics, K is held fixed. Technically, the sequence of models considered is the same as used to derive the local asymptotic power of the first-stage F test (a "Pitman drift" parameterization in which Π is in a $1/\sqrt{T}$ neighborhood of 0). Staiger and Stock (1997) showed that under general conditions on the errors and with random instruments, many results that hold exactly in the fixed-instrument, normal-error model can be reinterpreted as holding asymptotically, with simplifications arising from the consistency of $Z'Z/T$ and of the estimator for σ_v^2.

3. EMPIRICAL EXAMPLES

3.1 Estimating the Returns to Education

In an influential article, Angrist and Krueger (1991) proposed using the quarter of birth as an instrument to circumvent ability bias in estimating the returns to education. The date of birth, they argued, should be uncorrelated with ability, so that quarter of birth is exogenous; because of mandatory schooling laws, quarter of birth should also be relevant. With large samples from the U.S. census, they estimated the returns to education by TSLS, using as instruments quarter of birth and its interactions with state and year of birth binary variables, for as many as 178 instruments.

Surprisingly, despite the large number of observations (329,000 or more), the instruments, taken together, are weak in some of the Angrist–Krueger regressions. This point was first made by Bound et al. (1995), who provided Monte Carlo results showing that in some specifications, similar point estimates and standard errors obtain if each individual's true quarter of birth is replaced by a randomly generated quarter of birth. Because the results with the randomly generated quarter of birth must be spurious, this suggests that the results with the true quarter of birth are misleading. The source of these misleading inferences is weak instruments; in some specifications, the first-stage F statistic is less than 2, suggesting that μ^2/K might be 1 or less (recall that $E(F) - 1 \cong \mu^2/K$). In these specifications, there are a few strong instruments (the quarter of birth binary variables) and many weak ones (their interactions with state and year), resulting in a combined set of instruments that is weak. An important conclusion is that it is not helpful to think of weak instruments as a "finite-sample" problem that can be ignored if one has many observations.

3.2 The Log-Linearized Euler Equation in the Consumption-Based Capital Asset-Pricing Model

The first empirical application of GMM was Hansen and Singleton's (1982) investigation of the consumption-based capital asset pricing model (CCAPM). In its log-linearized form, the first-order condition of the CCAPM with constant relative risk aversion can be written as

$$E[(r_{t+1} + \alpha - \gamma \Delta c_{t+1})|Z_t] = 0, \tag{5}$$

where γ is the coefficient of relative risk aversion (here also the inverse of the elasticity of intertemporal substitution), Δc_{t+1} is the growth rate of consumption, r_{t+1} is the log gross return on some asset, α is a constant, and Z_t is a vector of variables in the information set at time t (Hansen and Singleton 1983; Campbell 2001 for a survey).

The coefficients of (5) can be estimated by GMM using Z_t as an instrument. One way to proceed is to use TSLS with r_{t+1} as the dependent variable; another is to apply TSLS with Δc_{t+1} as the dependent variable; and a third is to use a method, such as limited-information maximum likelihood (LIML), that is invariant to the normalization. Under standard fixed-model asymptotics, these estimators are asymptotically equivalent, so it should not matter which method is used. However, as discussed in detail by Neely, Roy, and Whiteman (2001) and Yogo (2002), this does matter greatly in practice, with point estimates of γ ranging from small (Hansen and Singleton 1982, 1983) to very large (Hall 1988; Campbell and Mankiw 1989).

The first-stage F statistics in these regressions are frequently less than 5 (Yogo 2002), and it appears that weak instruments can explain many of these seemingly contradictory results (Stock and Wright 2000; Neely et al. 2001). For an instrument to be strong, it must be a good predictor of either consumption growth or an asset return, depending on the normalization, but both are notoriously difficult to predict. So finding weak instruments in this application should not be a surprise.

3.3 Macroeconometric Examples

Weak identification can also be a concern in GMM estimation of macroeconomic equations with expectational terms. For example, Ma (2002) and Mavroeidis (2001) suggested that weak instruments can be an issue in GMM estimation of the hybrid New Keynesian Phillips curve (Fuhrer and Moore 1995; Gali and Gertler 1999). Other macroeconomic applications that confront weak identification include estimates of New Keynesian output equations (Fuhrer and Rudebusch 2002) and some structural vector autoregressions (Pagan and Robertson 1998).

4. DETECTION OF WEAK INSTRUMENTS

This section discusses methods for detecting weak instruments. In general, the linear IV regression model has n endogenous regressors, so that Y and v in (2) are $T \times n$. The methods for detecting weak instruments (and the definition of the concentration parameter) depend on n. We first discuss inference based on the first-stage F statistic when there is a single endogenous regressor, then turn to the case of $n > 1$. The section concludes with an alternative approach to inference about weak instruments proposed by Hahn and Hausman (2002).

To keep things simple, the formulas in Sections 4–6 apply to the case in which there are no included exogenous regressors. These formulas and methods, however, generally extend to the case of included exogenous regressors by replacing y, Y, and Z by the residuals from their projection onto the included exogenous regressors and by modifying the degrees of freedom as needed. Unless noted otherwise, the methods discussed in Sections 4–6 do not require fixed instruments and normally distributed errors for their asymptotic justification.

4.1 The First-Stage *F* Statistic

Before discussing how to use the first-stage F statistic to detect weak instruments, we need to provide a precise definition of weak instruments.

4.1.1. A Definition of Weak Instruments. A practical approach is to define a set of instruments to be weak if μ^2/K is small enough that inferences based on conventional normal approximating distributions are misleading. In this approach, the definition of weak instruments depends on the purpose to which the instruments are put, combined with the researcher's tolerance for departures from the usual standards of inference (i.e., bias, size of tests). For example, suppose that one is using TSLS and want its bias to be small. Accordingly, one measure of whether a set of instruments is strong is whether μ^2/K is sufficiently large so that the TSLS relative bias (as defined in Sec. 2) is at most (say) 10%; if not, then the instruments are deemed weak. Alternatively, if interested in hypothesis testing, one could define instruments to be strong if μ^2/K is large enough that a 5% hypothesis test rejects no more than (say) 15% of the time; otherwise, the instruments are weak. These two definitions—one based on relative bias and the other based on size—in general yield different threshold values of μ^2/K; thus instruments might be weak if used for one application, but not if used for another.

Here we consider the two definitions of weak instruments in the previous paragraph: The TSLS relative bias could exceed 10%, or the actual size of the nominal 5% TSLS t test could exceed 15%. As shown by Stock and Yogo (2001), under weak-instrument asymptotics, each of these definitions implies a threshold value of μ^2/K. If the actual value of μ^2/K exceeds this threshold, then the instruments are strong (e.g., TSLS relative bias is $<$10%). Otherwise, the instruments are weak.

4.1.2. Ascertaining Whether Instruments Are Weak Using the First-Stage F Statistic. In the fixed-instrument, normal-error model, or, alternatively, under weak-instrument asymptotics, the distribution of the first-stage F statistic depends only on μ^2/K and K. Hence the F statistic is useful for making inference about μ^2/K. As Hall, Rudebusch, and Wilcox (1996) showed in Monte Carlo simulations, simply using F to test the hypothesis of nonidentification ($\Pi = 0$) is an inadequate screen for problems caused by weak instruments. Instead, we follow Stock and Yogo (2001) and use F to test the null hypothesis that μ^2/K is less than or equal to the weak-instrument threshold against the alternative that it exceeds the threshold.

For selected values of K, Table 1 reports weak-instrument threshold values of μ^2/K and critical values of F for testing the null hypothesis that instruments are weak. For example, under the TSLS relative bias definition of weak instruments, if $K = 5$, then the threshold value of μ^2/K is 5.82, and the test that $\mu^2/K \le 5.82$ rejects in favor of the alternative that $\mu^2/K > 5.82$ if $F \ge 10.83$. Evidently the first-stage F statistic must be large, typically exceeding 10, for TSLS inference to be reliable.

Table 1. Selected Critical Values for Weak Instrument Tests for TSLS Based on the First-stage F statistic

Number of instruments (K)	Relative bias > 10%: Threshold μ^2/K	Relative bias > 10%: F statistic 5% critical value	Actual size of 5% test > 15%: Threshold μ^2/K	Actual size of 5% test > 15%: F statistic 5% critical value
1			1.82	8.96
2			4.62	11.59
3	3.71	9.08	6.36	12.83
5	5.82	10.83	9.20	15.09
10	7.41	11.49	15.55	20.88
15	7.94	11.51	21.69	26.80

NOTE: The second column contains the smallest values of μ^2/K that ensure that the bias of TSLS is no more than 10% of the inconsistency of OLS. The third column contains the 5% critical values applicable when the first-stage F statistic is used to test the null that μ^2/K is less than or equal to the value in the second column against the alternative that μ^2/K exceeds that value. The final two columns present the analogous weak-instrument thresholds and critical values when weak instruments are defined so that the usual nominal 5% TSLS t test of the hypothesis $\beta = \beta_0$ has size potentially exceeding 15%. (Source: Stock and Yogo 2001.)

4.2 Extension of the First-Stage F Statistic to $n > 1$

When there are multiple endogenous regressors, the concentration parameter is a $K \times K$ matrix, $\Sigma_{VV}^{1/2'}\Pi'Z'Z\Pi\Sigma_{VV}^{1/2}$, where Σ_{VV} is the covariance matrix of the vector of errors v_t. To avoid introducing new notation, we refer to the concentration parameter as μ^2 in both the scalar and matrix cases. The quality of the usual normal approximation is governed by the matrix μ^2/K. Because the predicted values of Y from the first-stage regression can be highly correlated, for the usual normal approximations to be good, it is not sufficient that some elements of μ^2/K are large. Rather, the matrix μ^2/K must be large in the sense that its smallest eigenvalue is large.

From a statistical perspective, when $n > 1$, the n first-stage F statistics are not sufficient for the concentration matrix even with fixed regressors and normal errors (see Shea 1997 for a discussion). Instead, inference about μ^2 can be based on the $n \times n$ matrix analog of the first-stage F statistic,

$$G_T = \widehat{\Sigma}_{VV}^{-1/2'}Y'P_ZY\widehat{\Sigma}_{VV}^{-1/2}/K, \tag{6}$$

where $\widehat{\Sigma}_{VV} = Y'M_ZY/(T-K)$, $M_Z = I - P_Z$, and I is a conformable identity matrix. Under weak-instrument asymptotics, $E(G_T) \to \mu^2/K + I$. Cragg and Donald (1993) proposed using G_T to test for partial identification (cf. Choi and Phillips 1992)—specifically, testing the hypothesis that the matrix Π has rank L against the alternative that it has rank greater than L, where $L < n$. From the perspective of IV inference, mere instrument relevance is insufficient; instead, the instruments must be strong in the sense that μ^2/K is large. Accordingly, Stock and Yogo (2001) considered the problem of testing the null hypothesis that a set of instruments is weak against the alternative that they are strong, where instruments are defined to be strong if conventional TSLS inference is reliable for any linear combination of the coefficients. By focusing on the worst-behaved linear combination, this approach is conservative but tractable, and Stock and Yogo provided tables of critical values, analogous to those in Table 1, based on the minimum eigenvalue of G_T.

4.3 A Test of the Null of Strong Instruments

The methods discussed so far have been tests of the null of weak instruments. Hahn and Hausman (2002) reversed the null and alternative and proposed a test of the null that the instruments are strong against the alternative that they are weak. They noted that when there is a single endogenous regressor ($n = 1$) and the instruments are strong, normalization of the regression (the choice of dependent variable) should not matter. Thus the TSLS estimator in the forward regression of y on Y and the inverse of the TSLS estimator in the reverse regression of Y on y are asymptotically equivalent [to order $o_p(T^{-1/2})$] with strong instruments, but this is not the case if the instruments are weak. Accordingly, Hahn and Hausman (2002) developed a statistic comparing the forward and reverse regression estimators (and their extensions when $n = 2$). They suggested that if this statistic rejects the null hypothesis, then a researcher should conclude that his or her instruments are weak. Otherwise, the researcher can treat the instruments as strong.

5. FULLY ROBUST INFERENCE WITH WEAK INSTRUMENTS

This section discusses hypothesis tests and confidence sets for β that are fully robust to weak instruments, in the sense that these procedures have the correct size or coverage rates regardless of the value of μ^2 (including $\mu^2 = 0$) when the sample size is large (specifically, under weak-instrument asymptotics). We focus on the case where $n = 1$, but these methods generalize to joint inference about β when $n > 1$.

Several fully robust tests have been proposed; consistent with earlier Monte Carlo studies, the results here suggest that none appears to dominate the others. Moreira (2001) provided a theoretical explanation of this in the context of the fixed-instrument, normal-error model. In that model, there is no uniformly most powerful test of the hypothesis $\beta = \beta_0$, a result that also holds more generally under weak-instrument asymptotics. In this light, the various fully robust procedures represent trade-offs, with some working better than others, depending on the true parameter values.

5.1 A Family of Fully Robust Gaussian Tests

Moreira (2001) considered the system (1) and (2) with fixed instruments and normally distributed errors. Suppose that the reduced-form equation for y is $y = Z\Pi\beta + w$. Let Ω denote the covariance matrix of the reduced-form errors, $[w_t\ v_t]'$, and for now suppose that Ω is known. We are interested in testing the hypothesis $\beta = \beta_0$.

Moreira (2001) showed that under these assumptions, the statistics

$$\mathcal{S} = \frac{(Z'Z)^{-1/2} Z' \underline{Y} b_0}{\sqrt{b_0' \Omega b_0}} \quad \text{and} \quad \mathcal{T} = \frac{(Z'Z)^{-1/2} Z' \underline{Y} \Omega^{-1} a_0}{\sqrt{a_0' \Omega^{-1} a_0}} \tag{7}$$

are sufficient for β and Π, where $\underline{Y} = [y\ Y]$, $b_0 = [1\ \ -\beta_0]'$, and $a_0 = [\beta_0\ \ 1]'$. Thus for the purpose of testing $\beta = \beta_0$, it suffices to consider test statistics that are functions of only $\mathcal{S}$ and $\mathcal{T}$, say $g(\mathcal{S}, \mathcal{T})$. Moreover, under the null hypothesis $\beta = \beta_0$, the distribution of $\mathcal{T}$ depends on Π, but the distribution of $\mathcal{S}$ does not; thus, under the null hypothesis, $\mathcal{T}$ is sufficient for Π. It follows that a test of $\beta = \beta_0$ based on $g(\mathcal{S}, \mathcal{T})$ is similar if its critical value is computed from the conditional distribution of $g(\mathcal{S}, \mathcal{T})$ given $\mathcal{T}$. Moreira (2001) also derived an infeasible power envelope for similar tests under the further assumption that Π is known. In practice, Π is not known; when $K > 1$, feasible tests cannot achieve the power envelope, and there is no uniformly most powerful test of $\beta = \beta_0$.

In practice, Ω is unknown, so the statistics in (7) cannot be computed. However, under weak-instrument asymptotics, Ω can be estimated consistently under the null and, moreover, the results in the preceding paragraph generalize to stochastic instruments and nonnormal errors. Accordingly, let $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{T}}$ denote $\mathcal{S}$ and $\mathcal{T}$ evaluated with $\widehat{\Omega} = \underline{Y}' M_Z \underline{Y}/(T-K)$ replacing Ω, where $M_Z = I - P_Z$. We refer to Moreira's (2001) family of tests, based on statistics of the form $g(\widehat{\mathcal{S}}, \widehat{\mathcal{T}})$, as Gaussian similar tests.

5.2 Three Gaussian Similar Tests

We now turn to three Gaussian similar tests: the Anderson–Rubin (AR) statistic, Kleibergen statistic, and Moreira statistic.

5.2.1. The Anderson–Rubin Statistic. More than 50 years ago, Anderson and Rubin (1949) proposed testing the null hypothesis $\beta = \beta_0$ using the statistic

$$AR(\beta_0) = \frac{(y - Y\beta_0)' P_Z (y - Y\beta_0)/K}{(y - Y\beta_0)' M_Z (y - Y\beta_0)/(T-K)} = \frac{\widehat{\mathcal{S}}'\widehat{\mathcal{S}}}{K}. \tag{8}$$

One definition of the LIML estimator is that it minimizes $AR(\beta)$.

With fixed instruments and normal errors, the quadratic forms in the numerator and denominator of (8) are independent chi-squared random variables under the null hypothesis, and $AR(\beta_0)$ has an exact $F_{K,T-K}$ null distribution. Under the more general conditions of weak-instrument asymptotics, $AR(\beta_0) \xrightarrow{d} \chi_k^2/K$ under the null hypothesis, regardless of the value of μ^2/K. Thus the AR statistic provides a fully robust test of the hypothesis $\beta = \beta_0$.

The AR statistic is profligate in its use of overidentifying restrictions in the sense that the numerator projects $y - Y\beta_0$ on Z rather than on a subspace of Z, leading to a loss of power relative to the infeasible power envelope when β is overidentified. Moreover, the AR statistic can reject either because $\beta \neq \beta_0$ or because the instrument orthogonality conditions fail, so inference based on the AR statistic differs from inference based on conventional GMM test statistics, for which the maintained hypothesis is that the instruments are valid. For these reasons, other statistics have been proposed for testing $\beta = \beta_0$ with the aim of improving power relative to $AR(\beta_0)$ when β is overidentified.

5.2.2. Kleibergen's Statistic. Kleibergen (2001) proposed the statistic

$$K(\beta_0) = \frac{(\widehat{\mathcal{S}}'\widehat{\mathcal{T}})^2}{\widehat{\mathcal{T}}'\widehat{\mathcal{T}}}, \tag{9}$$

which, following Moreira (2001), we have written in terms of $\widehat{\mathcal{S}}$ and $\widehat{\mathcal{T}}$. If $K = 1$, then $K(\beta_0) = AR(\beta_0)$. Kleibergen showed that under either conventional or weak-instrument asymptotics, $K(\beta_0)$ has a χ_1^2 null limiting distribution.

5.2.3. Moreira's Statistic. Moreira (2002) proposed testing $\beta = \beta_0$ using the conditional likelihood ratio test statistic

$$M(\beta_0) = \frac{1}{2}\Bigg(\widehat{\mathcal{S}}'\widehat{\mathcal{S}} - \widehat{\mathcal{T}}'\widehat{\mathcal{T}} + \sqrt{(\widehat{\mathcal{S}}'\widehat{\mathcal{S}} + \widehat{\mathcal{T}}'\widehat{\mathcal{T}})^2 - 4\big[(\widehat{\mathcal{S}}'\widehat{\mathcal{S}})(\widehat{\mathcal{T}}'\widehat{\mathcal{T}}) - (\widehat{\mathcal{S}}'\widehat{\mathcal{T}})^2\big]}\Bigg). \tag{10}$$

The (weak-instrument) asymptotic distribution of $M(\beta_0)$ under the null, conditional on $\widehat{\mathcal{T}} = \tau$, is nonstandard and depends on β_0 and τ. Moreira (2002) suggested computing the null distribution by Monte Carlo simulation.

5.3 Conservative Tests

Staiger and Stock (1997) suggested testing $\beta = \beta_0$ using a Bonferroni test. Wang and Zivot (1998) and Zivot, Startz, and Nelson (1998) proposed a modification of conventional GMM statistics in which σ_u^2 is estimated under the null hypothesis.

Under weak-instrument asymptotics, these tests are conservative (i.e., their size is less than their significance level for some values of the parameters). Numerical analysis suggests that these tests tend to have lower power than the Gaussian similar tests.

5.4 Power Comparisons

The asymptotic power functions of the AR, Kleibergen, and Moreira tests depend on μ^2/K, ρ [the correlation between u and v in (1) and (2)], and K, as well as on the true value of β. We consider two values of $\mu^2/K : \mu^2/K = 1$, which corresponds to very weak instruments (nearly unidentified), and $\mu^2/K = 5$, which corresponds to moderately weak instruments. The two values of ρ considered correspond to moderate endogeneity ($\rho = .5$) and very strong endogeneity ($\rho = .99$, as used in Fig. 1).

Figure 2 presents weak-instrument asymptotic power functions for $K = 5$ instruments, so the degree of overidentification is 4. The power depends on $\beta - \beta_0$ but not on β_0, so Figure 2 applies to general β_0. The shaded region is the area between Moreira's (2001) infeasible asymptotic Gaussian power envelope and the power function of the AR test; the challenge for newly proposed fully robust tests is to have power functions as close to the top of this region as possible. When $\mu^2/K = 1$ and $\rho = .5$, all tests have poor power for all values of the parameter space—a reassuring result given how weak the instruments are; moreover, all tests have power functions that are far from the infeasible power envelope. Notably, the power functions do not increase monotonically in $|\beta - \beta_0|$. When $\mu^2/K = 5$, the M test (but not the K test) approaches the infeasible envelope for both values of ρ.

Figure 3 presents the corresponding power functions for many instruments ($K = 50$). In all cases, the M test is within or toward the top of the shaded region; this is mainly (but not always) the case for the K test, which has a power function that, oddly, descends substantially for $\beta \ll \beta_0$. As Figure 3 makes clear, when K is large, the AR test has relatively low power (arising from its inefficient use of overidentifying restrictions), and substantial power improvements are possible, particularly by using the M test.

5.5 Robust Confidence Sets

Due to the duality between hypothesis tests and confidence sets, these tests can be used to construct fully robust confidence sets. For example, a fully robust 95% confidence set can be constructed as the set of β_0 for which the AR statistic, $AR(\beta_0)$, fails to reject at the 5% significance level. In general, this approach requires evaluating the test statistic for all points in the parameter space, although for some statistics the confidence interval can be obtained by solving a polynomial equation.

When the instruments are weak, these sets can have infinite volume. For example, because the AR statistic is a ratio of quadratics, it can have a finite maximum, and when $\mu^2 = 0$, any point in the parameter space will be contained in the AR confidence set with probability 95%. This does not imply that

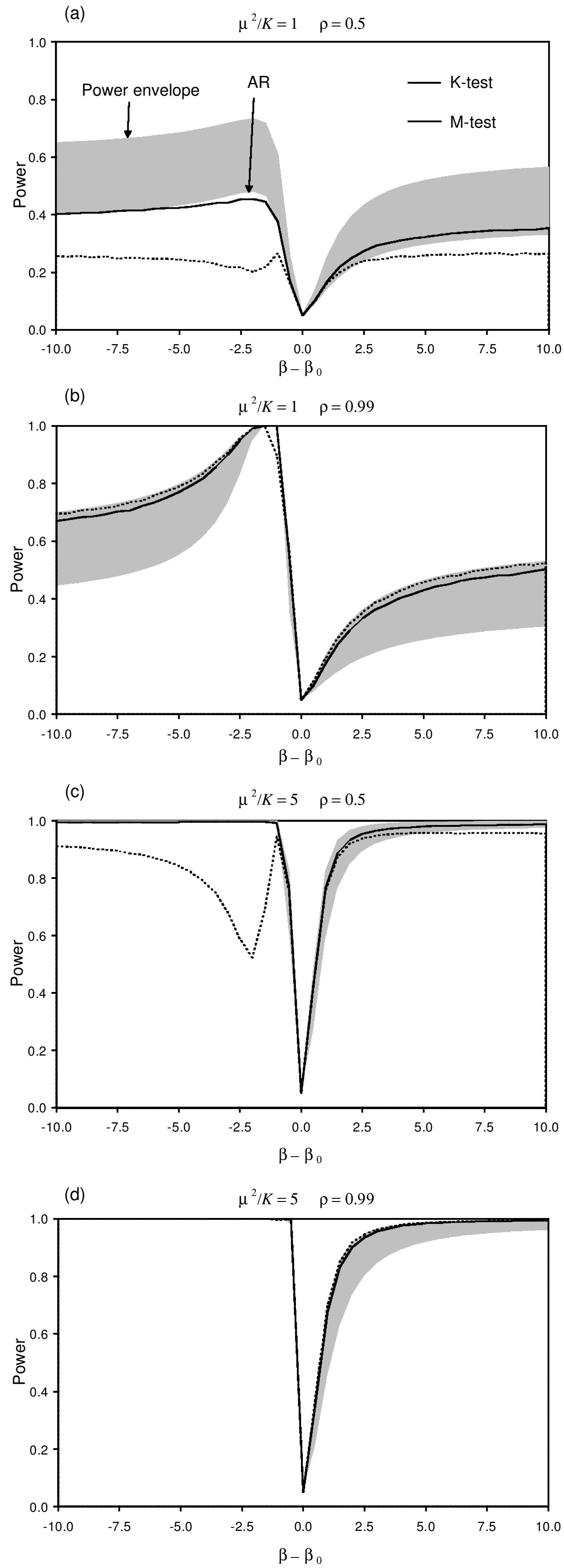

Figure 2. Weak-Instrument Asymptotic Power of Gaussian Similar Tests for K = 5 Instruments. The upper boundary of the shaded area is the Gaussian power envelope, the lower boundary is the power of the AR test. The other two power functions are for Kleibergen's and Moreira's tests.

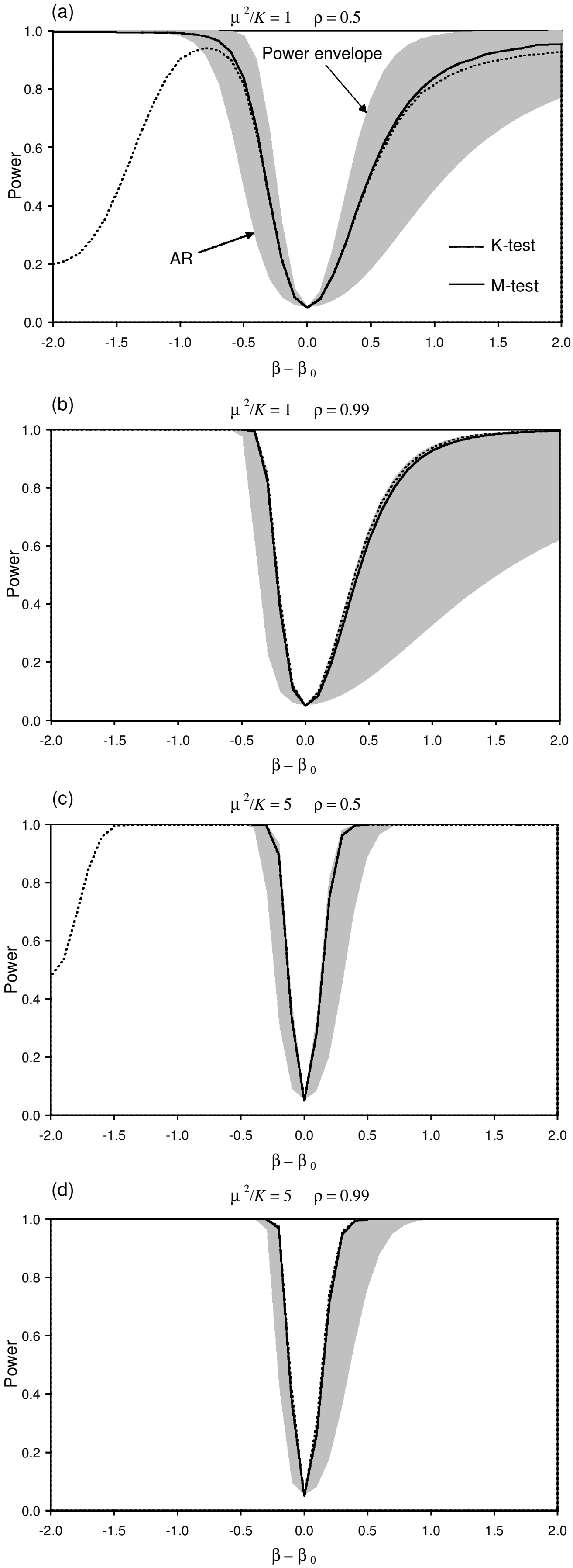

Figure 3. Weak-Instrument Asymptotic Power of Gaussian Similar Tests for $K = 50$ Instruments. See the legend to Figure 2.

these methods waste information or are unnecessarily imprecise; rather, if instruments are weak, then there simply is limited information to use to make inferences about β. This point was made formally by Dufour (1997), who showed that under weak-instrument asymptotics, a confidence set for β must have infinite expected volume if it is to have nonzero coverage uniformly in the parameter space, as long as μ^2 is fixed and finite. This infinite expected volume condition is shared by confidence sets constructed using any of the fully robust methods of this section (see Zivot et al. 1998 for further discussion).

6. PARTIALLY ROBUST INFERENCE WITH WEAK INSTRUMENTS

Although the fully robust tests discussed in the previous section always control size, they can be difficult to compute. Moreover, for $n > 1$, they do not readily provide point estimates, and confidence intervals for individual elements of β must be obtained by conservative projection methods. The methods described in this section are relatively easy to compute, and inference proceeds using conventional normal fixed-model asymptotic approximations. These methods are partially robust to weak instruments in the sense that they are more reliable than TSLS when instruments are weak.

6.1 k-Class Estimators

The k-class estimator of β is $\hat{\beta}(k) = [Y'(I - kM_Z)Y]^{-1}[Y'(I - kM_Z)y]$. This class includes TSLS (for which $k = 1$), LIML, and some alternatives that improve on TSLS when instruments are weak.

6.1.1. Limited-Information Maximum Likelihood. LIML is a k-class estimator where $k = k_{\text{LIML}}$ is the smallest root of the determinantal equation $|\underline{Y}'\underline{Y} - k\underline{Y}'M_Z\underline{Y}| = 0$. Although the mean of the LIML estimator does not exist because its distribution has fat tails, its median is typically much closer to β than is the mean or median of TSLS. In the fixed-instrument, normal-error model, the bias of TSLS increases with K, but the bias of LIML does not (Rothenberg 1984). When the instruments are fixed and the errors are symmetrically distributed, LIML is the best median-unbiased k-class estimator to second order (Rothenberg 1983). Moreover, unlike TSLS, LIML is consistent under many-instrument asymptotics (Bekker 1994).

6.1.2. Fuller-k Estimator. Fuller (1977) proposed an alternative k-class estimator that sets $k = k_{\text{LIML}} - b/(T - K)$, where b is a positive constant. With fixed instruments and normal errors, the Fuller-k estimator with $b = 1$ is best unbiased to second order (Rothenberg 1984). In Monte Carlo simulations, Hahn et al. (2001a) reported substantial reductions in bias and mean squared error (MSE) using Fuller-k estimators, relative to TSLS and LIML, when instruments are weak.

6.1.3. Bias-Adjusted Two-Stage Least Squares. Donald and Newey (2001) considered a bias-adjusted TSLS estimator (BTSLS), a k-class estimator with $k = T/(T - K + 2)$, modifying an estimator previously proposed by Nagar (1959). Rothenberg (1984) showed that BTSLS is unbiased to second order in the fixed-instrument, normal-error model. Donald and Newey provided expressions for the second-order asymptotic

MSE of BTSLS, TSLS, and LIML as a function of the number of instruments K. In Monte Carlo simulations, these authors found that selecting the number of instruments to minimize the second-order MSE generally improves performance. Chao and Swanson (2001) derived the bias and MSE of TSLS under weak-instrument asymptotics, modified to allow the number of instruments to increase with the sample size. They reported improvements in Monte Carlo simulations by incorporating bias adjustments.

6.1.4. *Jackknife Instrumental Variables.* Angrist, Imbens, and Krueger (1999) proposed the jackknife instrumental variables estimator (JIVE), $\hat{\beta}^{\text{JIVE}} = (\widetilde{Y}'Y)^{-1}\widetilde{Y}'y$, where the ith row of $\widetilde{Y}$ is $Z_i\widehat{\Pi}_{-i}$ and $\widehat{\Pi}_{-i}$ is the estimator of Π computed using all but the ith observation. They showed that JIVE and TSLS are asymptotically equivalent under conventional fixed-model asymptotics. Calculations drawing on work of Chao and Swanson (2002) reveal that under weak-instrument asymptotics, JIVE is asymptotically equivalent to a k-class estimator with $k = 1 + K/(T-K)$. Theoretical calculations (Chao and Swanson 2002) and Monte Carlo simulations (Angrist, Imbens, and Krueger 1999; Blomquist and Dahlberg 1999) indicate that JIVE improves on TSLS when there are many instruments.

6.2 Comparisons

One way to assess how robust an estimator or test is to weak instruments is to characterize the size of its weak instrument region. When $n = 1$, this can be done by computing the critical value of the first-stage F statistic testing (at the 5% level) the null hypothesis that μ^2/K is too small to ensure a desired degree of reliability under weak-instrument asymptotics (i.e., the instruments are weak) against the alternative that it exceeds the threshold value of μ^2/K (i.e., the instruments are strong). This is the approach taken in Table 1 for TSLS, and Figure 4 applies it to the other estimators discussed in this section. In Figure 4(a), the weak-instrument set is defined to be the set of μ^2/K such that the relative bias of the estimator exceeds 10%; in Figure 4(b), the weak-instrument set is instead defined so that a nominal 5% test of $\beta = \beta_0$, based on the relevant t statistic, can have size exceeding 15%. In the context of Figure 4, the smaller the critical values, the more robust the procedure.

As Figure 4 shows, LIML, BTSLS, JIVE, and the Fuller-k estimator (with $b = 1$) generally have smaller critical values than TSLS. In this sense, these four estimators are more robust to weak instruments than TSLS. In contrast to TSLS, these critical values decrease as a function of K. For $K \geq 10$, the critical values of the first-stage F statistic fall to 5 or less, well below those for TSLS. In this sense, these partially robust methods evidently provide relatively reliable alternatives in applications with weak instruments.

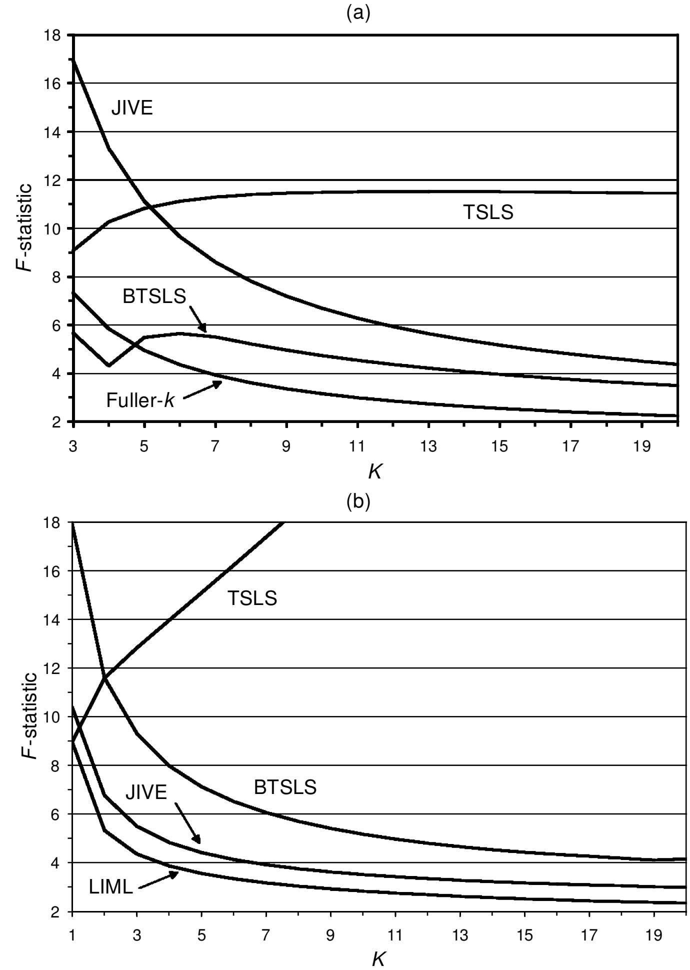

Figure 4. Critical Values for Weak-Instrument Tests Based on the First-Stage F Statistic for the TSLS, LIML, BTSLS, JIVE, and Fuller-k Estimators As a Function of the Number of Instruments (K). The critical value is for a 5% test of the null hypothesis that the instruments are weak, de' ned as (a) the weak-instrument asymptotic relative bias of the estimator exceeds 10% and (b) the weak-instrument asymptotic size of the 5% Wald test can exceed 15%.

7. GENERALIZED METHOD OF MOMENTS INFERENCE IN GENERAL NONLINEAR MODELS

It has been recognized for some time that the usual large-sample normal approximations to GMM statistics in general nonlinear models can provide poor approximations to exact sampling distributions in problems of applied interest. For example, Hansen et al. (1996) examined GMM estimators of various intertemporal asset pricing models using a Monte Carlo design calibrated to match U.S. data. They found that in many cases, inferences based on the usual normal approximations are misleading (see also Tauchen 1986; Kocherlakota 1990; Ferson and Foerester 1994; Smith 1999). As discussed in Section 3.2, weak instruments are a plausible source of these problems.

The methods of Sections 4–6 apply to the linear IV model with homoscedastic, serially uncorrelated errors. This section provides a nontechnical discussion of methods that apply when the errors are heteroscedastic or serially correlated and/or when the model is nonlinear, that is, extensions of the linear methods for iid data to general GMM. We begin by briefly discussing the problems posed by weak instruments in nonlinear GMM and suggest that a better term in this context is weak identification. We then briefly survey the quite incomplete literature on detection of weak identification and on procedures that are fully or partially robust to weak identification.

7.1 Weak Identification in Nonlinear GMM

In GMM, the $n \times 1$ parameter vector θ is identified by the G conditional mean restrictions $E[h(Y_t, \theta_0)|Z_t] = 0$, where θ_0 is the true value of θ and Z_t is a K-vector of instruments; this in turn implies $E[\phi_t(\theta_0)] = 0$, where $\phi_t(\theta) = h(Y_t, \theta) \otimes Z_t$. If the instruments are relevant, then $E[h(Y_t, \theta) \otimes Z_t] \neq 0$ for $\theta \neq \theta_0$, a necessary condition for θ to be identified. In the linear model, weak instruments arise when $E[h(Y_t, \theta) \otimes Z_t]$ is nearly 0 for $\theta \neq \theta_0$; that is, when Z_t is nearly uncorrelated with the model error term even at false values of θ. More generally, in nonlinear GMM, if $E[h(Y_t, \theta) \otimes Z_t]$ is nearly 0 for $\theta \neq \theta_0$, then θ can be thought of as being weakly identified.

Because there is no exact sampling theory for GMM estimators, formal treatments of weak identification in GMM rely on asymptotics. One approach is to use stochastic expansions in orders of $T^{1/2}$; however, as in the linear case, the resulting approximations seem likely to be poor when identification is very weak. A second approach (Stock and Wright 2000) is to use an asymptotic nesting in which, loosely speaking, the GMM version of the concentration parameter is fixed as $T \to \infty$. This yields a stochastic process representation of the limiting objective function, which in the linear case simplifies to the weak-instrument asymptotics discussed in Section 2.2.

7.2 Detecting Weak Identification

An implication of weak identification is that GMM estimators can exhibit a variety of pathologies. For example, two-step GMM estimators and iterated GMM point estimators can be quite different and can produce quite different confidence sets. If identification is weak, then GMM estimates can be sensitive to the addition of instruments or to changes in the sample. If these features are present in an empirical application, then they can be symptomatic of weak identification.

The only formal test for weak identification in nonlinear GMM of which we are aware is that proposed by Wright (2001). In the conventional asymptotic theory of GMM, the identification condition requires the gradient of $\phi_t(\theta_0)$ to have full column rank. Wright (2001) proposed a test of the hypothesis of a complete failure of this rank condition. Thus Wright's test, like Cragg and Donald's (1993) in the linear model, is strictly a test for nonidentification or underidentification, not for weak identification.

7.3 Procedures That Are Fully Robust to Weak Identification

We are aware of only two fully robust methods for testing $\theta = \theta_0$ in nonlinear GMM: a nonlinear AR statistic and Kleibergen's statistic.

7.3.1. Nonlinear Anderson–Rubin Statistic. Because the numerator and denominator of the AR statistic (8) are evaluated at the true parameter value, it has a weak-instrument asymptotic $F_{K,\infty}$ distribution even if the unknown parameters are poorly identified. This observation suggests tests of $\theta = \theta_0$ based on the nonlinear analog of the AR statistic, which is the so-called continuous-updating GMM objective function (Hansen et al. 1996) in which the weight matrix is evaluated at the same parameter value as the numerator:

$$S_T^{CU}(\theta) = \left[\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_t(\theta)\right]' \widehat{W}(\theta)^{-1}\left[\sqrt{\frac{1}{T}}\sum_{t=1}^{T}\phi_t(\theta)\right], \quad (11)$$

where $\widehat{W}(\theta) = T^{-1}\sum_{t=1}^{T}[\phi_t(\theta) - \bar{\phi}(\theta)][\phi_t(\theta) - \bar{\phi}(\theta)]'$ and $\bar{\phi}(\theta) = T^{-1}\sum_{t=1}^{T}\phi_t(\theta)$. [If $\phi_t(\theta)$ is serially correlated, then $\widehat{W}(\theta)$ is replaced by an estimator of the spectral density of $\phi_t(\theta)$ at frequency 0.]

Under the null hypothesis $\theta = \theta_0$, $S_T^{CU}(\theta_0)$ is asymptotically χ^2_{GK} distributed, whether identification is weak or strong (Stock and Wright 2000). If the instruments are relevant, then under the alternative that $\theta \neq \theta_0$, the "numerator moments" of $S_T^{CU}(\theta_0)$ have nonzero expectation. A confidence set for θ is computed by inverting the $S_T^{CU}(\theta)$ statistic numerically (see Stock and Wright 2000; Ma 2002 for examples).

7.3.2. Kleibergen's Statistic. Kleibergen (2002) proposed testing the hypothesis $\theta = \theta_0$ using a generalization of $K(\beta_0)$ and showed that the proposed statistic has a χ^2_n distribution under both conventional asymptotics and the weak-identification asymptotics of Stock and Wright (2000). Kleibergen found in Monte Carlo simulations that his proposed statistic generally gives a more powerful test than $S_T^{CU}(\theta_0)$, consistent with the improvement of the K test over the AR test reported in Section 5.4.

7.4 Procedures That Are Partially Robust to Weak Identification

Because there are estimators that improve on TSLS when instruments are weak in the linear case, it stands to reason that there should be estimators that improve on two-step GMM in the nonlinear case. The limited work in this area to date has yielded some promising results. Two GMM estimators that appear to be partially robust to weak instruments are the continuous-updating estimator (CUE) (Hansen et al. 1996) and generalized empirical likelihood (GEL) estimator (Smith 1997). The CUE minimizes $S_T^{CU}(\theta)$ in (11). In the linear model, the CUE is asymptotically equivalent to LIML under weak-instrument and conventional asymptotics if the errors are homoscedastic. GEL estimators represent a family of estimators that contain empirical likelihood (Owen 1988; DiCiccio, Hall, and Romano 1991), the CUE, and other estimators. The GEL estimators have good properties in stochastic expansions (Rothenberg 1999; Newey and Smith 2001). For example, all GEL estimators are like LIML, BTSLS, JIVE, and the Fuller-k estimator in the linear model, in the sense that their second-order bias is less than that of the two-step GMM estimator. Work on GEL estimators in the context of weak instruments is promising but young; the reader is referred to Imbens (2002) for further discussion.

8. CONCLUSIONS

Many of the extensions of GMM since Hansen's (1982) and Hansen and Singleton's (1982) seminal work can be seen as attempts to improve the performance of GMM in circumstances of practical interest to empirical economists. One such circumstance is the presence of weak instruments or weak identification.

Despite the evolving nature of the literature, this survey suggests that there are some useful methods that practitioners can adopt to address concerns about weak instruments. In the linear IV model with homoscedastic errors and one endogenous regressor, applied researchers should at least use the tools of Section 4 to assess whether weak instruments potentially are a problem in a given application, for example, by checking the first-stage F statistic. If the first-stage F statistic is small, say <10, and if the errors appear to be homoscedastic and serially uncorrelated, then either a fully robust method (our preference) from Section 5 or a partially robust method from Section 6 can be used. Even if $F > 10$, it is prudent to check the results using LIML, BTSLS, JIVE, or the Fuller-k estimator, especially when the number of instruments is large. In the GMM case (i.e., the moments are nonlinear in the parameters and/or the errors are heteroscedastic or serially correlated), then one or more of the methods of Sections 7.3 and 7.4 can be used.

There are a number of related topics that, because of space limitations, have not been discussed in this survey. Because we have focused on weak instruments, we did not discuss the problem of estimation when some instruments are strong and others are weak. In that circumstance, one way to proceed is to try to cull the weak instruments from the strong and to use only the strong (see Hall and Inoue 2001; Hall and Peixe 2001; Donald and Newey 2001). A second omitted topic is estimation of linear panel data models with a lagged dependent variable, in which instruments (lags) are weak if the lag coefficient is almost 1; recent work in this area includes that of Kiviet (1995), Alonso-Borrego and Arellano (1996), and Hahn et al. (2001b). A third omitted issue is combining weak instruments with a failure of exogeneity restrictions (as emphasized by Bound et al. 1995). On these and related topics, much work remains.

ACKNOWLEDGMENTS

The authors thank Joshua Angrist, Whitney Newey, Adrian Pagan, Marcelo Moreira, and Eric Swanson for comments on an earlier draft. This research was supported in part by National Science Foundation grant SBR-9730489.

[Received June 2002. Revised June 2002.]

REFERENCES

Alonso-Borrego, C., and Arellano, M. (1996), "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data," *Journal of Business & Economic Statistics*, 17, 36–49.

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

Anderson, T. W., and Rubin, H. (1949), "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, 20, 46–63.

Angrist, J. D., Imbens, G. W., and Krueger, A. B. (1999), "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67.

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

Bekker, P. A. (1994), "Alternative Approximations to the Distribution of Instrumental Variables Estimators," *Econometrica*, 62, 657–681.

Bekker, P. A., and van der Ploeg, J. (1999), "Instrumental Variable Estimation Based on Grouped Data," manuscript, University of Groningen, Dept. of Economics.

Blomquist, S., and Dahlberg, M. (1999), "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments With Weak Instruments," *Journal of Applied Econometrics*, 14, 69–88.

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

Buse, A. (1992), "The Bias of Instrumental Variable Estimators," *Econometrica*, 60, 173–180.

Campbell, J. Y. (2001), "Consumption-Based Asset Pricing," in *Handbook of the Economics of Finance*, forthcoming.

Campbell, J. Y., and Mankiw, N. G. (1989), "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," *NBER Macroeconomics Annual*, 4, 185–216.

Chao, J., and Swanson, N. R. (2001), "Bias and MSE Analysis of the IV Estimator Under Weak Identification With Application to Bias Correction," unpublished manuscript, Purdue University.

——— (2002), "Consistent Estimation With a Large Number of Weak Instruments," unpublished manuscript, Purdue University.

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

Cragg, J. G., and Donald, S. G. (1993), "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

DiCiccio, T., Hall, P., and Romano, J. (1991), "Empirical Likelihood is Bartlett-Correctable," *The Annals of Statistics*, 19, 1053–1061.

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

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

Ferson, W. E., and Foerester, S. R. (1994), "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," *Journal of Financial Economics*, 36, 29–55.

Fuhrer, J. C., and Moore, G. R. (1995), "Inflation Persistence," *Quarterly Journal of Economics*, 110, 127–159.

Fuhrer, J. C., and Rudebusch, G. D. (2002), "Estimating the Euler Equation for Output," unpublished manuscript, Federal Reserve Bank of Boston.

Fuller, W. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–953.

Gali, J., and Gertler, M. (1999), "Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, 44, 195–222.

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

Hahn, J., Hausman, J., and Kuersteiner, G. (2001a), "Higher Order MSE of Jackknife 2SLS," unpublished manuscript, Massachusetts Institute of Technology, Dept. of Economics.

——— (2001b), "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models With Fixed Effects," unpublished manuscript, Massachusetts Institute of Technology, Dept. of Economics.

Hall, A. R., and Inoue, A. (2001), "A Canonical Correlations Interpretation of Generalized Method of Moments Estimation With Applications to Moment Selection," unpublished manuscript, North Carolina State University.

Hall, A. R., and Peixe, F. P. M. (2001), "A Consistent Method for the Selection of Relevant Instruments," unpublished manuscript, North Carolina State University.

Hall, A. R., Rudebusch, G. D., and Wilcox, D. W. (1996), "Judging Instrument Relevance in Instrumental Variables Estimation," *International Economic Review*, 37, 283–289.

Hall, R. E. (1988), "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339–357.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996), "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

Hansen, L. P., and Singleton, K. J. (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

——— (1983), "Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–265.

Imbens, G. W. (2002), "Generalized Method of Moments and Empirical Likelihood," *Journal of Business & Economic Statistics*, forthcoming.

Kiviet, J. F. (1995), "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 1–268.

Kleibergen, F. (2001), "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, forthcoming.

——— (2002), "Testing Parameters in GMM Without Assuming That They Are Identified," unpublished manuscript, University of Amsterdam, Dept. of Economics.

Kocherlakota, N. (1990), "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.

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

Ma, A. (2002), "GMM Estimation of the New Phillips Curve," *Economics Letters*, 76, 411–417.

Mavroeidis, S. (2001), "Identification and Misspecification Issues in Forward Looking Monetary Models," unpublished manuscript, Oxford University, Dept. of Economics.

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

——— (2002), "A Conditional Likelihood Ratio Test for Structural Models," unpublished manuscript, University of California Berkeley, Dept. of Economics.

Morimune, K. (1983), "Approximate Distributions of 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.

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

Nelson, C. R., and Startz, R. (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 Variable Estimator and Its t Ratio When the Instrument Is a Poor One," *Journal of Business*, 63, S125–S140.

Newey, W. K., and Smith, R. J. (2001), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," unpublished manuscript, Massachusetts Institute of Technology, Dept. of Economics.

Owen, A. (1988), "Empirical Likelihood Ratios Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

Pagan, A. R., and Robertson, J. C. (1998), "Structural Models of the Liquidity Effect," *The Review of Economics and Statistics*, 80, 202–217.

Phillips, P. C. B. (1984), "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, Vol. 1, eds. Z. Griliches and M. D. Intriligator, Amsterdam: North-Holland.

Richardson, D. H. (1968), "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214–1226.

Rothenberg, T. J. (1983), "Asymptotic Properties of Some Estimators in Structural Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, eds. S. Karlin, T. Amemiya, and L. A. Goodman, New York: Academic Press.

——— (1984), "Approximating the Distribution of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, Vol. 2, eds. Z. Griliches and M. D. Intriligator, Amsterdam: North-Holland.

——— (1999), "Higher Order Properties of Empirical Likelihood for Simultaneous Equations," unpublished manuscript, University of California Berkeley, Dept. of Economics.

Sawa, T. (1969), "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–936.

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

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Smith, D. C. (1999), "Finite Sample Properties of Tests of the Epstein-Zin Asset Pricing Model," *Journal of Econometrics*, 93, 113–148.

Smith, R. (1997), "Alternative Semiparametric Likelihood Approaches to Generalized Method of Moments Estimation," *Economic Journal*, 107, 503–519.

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

Stock, J. H., and Wright, J. H. (2000), "GMM With Weak Identification," *Econometric*a, 68, 1055–1096.

Stock, J. H., and Yogo, M. (2001), "Testing for Weak Instruments in Linear IV Regression," unpublished manuscript, Harvard University.

Tauchen, G. (1986), "Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained From Financial Market Data," *Journal of Business & Economic Statistics*, 4, 397–425.

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

Wright, J. H. (2001), "Detecting Lack of Identification in GMM," *Econometric Theory*, forthcoming.

Yogo, M. (2002), "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," unpublished manuscript, Harvard University, Dept. of Economics.

Zivot, E., Startz, R., and Nelson, C. R. (1998), "Valid Confidence Intervals and Inference in the Presence of Weak Instruments," *International Economic Review*, 39, 1119–1246.