

# GMM, Weak Instruments, and Weak Identification

James H. Stock

Kennedy School of Government, Harvard University  
and the National Bureau of Economic Research

Jonathan Wright

Federal Reserve Board, Washington, D.C.

and

Motohiro Yogo\*

Department of Economics, Harvard University

June 2002

## ***ABSTRACT***

Weak instruments arise when the instruments in linear IV regression are weakly correlated with the included endogenous variables. In nonlinear GMM, weak instruments correspond to weak identification of some or all of the unknown parameters. Weak identification leads to non-normal 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.

\*Prepared for the *Journal of Business and Economic Statistics* symposium on GMM. This research was supported in part by NSF grant SBR-0214131.

## 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. Of course, at a mathematical level, these two requirements are the same, but conceptually the starting point is different. The shift from debatable (in Sims' (1980) words, "incredible") exclusion restrictions to first order conditions derived from economic theory has proven to be a highly productive way to think about candidate instrumental variables in a wide variety of applications. Accordingly, careful consideration of instrument exogeneity now is a standard part of a well-done empirical analysis using GMM.

Instrument exogeneity, however, is only one of the criteria necessary for an instrument to be valid, and 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," 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 paper surveys weak instruments and its counterpart in nonlinear GMM, weak identification. We have five main themes:

1. If instruments are weak, then the sampling distributions of GMM and IV statistics are in general non-normal and standard GMM and IV point estimates, hypothesis tests, and confidence intervals are unreliable.
2. Weak instruments are commonplace in empirical research. This should not be a surprise. Finding exogenous instruments is hard work, and the features that make an instrument plausibly exogenous – for example, 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) provide an empirical example, based on an important article by Angrist and Krueger (1991), in which weak instrument problems arise despite having 329,000 observations. In a formal mathematical sense, the strength of the instruments, as measured by the so-called concentration parameter, plays the role of the sample size in determining the quality of the usual normal approximation.
4. If you have weak instruments, you do not need to abandon your empirical research, but neither should you use conventional GMM or IV methods. Various tools are available for handling weak instruments in the linear IV

regression model, and although research in this area is far from complete, a judicious use of these tools can result in reliable inferences.

5. What to do about weak instruments – even how to recognize them – is a much more difficult problem in general nonlinear GMM than in linear IV regression, and much theoretical work remains. Still, it is possible to make some suggestions for empirical practice.

This survey emphasizes the linear IV regression model, simply because much more is known about weak instruments in this case. We begin in Section 2 with a primer on weak instruments in linear IV regression. With this as background, Section 3 discusses three important empirical applications that confront the challenge of weak instruments. Sections 4 – 6 discuss recent econometric approaches to weak instruments: their detection (Section 4); methods that are fully robust to weak instruments, at least in large samples (Section 5); and methods that are somewhat simpler to use and are partially robust to weak instruments (Section 6). Section 7 turns to weak identification in nonlinear GMM, its consequences, and methods for detecting and handling weak identification. Section 8 concludes.

As we see in the next section, many of the key ideas of weak instruments have been understood for decades and can be explained in the context of the simplest IV regression model. This said, 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 out of date.

## 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 i.i.d., normally distributed errors. This section therefore begins by using this model to show how weak instruments lead to TSLS having a non-normal sampling distribution, regardless of the sample size. The fixed-instrument, normal-error model, however, has strong assumptions that are empirically unrealistic. This section therefore concludes with a synopsis of asymptotic methods that weaken these strong assumptions yet attempt to retain the insights gained from the finite sample distribution theory.

### 2.1 The Linear Gaussian IV Regression Model with a Single Regressor

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

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

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

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  $[u_t \ v_t]'$  are assumed to be i.i.d. and normally distributed  $N(0, \Sigma)$ ,  $t = 1, \dots, T$ , where the elements of  $\Sigma$  are  $\sigma_u^2$ ,  $\sigma_{uv}$ , and  $\sigma_v^2$ .

Equation (1) is the structural equation, and  $\beta$  is the scalar parameter of interest. Equation (2) relates the included endogenous variable to the instruments.

***The concentration parameter.*** The concentration parameter,  $\mu^2$ , is a unitless measure of the strength of the instruments, and is defined as

$$\mu^2 = \Pi'Z'Z\Pi/\sigma_v^2. \quad (3)$$

A useful interpretation of  $\mu^2$  is in terms of  $F$ , the  $F$ -statistic testing the hypothesis that  $\Pi=0$  in (2); because  $F$  is the  $F$ -statistic testing for nonzero coefficients on the instruments in the first stage of TSLS, it is commonly called the “first-stage  $F$ -statistic,” terminology that is adopted here. Let  $\tilde{F}$  be the infeasible counterpart of  $F$ , computed using the true value of  $\sigma_v^2$ ; then  $\tilde{F}$  has a noncentral  $\chi_K^2/K$  distribution with noncentrality parameter  $\mu^2/K$ , and  $E(\tilde{F}) = \mu^2/K + 1$ . If the sample size is large,  $F$  and  $\tilde{F}$  are close, so  $E(F) \cong \mu^2/K + 1$ ; that is, the expected value of the first-stage  $F$ -statistic is approximately  $1 + \mu^2/K$ . Thus, larger values of  $\mu^2/K$  shift out the distribution of the first-stage  $F$ -statistic. Said differently,  $F - 1$  can be thought of as the sample analog of  $\mu^2/K$ .

***An expression for the TSLS estimator.*** The TSLS estimator of  $\beta$  is,

$$\hat{\beta}^{\text{TSLS}} = \frac{Y'P_Z y}{Y'P_Z Y}, \quad (4)$$

where  $P_Z = Z(Z'Z)^{-1}Z'$ . Rothenberg (1984) presents a useful expression for the TSLS estimator, which obtains after a few lines of algebra. First substitute the expression for  $Y$  in (2) into (4) to obtain,

$$\hat{\beta}^{\text{TSLS}} - \beta = \frac{\Pi'Z'u + v'P_Z u}{\Pi'Z'Z\Pi + 2\Pi'Z'v + v'P_Z v}. \quad (5)$$

Now define,

$$z_u = \frac{\Pi'Z'u}{\sigma_u \sqrt{\Pi'Z'Z\Pi}}, \quad z_v = \frac{\Pi'Z'v}{\sigma_v \sqrt{\Pi'Z'Z\Pi}},$$

$$S_{uv} = \frac{v'P_Z u}{\sigma_v \sigma_u}, \quad S_{vv} = \frac{v'P_Z v}{\sigma_v^2}.$$

Under the assumptions of fixed instruments and normal errors, the distributions of  $z_u$ ,  $z_v$ ,  $S_{uv}$ , and  $S_{vv}$  do not depend on the sample size  $T$ :  $z_u$  and  $z_v$  are standard normal random variables with a correlation equal to  $\rho$ , the correlation between  $u$  and  $v$ , and  $S_{uv}$  and  $S_{vv}$  are quadratic forms of normal random variables with respect to the idempotent matrix  $P_Z$ .

Substituting these definitions into (5), multiplying both sides by  $\mu$ , and collecting terms yields,

$$\mu(\hat{\beta}^{\text{TSLS}} - \beta) = (\sigma_u/\sigma_v) \frac{z_u + S_{uv}/\mu}{1 + 2z_v/\mu + S_{vv}/\mu^2}. \quad (6)$$

As this expression makes clear, the terms  $z_v$ ,  $S_{uv}$ , and  $S_{vv}$  result in the TSLS estimator having a non-normal distribution. If, however, the concentration parameter  $\mu^2$  is large, then the terms  $S_{uv}/\mu$ ,  $z_v/\mu$ , and  $S_{vv}/\mu^2$  will all be small in a probabilistic sense, so that the dominant term in (6) will be  $z_u$ , which in turn yields the usual normal approximation to the distribution of  $\hat{\beta}^{\text{TSLS}}$ . Formally,  $\mu^2$  plays the role in (6) that is usually played by the number of observations: as  $\mu^2$  gets large, the distribution of  $\mu(\hat{\beta}^{\text{TSLS}} - \beta)$  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 a good one, it is not enough to have many observations: the concentration parameter must be large.

***Bias of the TSLS estimator in the unidentified case.*** When  $\mu^2 = 0$  (equivalently, when  $\Pi = 0$ ), the instrument is not just weak but irrelevant, and the TSLS estimator is centered around the probability limit of the OLS estimator. To see this, use (5) and  $\Pi = 0$  to obtain,  $\hat{\beta}^{\text{TSLS}} - \beta = v'P_Z u / v'P_Z v$ . It is useful to write  $u = E(u|v) + \eta$  where, because  $u$  and  $v$  are jointly normal,  $E(u|v) = (\sigma_{uv}/\sigma_v^2)v$  and  $\eta$  is normally distributed. Moreover,  $\sigma_{uv} = \sigma_{uY}$  and  $\sigma_v^2 = \sigma_Y^2$  because  $\Pi = 0$ . Thus  $u = (\sigma_{uY}/\sigma_Y^2)v + \eta$ , so

$$\hat{\beta}^{\text{TSLS}} - \beta = \frac{v'P_Z[(\sigma_{uY}/\sigma_Y^2)v + \eta]}{v'P_Z v} = \frac{\sigma_{uY}}{\sigma_Y^2} + \frac{v'P_Z \eta}{v'P_Z v}. \quad (7)$$

Because  $\eta$  and  $v$  are independently distributed and  $Z$  is fixed,  $E[(v'P_Z \eta / v'P_Z v)|v] = 0$ .

Suppose that  $K \geq 3$ , so that the first moment of the final ratio in (7) exists; then by the law of iterated expectations,



$$E(\hat{\beta}^{\text{TSL}} - \beta) = \frac{\sigma_{uY}}{\sigma_Y^2}. \quad (8)$$

The right hand side of (8), is the inconsistency of the OLS estimator when  $Y$  is correlated with the error term; that is,  $\hat{\beta}^{\text{OLS}} \xrightarrow{p} \beta + \sigma_{uY}/\sigma_Y^2$ . Thus, when  $\mu^2 = 0$ , the expectation of the TSL estimator is the probability limit of the OLS estimator.

The result in (8) applies to the limiting case of irrelevant instruments; with weak instruments, the TSL estimator is biased towards the probability limit of the OLS estimator. Specifically, define the “relative bias” of TSL to be the bias of TSL relative to the inconsistency of OLS, that is,  $E(\hat{\beta}^{\text{TSL}} - \beta)/\text{plim}(\hat{\beta}^{\text{OLS}} - \beta)$ . Then the TSL relative bias is approximately inversely proportional to  $1 + \mu^2/K$  (this result holds whether or not the errors are not normally distributed (Buse (1992))). Hence, the relative bias decreases monotonically in  $\mu^2/K$ .

*Numerical examples.* Figures 1a and 1b respectively present the densities of the TSL estimator and its  $t$ -statistic for various values of the concentration parameter, when the true value of  $\beta$  is zero. The other parameter values mirror those in Nelson and Startz (1990a, 1990b), with  $\sigma_u = \sigma_v = 1$ ,  $\rho = .99$ , and  $K = 1$  (one instrument). For small values of  $\mu^2$ , such as the value of .25 considered by Nelson and Startz, the distributions are strikingly non-normal, even bimodal. As  $\mu^2$  increases, the distributions eventually approach the usual normal limit.

Under the assumptions of fixed instruments and normal errors, the distributions in Figure 1 depend on the sample size only through the concentration parameter; for a

given value of the concentration parameter, the distributions do not depend on the sample size. Thus, these figures illustrate the point made more generally by (6) that the quality of the usual normal approximation depends on the size of the concentration parameter.

The Nelson–Startz results build on a large literature on the exact distribution of IV estimators under the assumptions of fixed exogenous variables and i.i.d. normal errors (e.g. the exact distribution of the TSLS estimator was obtained by Richardson (1968) and Sawa (1969) for the case of a single right hand side endogenous regressor). From a practical perspective, this literature has two drawbacks. First, the expressions for distributions in this literature, comprehensively reviewed by Phillips (1984), are among the most offputting in econometrics and pose substantial computational challenges. Second, because the assumptions of fixed instruments and normal errors are inappropriate in applications, it is unclear how to apply these results to the sorts of problems encountered in practice. To obtain more generally useful results, researchers have focused on asymptotic approximations, which we now briefly review.

## **2.2 Asymptotic Approximations**

Asymptotic distributions can provide good approximations to exact sampling distributions. Conventionally, asymptotic limits are taken for a fixed model as the sample size gets large, but this is not the only approach and for some problems this is not the best approach, in the sense that it does not necessarily 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 large-sample normal approximations can be quite poor when the concentration parameter is small, even if the number of observations

is large. For this reason, there have been various alternative asymptotic methods used to analyze IV statistics in the presence of weak instruments; three such methods are Edgeworth expansions, large-model asymptotics, and weak-instrument asymptotics. Because the concentration parameter plays a role in these distributions akin to the sample size, all these methods aim to improve the quality of the approximations when  $\mu^2/K$  is small, but the number of observations is large.

***Edgeworth expansions.*** An Edgeworth expansion is a representation of the distribution of the statistic of interest in powers of  $1/\sqrt{T}$ ; Edgeworth and related expansions of IV estimators are reviewed by Rothenberg (1984). As Rothenberg (1984) points 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 upon the conventional normal approximation when  $\mu$  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) suggests the Edgeworth approximation is “excellent” for  $\mu^2 > 50$  and “adequate” for  $\mu^2$  as small as 10, as long as the number of instruments is small (less than  $\mu$ ).

***Many-instrument asymptotics.*** Although the problem of many instruments and weak instruments might at first seem different, they are in fact related: if there were many strong instruments, then the adjusted  $R^2$  of the first-stage regression would be nearly one, so a small first-stage adjusted  $R^2$  suggests that many of the instruments are weak. Bekker (1994) formalized this notion by developing asymptotic approximations for a sequence of models with fixed instruments and Gaussian errors, in which the number of instruments,  $K$ , is proportional to the sample size and  $\mu^2/K$  converges to a

constant, finite limit; similar approaches were taken by Anderson (1976), Kunitomo (1980), and Morimune (1983). The limiting distributions obtained from this approach generally are normal and simulation evidence suggests that these approximations are good for moderate as well as large values of  $K$ , although they cannot capture the non-normality evident in the Nelson–Startz example in Figure 1. Distributions obtained using this approach generally depend on error distribution (see Bekker and van der Ploeg (1999)), so some procedures justified using many-instrument asymptotics require adjustments if the errors are non-normal. Rate and consistency results are, however, more robust to non-normality (see Chao and Swanson (2002)).

***Weak-instrument asymptotics.*** Like many-instrument asymptotics, weak-instrument asymptotics (Staiger and Stock (1997)) considers a sequence of models chosen to keep  $\mu^2/K$  constant; unlike many-instrument asymptotics,  $K$  is held fixed. Technically, the sequence of models considered is the same as is used to derive the local asymptotic power of the first-stage  $F$  test (a “Pitman drift” parameterization in which  $\Pi$  is in a  $1/\sqrt{T}$  neighborhood of zero). Staiger and Stock (1997) show that, under general conditions on the errors and with random instruments, the representation in (6) can be reinterpreted as holding asymptotically in the sense that the sample size  $T$  tends to infinity while  $\mu^2/K$  is fixed. More generally, many of the results from the fixed-instrument, normal-error model apply to models with random instrument and nonnormal errors, with certain simplifications arising from the consistency of the estimator of  $\sigma_v^2$ .

### **2.3 Weak Instruments with Multiple Regressors**

The linear IV model with multiple regressors is,

$$y = Y\beta + X\gamma + u \quad (9)$$

$$Y = Z\Pi + X\Phi + V, \quad (10)$$

where  $Y$  is now a  $T \times n$  vector of observations on  $n$  endogenous variables,  $X$  is a  $T \times J$  matrix of included exogenous variables, and  $V$  is a  $T \times n$  vector of disturbances with covariance matrix  $\Sigma_{VV}$ ; as before,  $Z$  is a  $T \times K$  matrix of instruments and  $u$  is  $T \times 1$ . The concentration parameter is now a  $K \times K$  matrix. Expressed in terms population moments, the concentration matrix is,  $\Sigma_{VV}^{-1/2} \Pi' \Sigma_{ZX} \Pi \Sigma_{VV}^{1/2}$ , where  $\Sigma_{ZX} = \Sigma_{ZZ} - \Sigma_{ZX} \Sigma_{XX}^{-1} \Sigma_{XZ}$ . To avoid introducing new notation, we refer to the concentration parameter as  $\mu^2$  in both the scalar and matrix case.

Exact distribution theory for IV statistics with multiple regressors is quite complicated, even with fixed exogenous variables and normal errors. Somewhat simpler expressions obtain using weak instrument asymptotics (under which reduced form moments, such as  $\Sigma_{VV}$  and  $\Sigma_{ZX}$  are consistently estimable). The quality of the usual normal approximation is governed by the matrix  $\mu^2/K$ . 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 enough for a few elements of  $\mu^2/K$  to be large. Rather, it is necessary for the matrix  $\mu^2/K$  to be large in the sense that its smallest eigenvalue is large.

The notation for IV estimators with included exogenous regressors  $X$  (equations (9) and (10)) is cumbersome. Sections 4 – 6 therefore focus on the case of no included

exogenous variables. The discussion, however, extends directly to the case of included exogenous regressors, and the formulas generally extend by replacing  $y$ ,  $Y$ , and  $Z$  by the residuals from their projection onto  $X$  and by modifying the degrees of freedom as needed.

### **3. Three Empirical Examples**

There are many examples of weak instruments in empirical work. Here, we mention three from labor economics, finance, and macroeconomics.

#### **3.1 Estimating the Returns to Education**

In an influential paper, Angrist and Krueger (1991) proposed using the quarter of birth as an instrumental variable 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. Using large samples from the U.S. census, they therefore estimated the returns to education by TSLS, using quarter of birth and its interactions with state and year of birth binary variables. Depending on the specification, they had as many as 178 instruments.

Surprisingly, despite the large number of observations (329,000 observations or more), in some of the Angrist–Krueger regressions the instruments are weak. This point was first made by Bound, Jaeger, and Baker (1995), who (among other things) provide Monte Carlo results showing that similar point estimates and standard errors obtain in

some specifications 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  $\mu^2/K$  might be one or less (recall that  $E(F) - 1 \cong \mu^2/K$ ). An important conclusion is that it is not helpful to think of weak instruments as a "finite sample" problem that can be ignored if you have sufficiently many observations.

### 3.2 The Log-Linearized Euler Equation in the CCAPM

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,

$$E[(r_{t+1} + \alpha - \gamma^{-1}\Delta c_{t+1})|Z_t] = 0, \quad (11)$$

where  $\gamma$  is the coefficient of risk aversion (here, also the inverse of the intertemporal elasticity of substitution),  $\Delta c_{t+1}$  is the growth rate of consumption,  $r_t$  is the log gross return on some asset,  $\alpha$  is a constant, and  $Z_t$  is a vector of variables in the information set at time  $t$  (Hansen and Singleton (1983); see Campbell (2001) for a textbook treatment).

The coefficients of (11) can be estimated by GMM using  $Z_t$  as an instrument. One way to proceed is to use TSLS with  $r_{i,t+1}$  as the dependent variable; another is to

apply TSLS with  $\Delta c_{t+1}$  as the dependent variable; a third is to use a method, such as 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. But, as discussed in detail in Neely, Roy, and Whiteman (2001) and Yogo (2002), it matters greatly in practice, with point estimates of  $\gamma$  ranging from small (Hansen and Singleton (1982, 1983) to very large (Hall (1988), Campbell and Mankiw (1989)).

Although one possible explanation for these disparate empirical findings is that the instruments are not exogenous – the restrictions in (11) fail to hold – another possibility is that the instruments are weak. Indeed, the analysis of Stock and Wright (2000), Neely, Roy, and Whiteman (2001), and Yogo (2002) suggests that weak instruments are part of the explanation for these seemingly contradictory results. It should not be a surprise that instruments are weak here: 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. In fact, the first-stage  $F$ -statistics in regressions based on (11) are frequently less than 5 (Yogo (2002)).

### **3.3 A Hybrid Phillips curve**

Forward-looking price setting behavior is a prominent feature of modern macroeconomics. The hybrid Phillips curve (e.g. Fuhrer and Moore (1995)) blends forward-looking and backward-looking behavior, so that prices are based in part on expected future prices and in part on past prices. According to a typical hybrid Phillips curve, inflation ( $\pi_t$ ) follows,



$$\pi_t = \beta_1 x_t + \beta_2 E_t \pi_{t+1} + \beta_3 \pi_{t-1} + \omega_t, \quad (12)$$

where  $x_t$  is a measure of demand pressures, such as the output gap,  $E_t \pi_{t+1}$  is the expected inflation rate in time  $t+1$  based on information available at time  $t$ , and  $\omega_t$  is a disturbance. Because  $E_t \pi_{t+1}$  is unobserved, the coefficients of (12) cannot be estimated by OLS but they can be estimated by GMM based on the moment condition,  $E[(\pi_t - \beta_1 x_t - \beta_2 \pi_{t+1} - \beta_3 \pi_{t-1}) | x_t, \pi_t, \pi_{t-1}, Z_t] = 0$ , where  $Z_t$  is a vector of variables known at date  $t$ . Efforts to estimate (12) using the output gap have met with mixed success (e.g. Roberts (1997), Fuhrer (1997)), although Gali and Gertler (1999) report plausible parameter estimates when marginal cost, as measured by labor's share, is used as  $x_t$ .

Although the papers in this literature do not discuss the possibility of weak instruments, recent work by Ma (2001) and Mavroeidis (2001) suggests that weak instruments could be a problem here. Here, too, this should not be a surprise: to be strong, the instrument  $Z_t$  must have substantial marginal predictive content for  $\pi_{t+1}$ , given  $x_t$ ,  $\pi_t$ , and  $\pi_{t-1}$ . However, the regression relating  $\pi_{t+1}$  to  $x_t$ ,  $\pi_t$ , and  $\pi_{t-1}$  is the "old" backwards-looking Phillips curve which, despite its theoretical limitations, is one of the most reliable tools for forecasting inflation (e.g. Stock and Watson (1999)); for  $Z_t$  to be a strong instrument, it must improve substantially upon a backward-looking Phillips curve.

#### 4. Detection of Weak Instruments

This section discusses two methods for detecting weak instruments, one based on the first-stage  $F$ -statistic, the other based on a statistic recently proposed by Hahn and Hausman (2002).

#### 4.1 The First-stage $F$ -statistic

Before discussing how to use the first-stage  $F$ -statistic to detect weak instruments, we need to say what, precisely, weak instruments are.

*A definition of weak instruments.* A practical approach to defining weak instruments is that instruments are weak if  $\mu^2/K$  is so small 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 (bias, size of tests). For example, suppose you are using TSLS and you want its bias to be small. One measure of the bias of TSLS is its bias relative to the inconsistency of OLS, as defined in Section 2. Accordingly, one measure of whether a set of instruments is strong is whether  $\mu^2/K$  is sufficiently large that the TSLS relative bias is, say, no more than 10%; if not, then the instruments are deemed weak. Alternatively, if you are interested in hypothesis testing, then you could define instruments to be strong if  $\mu^2/K$  is large enough that a 5% hypothesis test to reject no more than (say) 15% of the time; otherwise, the instruments are weak. These two definitions – one based on relative bias, one based on size – in general yield different sets of  $\mu^2/K$ ; 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, that is, that the TOLS bias could exceed 10%, or that the nominal 5% test of  $\beta = \beta_0$  based on the usual TOLS  $t$ -statistic has size that could exceed 15%. As shown by Stock and Yogo (2001), under weak-instrument asymptotics, each of these definitions implies a threshold value of  $\mu^2/K$ : if the actual value of  $\mu^2/K$  exceeds this threshold, then the instruments are strong (e.g. TOLS relative bias is less than 10%), otherwise the instruments are weak.

*Ascertaining whether instruments are weak using the first-stage  $F$ -statistic.* In the fixed-instrument/normal-error model, and also under weak-instrument asymptotics, the distribution of the first-stage  $F$ -statistic depends only on  $\mu^2/K$  and  $K$ , so that inference about  $\mu^2/K$  can be based on  $F$ . As Hall, Rudebusch, and Wilcox (1996) show in Monte Carlo simulations, however, simply using  $F$  to test the hypothesis of non-identification ( $H = 0$ ) is inadequate as a screen for problems caused by weak instruments. Instead, we follow Stock and Yogo (2001) and use  $F$  to test the null hypothesis that  $\mu^2/K$  is less than or equal to the weak-instrument threshold, against the alternative that it exceeds the threshold.

Table 1 summarizes weak-instrument threshold values of  $\mu^2/K$  and critical values for the first-stage  $F$ -statistic testing the null hypothesis that instruments are weak, for selected values of  $K$ . For example, under the TOLS relative bias definition of weak instruments, if  $K = 5$  then the threshold value of  $\mu^2/K$  is 5.82 and the test that  $\mu^2/K \leq 5.82$  rejects in favor of the alternative that  $\mu^2/K > 5.82$  if  $F \geq 10.83$ . It is evident from Table 1 that one needs large values of the first-stage  $F$ -statistic, typically exceeding 10, for TOLS inference to be reliable.

## 4.2 Extension of the First-stage $F$ -statistic to $n > 1$

As discussed in Section 2.3, when there are multiple included endogenous regressors ( $n > 1$ ) the concentration parameter  $\mu^2$  is a matrix. Because of possible multicollinearity among the predicted values of the endogenous regressors, in general the distribution of TSLS statistics depends on all the elements of  $\mu^2$ .

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

$$G_T = \hat{\Sigma}_{VV}^{-1/2} Y' P_Z Y \hat{\Sigma}_{VV}^{-1/2} / K, \quad (13)$$

where  $\hat{\Sigma}_{VV}^{-1/2} = Y' M_Z Y / (T - K)$ . Under weak-instrument asymptotics,  $E(G_T) \rightarrow \mu^2 / K + I_n$ , where  $I_n$  is the  $n \times n$  identity matrix.

Cragg and Donald (1993) proposed using  $G_T$  to test for partial identification, specifically, testing the hypothesis that the matrix  $\Pi$  has rank  $L$  against the alternative that it has rank greater than  $L$ , where  $L < n$ . If the rank of  $\Pi$  is  $L$ , then  $L$  linear combinations of  $\beta$  are identified, and  $n - L$  are not; Choi and Phillips (1992) discuss this situation, which they refer to as partial identification. The Cragg-Donald test statistic is  $\sum_{i=1}^{n-L} \lambda_i$ , where  $\lambda_i$  is the  $i^{\text{th}}$ -smallest eigenvalue of  $G_T$ . Under the null, this statistic has a

$\chi^2_{(n-L)(K-L)}/[(n-L)(K-L)]$  limiting distribution. For  $L = 0$  (so the null is complete non-identification), the statistic reduces to  $\text{trace}(G_T)$ , the sum of the  $n$  individual first-stage  $F$ -statistics. If  $L = n - 1$  (so the null is the smallest possible degree of underidentification), then the Cragg–Donald statistic is the smallest eigenvalue of  $G_T$ .

Although the Cragg–Donald statistic can be used to test for underidentification, from the perspective of IV inference, mere instrument relevance is insufficient, rather, the instruments must be strong in the sense that  $\mu^2/K$  must be large. Rejecting the null hypothesis of partial identification does not ensure reliable IV inference. Accordingly, Stock and Yogo (2001) consider 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 they provide tables of critical values, similar to those in Table 1, based on the minimum eigenvalue of  $G_T$ .

### 4.3 A Specification Test of a Null of Strong Instruments

The methods discussed so far have been tests of the null of weak instruments. Hahn and Hausman (2002) reverse the null and alternative and propose a test of the null that the instruments are strong, against the alternative that they are weak. Their procedure is conceptually straightforward. Suppose that there is a single included endogenous regressor ( $n = 1$ ) and that the instruments are strong, so that conventional normal approximations are valid. Then the normalization of the regression (the choice of dependent variable) should not matter asymptotically. 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 if instruments are weak, they are not. Accordingly, Hahn and Hausman (2002) developed a statistic comparing the estimators from the forward and reverse regressions (and the extension of this idea when  $n = 2$ ). They suggest that if this statistic rejects the null hypothesis, then a researcher should conclude that his or her instruments are weak, otherwise he or she can treat the instruments as strong.

## 5. Fully Robust Inference with Weak Instruments in the Linear Model

This section discusses hypothesis tests and confidence sets for  $\beta$  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  $\mu^2$  (including  $\mu^2 = 0$ ) when the sample size is large, specifically, under weak-instrument asymptotics. If  $n = 1$ , these methods also produce median-unbiased estimators of  $\beta$  as the limiting case of a 0% confidence interval. This discussion starts with testing, then concludes with the complementary problem of confidence sets. We focus on the case  $n = 1$ , but the methods can be generalized to joint inference about  $\beta$  when  $n > 1$ .

Several fully robust methods have been proposed, and Monte Carlo studies suggest that none appears to dominate the others. Moreira (2001) provides 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 asymptotically under weak instrument asymptotics. In this light, the various fully

robust procedures represent tradeoffs, with some working better than others, depending on the true parameter values.

### 5.1 Fully Robust Gaussian Tests

Moreira (2001) considers the system (1) and (2) with fixed instruments and normally distributed errors. The reduced-form equation for  $y$  is,

$$y = Z\Pi\beta + w. \quad (14)$$

Let  $\Omega$  denote the covariance matrix of the reduced-form errors,  $[w_t \ v_t]'$ , and for now suppose that  $\Omega$  is known. We are interested in testing the hypothesis  $\beta = \beta_0$ .

Moreira (2001) shows that, under these assumptions, the statistics  $(\mathcal{S}, \mathcal{T})$  are sufficient for  $\beta$  and  $\Pi$ , where

$$\mathcal{S} = \frac{(Z'Z)^{-1/2} Z'Yb_0}{\sqrt{b_0' \Omega b_0}} \quad \text{and} \quad \mathcal{T} = \frac{(Z'Z)^{-1/2} Z'Y\Omega^{-1}a_0}{\sqrt{a_0' \Omega^{-1}a_0}}, \quad (15)$$

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  $\Pi$  but the distribution of  $\mathcal{S}$  does not; thus, under the null hypothesis,  $\mathcal{T}$  is sufficient for  $\Pi$ . 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 derives an

infeasible power envelope for similar tests in the fixed-instrument, normal-error model, under the further assumption that  $\Pi$  is known. In practice,  $\Pi$  is not known so feasible tests cannot achieve this power envelope and, when  $K > 1$ , there is no uniformly most powerful test of  $\beta = \beta_0$ .

In practice,  $\Omega$  is unknown so the statistics in (15) cannot be computed. However, under weak-instrument asymptotics,  $\Omega$  can be estimated consistently. Accordingly, let  $\hat{\mathcal{S}}$  and  $\hat{\mathcal{T}}$  denote  $\mathcal{S}$  and  $\mathcal{T}$  evaluated with  $\hat{\Omega} = \underline{Y}'M_Z\underline{Y}/(T - K)$  replacing  $\Omega$ . Because Moreira's (2001) family of feasible similar tests, based on statistics of the form  $g(\hat{\mathcal{S}}, \hat{\mathcal{T}})$ , are derived under the assumption of normality, we refer to them as Gaussian similar tests.

## 5.2 Three Gaussian Similar Tests

We now turn to three Gaussian similar tests: the Anderson-Rubin statistic, Kleibergen's (2001) test statistic, and Moreira's (2002) conditional likelihood ratio (LR) statistic.

***The Anderson-Rubin Statistic.*** More than fifty years ago, Anderson and Rubin (1949) proposed testing the null hypothesis  $\beta = \beta_0$  in (9) using the statistic,

$$AR(\beta_0) = \frac{(y - Y\beta_0)' P_Z (y - Y\beta_0) / K}{(y - Y\beta_0)' M_Z (y - Y\beta_0) / (T - K)} = \frac{\hat{\mathcal{S}}'\hat{\mathcal{S}}}{K}. \quad (16)$$

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



Anderson and Rubin (1949) showed that, if the errors are Gaussian and the instruments are fixed, then this test statistic has an exact  $F_{K,T-K}$  distribution under the null, regardless of the value of  $\mu^2/K$ . 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  $\mu^2/K$ . Thus the AR statistic provides a fully robust test of the hypothesis  $\beta = \beta_0$ .

The AR statistic can reject either because  $\beta \neq \beta_0$  or because the instrument orthogonality conditions fail. In this sense, inference based on the AR statistic is different from inference based on conventional GMM standard errors, for which the maintained hypothesis is that the instruments are valid. For this reason, a variety of tests have been proposed that maintain the hypothesis that the instruments are valid.

***Kleibergen's Statistic.*** Kleibergen (2001) proposed testing  $\beta = \beta_0$  using the statistic,

$$K(\beta_0) = \frac{(\hat{S}'\hat{T})^2}{\hat{T}'\hat{T}}, \quad (17)$$

which, following Moreira (2001), we have written in terms of  $\hat{S}$  and  $\hat{T}$ . If  $K = 1$ , then  $K(\beta_0) = AR(\beta_0)$ . Kleibergen shows that under either conventional or weak-instrument asymptotics,  $K(\beta_0)$  has a  $\chi_1^2$  null limiting distribution.

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

$$M(\beta_0) = \frac{1}{2} \left( \hat{S}'\hat{S} - \hat{T}'\hat{T} + \sqrt{(\hat{S}'\hat{S} - \hat{T}'\hat{T})^2 - 4[(\hat{S}'\hat{S})(\hat{T}'\hat{T}) - (\hat{S}'\hat{T})^2]} \right). \quad (18)$$

The (weak-instrument) asymptotic conditional distribution of  $M(\beta_0)$  under the null, given  $\hat{T}$ , is nonstandard and depends on  $\beta_0$ , and Moreira (2002) suggests computing the distribution by Monte Carlo simulation.

### 5.3 Power Comparisons

We now turn to a comparison of the weak-instrument asymptotic power of the Anderson-Rubin, Kleibergen, and Moreira tests. The asymptotic power functions of these tests depend on  $\mu^2/K$ ,  $\rho$  (the correlation between  $u$  and  $v$  in Equations (1) and (2)), and  $K$ , as well as the true value of  $\beta$ . We consider two values of  $\mu^2/K$ :  $\mu^2/K = 1$  corresponds to very weak instruments (nearly unidentified), and  $\mu^2/K = 5$  corresponds to moderately weak instruments. The two values of  $\rho$  considered correspond to moderate endogeneity ( $\rho = .5$ ) and very strong endogeneity ( $\rho = .99$ ).

Figure 2 presents weak-instrument asymptotic power functions for  $K = 5$  instruments, so that the degree of overidentification is 4; the power depends on  $\beta - \beta_0$  but not on  $\beta_0$ , so Figure 2 applies to general  $\beta_0$ . Moreira's (2001) infeasible asymptotic Gaussian power envelope is also reported as a basis for comparison. When  $\mu^2/K = 1$  and  $\rho = .5$ , all tests have poor power for all values of the parameter space, a result that is reassuring given how weak the instruments are; moreover, all tests have power functions that are far from the infeasible power envelope. Another unusual feature of these tests is that the power function does not increase monotonically as  $\beta$  departs from  $\beta_0$ . The

relative performance of the three tests differs, depending on the true parameter values, and the power functions occasionally cross. Typically (but not always), the  $K(\beta_0)$  and  $M(\beta_0)$  tests outperform the  $AR(\beta_0)$  test.

Figure 3 presents the corresponding power functions for many instruments ( $K = 50$ ). For  $\mu^2/K = 5$ , both the  $K(\beta_0)$  and  $M(\beta_0)$  tests are close to the power envelope for most values of  $\beta - \beta_0$  (except, oddly, when  $\beta \ll \beta_0$ , the power function of the  $K(\beta_0)$  test descends from one to approximately one-half). As Figure 3 makes clear, when  $K$  is large the  $AR$  statistic has relatively low power and substantial power improvements are possible using the other statistics.

These figures demonstrate the power improvements possible, relative to the  $AR(\beta_0)$  statistic. In addition, they show that the  $K(\beta_0)$  and  $M(\beta_0)$  tests have power functions that are close to the infeasible power envelope even for moderate values of the concentration parameter, particularly when the degree of endogeneity is large.

#### 5.4 Conservative Tests

Two other tests that have been proposed are Bonferroni tests and a bounded score test. Under weak-instrument asymptotics, both methods are conservative in the sense that their size is less than their significance level for some values of the parameters.

***Bonferroni methods.*** Staiger and Stock (1997) show how the Bonferroni inequality can be used to construct a test of the hypothesis that  $\beta = \beta_0$  that is robust to weak instruments. Their proposal entails using the first-stage  $F$ -statistic to obtain a 97.5% confidence interval for  $\mu^2/K$ , then computing confidence intervals for  $\beta$  as the

union of the TSLS (or LIML) 97.5% confidence intervals, conditional on the non-rejected values of  $\mu^2/K$ .

***The Wang-Zivot (1998) Score Test.*** Wang and Zivot (1998) and Zivot, Startz and Nelson (1998) consider testing  $\beta = \beta_0$  using modifications of conventional GMM test statistics, in which  $\sigma_u^2$  is estimated under the null hypothesis. Using weak-instrument asymptotics, they show that although these statistic are not asymptotically pivotal, their null distributions are bounded by a  $F_{K,\infty}$  distribution (the bound is tight if  $K = 1$ ), which permits a valid, but conservative, test.

## **5.5 Robust Confidence Sets**

By 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  $\beta_0$  for which the Anderson-Rubin (1949) 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 (a quadratic, in the case of the  $AR$  statistic). Because the tests in this section are fully robust to weak instruments, the confidence sets constructed by inverting the tests are fully robust.

As a general matter, when the instruments are weak, these sets can have infinite volume. For example, because the  $AR$  statistic is a ratio of quadratic forms, 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 these methods

waste information, or are unnecessarily imprecise; on the contrary, they reflect the fact that, if instruments are weak, there simply is limited information to use to make inferences about  $\beta$ . This point is made formally by Dufour (1997), who shows that under weak instrument asymptotics a confidence set for  $\beta$  must have infinite expected volume, if it is to have nonzero coverage uniformly in the parameter space, as long as  $\mu^2$  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; for additional discussion see Zivot, Startz and Nelson (1998).

## **6. Partially Robust Inference with Weak Instruments**

Although the fully robust tests discussed in the previous section always control size, they have some practical disadvantages, for example some are difficult to compute. Moreover, for  $n > 1$  they do not readily provide point estimates, and confidence intervals for individual elements of  $\beta$  must be obtained by conservative projection methods. For this reason, some researchers have investigated inference that is partially robust to weak instruments. Recall from Section 4 that we stated that whether instruments are weak depends on the task to which they are put. Accordingly, one way to frame the investigation of partially robust methods is to push the ‘weak instrument threshold’ as far as possible below that needed for TSLS to be reliable, yet not require that the method produce valid inference in the completely unidentified case.

### **6.1 $k$ -class Estimators**

The  $k$ -class estimator of  $\beta$  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 upon TSLS when instruments are weak.

**LIML.** LIML is a  $k$ -class estimator where  $k = k_{\text{LIML}}$  is the smallest root of the determinantal equation  $|\underline{Y}'\underline{Y} - k\underline{Y}'\underline{M}_Z\underline{Y}| = 0$ . The mean of distribution of the LIML estimator does not exist, and its median is typically much closer to  $\beta$  than is the mean or median of TSLS. In the fixed-instrument, normal-error case the bias of TSLS increases with  $K$ , but the bias of LIML does not (Rothenberg (1984)). When the errors are symmetrically distributed and instruments are fixed, LIML is the best median-unbiased  $k$ -class estimator to second order (Rothenberg (1983)). Moreover, LIML is consistent under many-instrument asymptotics (Bekker (1994)), while TSLS is not. This bias reduction comes at a price, as the LIML estimator has fat tails. For example, LIML generally has a larger interquartile range than TSLS when instruments are weak (e.g. Hahn, Hausman and Kuersteiner (2001a)).

**Fuller- $k$  estimators.** Fuller (1977) proposed an alternative  $k$ -class estimator which 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, Hausman, and Kuersteiner (2001a) report substantial reductions in bias and mean squared error, relative to TSLS and LIML, using Fuller-  $k$  estimators when instruments are weak.

**Bias-adjusted TSLS.** Donald and Newey (2001) consider a bias-adjusted TSLS estimator (BTSLS), which is a  $k$ -class estimator with  $k = T/(T - K + 2)$ , modifying an estimator previously proposed by Nagar (1959). Rothenberg (1984) shows that BTSLS is

unbiased to second order in the fixed-instrument, normal-error model. Donald and Newey provide expressions for the second-order asymptotic mean square error (MSE) of BTSLs, TSLS and LIML, as a function of the number of instruments  $K$ . They propose choosing the number of instruments to minimize the second-order MSE. In Monte-Carlo simulations, they find that selecting the number of instruments in this way generally improves performance. Chao and Swanson (2001) develop analogous expressions for bias and mean squared error of TSLS under weak instrument asymptotics, modified to allow the number of instruments to increase with the sample size; they report improvements in Monte Carlo simulations by incorporating bias adjustments.

### 6.3 The SSIV and JIVE Estimators

Angrist and Krueger (1992, 1995) proposed eliminating the bias of TSLS by splitting the sample into two independent subsamples, then running the first-stage regression on one subsample and the second stage regression on the other. The SSIV estimator is given by an OLS regression of  $y^{[1]}$  on  $\hat{\Pi}^{[2]} Z^{[1]}$ , where the superscript denotes the subsample. Angrist and Krueger (1995) show that SSIV is biased towards zero, rather than towards the OLS probability limit.

Because SSIV does not use the sample symmetrically and appears to waste information, Angrist, Imbens and Krueger (1999) proposed a jackknife instrumental variables (JIVE) estimator. The JIVE estimator is  $\hat{\beta}^{\text{JIVE}} = (\tilde{Y}'Y)^{-1}\tilde{Y}'y$ , where the  $i^{\text{th}}$  row of  $\tilde{Y}$  is  $Z_i\hat{\Pi}_{-i}$  and  $\hat{\Pi}_{-i}$  is the estimator of  $\Pi$  computed using all but the  $i^{\text{th}}$  observation. Angrist, Imbens and Krueger show that JIVE and TSLS are asymptotically equivalent under conventional fixed-model asymptotics. Calculations 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), and Blomquist and Dahlberg (1999)) indicate that JIVE improves upon TSLS when there are many instruments

## 6.4 Comparisons

One way to assess the extent to which a proposed estimator or test is robust to weak instruments is to characterize the size of the weak instrument region. When  $n = 1$ , this can be done by computing the value of  $\mu^2/K$  needed to ensure that inferences attain a desired degree of reliability. This was the approach taken in Table 1, and here we extend this approach to some of the estimators discussed in this section.

Figure 4 reports boundaries of asymptotic weak-instrument sets, as a function of  $K$ , for various estimators (for computational details, see Stock and Yogo (2001)). In Figure 4a, the weak-instrument set is defined to be the set of  $\mu^2/K$  such that the relative bias of the estimator (relative to the inconsistency of OLS) exceeds 10%; the values of  $\mu^2/K$  plotted are the largest values that meet this criterion. In Figure 4b, the weak-instrument set is defined so that a nominal 5% test of  $\beta = \beta_0$ , based on the relevant  $t$ -statistic, rejects more than 15% of the time (that is, has size exceeding 15%).

Two features of Figure 4 are noteworthy. First, LIML, BTSLs, JIVE, and the Fuller- $k$  estimator have much smaller weak-instrument thresholds than TSLS; in this sense, these four estimators are more robust to weak instruments than TSLS. Second, these thresholds do not increase as a function of  $K$ , whereas the TSLS threshold increases in  $K$ , a reflection of its greater bias as the number of instruments increases.



## 7. GMM 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, Heaton and Yaron (1996) examine GMM estimators of various intertemporal asset pricing models using a Monte Carlo design calibrated to match U.S. data. They find that, in many cases, inferences based on the usual normal distributions are misleading (also see Tauchen (1986), Kocherlakota (1990), Ferson and Foerester (1994) and Smith (1999)).

The foregoing discussion of weak instruments in the linear model suggests that weak instruments could be one possible reason for the failure of the conventional normal approximations in nonlinear GMM. In the linearized CCAPM Euler equation (11), both the log gross asset return  $r_t$  and the growth rate of consumption  $\Delta c_t$  are difficult to predict; thus, as argued by Stock and Wright (2000) and Neely, Roy, and Whiteman (2001), it stands to reason that estimation of the original nonlinear Euler equation by GMM also suffers from weak instruments. But making this intuition precise in the general nonlinear setting is difficult: the machinery discussed in the previous sections relies heavily on the linear regression model. In this section, 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. The general approaches to handling the problem of weak identification are the same in the nonlinear setting as the linear

setting: detection of weak identification, procedures that are fully robust to weak identification, and procedures that are partially robust. As shall be seen, the literature on this topic is quite incomplete.

## 7.1 Consequences of Weak Identification in Nonlinear GMM

In GMM estimation, the  $n \times 1$  parameter vector  $\theta$  is identified by the  $G$  conditional mean conditions  $E[h(Y_t, \theta_0)|Z_t] = 0$ , where  $\theta_0$  is the true value of  $\theta$  and  $Z_t$  is a  $K$ -vector of instruments. The GMM estimator is computed by minimizing

$$S_T(\theta) = \left[ \sqrt{\frac{1}{T}} \sum_{t=1}^T \phi_t(\theta) \right]' W(\theta)^{-1} \left[ \sqrt{\frac{1}{T}} \sum_{t=1}^T \phi_t(\theta) \right], \quad (19)$$

where  $\phi_t(\theta) = h(Y_t, \theta) \otimes Z_t$  is  $r \times 1$ ,  $W(\theta)$  is a  $r \times r$  positive definite matrix, and  $r = GK$ . In the two-step GMM estimator,  $W(\theta)$  initially is set to the identity matrix, yielding the estimator  $\hat{\theta}^{(1)}$ , and the second step uses  $\hat{W}(\hat{\theta}^{(1)})$ , where

$$\hat{W}(\theta) = \frac{1}{T} \sum_{t=1}^T [\phi_t(\theta) - \bar{\phi}(\theta)][\phi_t(\theta) - \bar{\phi}(\theta)]' \quad (20)$$

and  $\bar{\phi}(\theta) = T^{-1} \sum_{t=1}^T \phi_t(\theta)$ . (Here, we assume that  $\phi_t(\theta)$  is serially uncorrelated, otherwise  $\hat{W}(\theta)$  is replaced by an estimator of the spectral density of  $\phi_t(\theta)$  at frequency zero.) The iterated GMM estimator continues this process, evaluating  $\hat{W}(\theta)$  at the

previous estimate of  $\theta$  until convergence (see Ferson and Foerester (1994)). For additional details about GMM estimation, see Hayashi (2000).

To understand weak identification in nonlinear GMM, it is useful to return for a moment to the linear model, in which case we can write the moment condition as

$$E[(Y_{1t} - \theta_0 Y_{2t})Z_t] = 0 \quad (21)$$

where, in a slight shift in notation,  $Y_{1t}$  and  $Y_{2t}$  correspond to  $y_t$  and  $Y_t$  in equations (1) and (2) and the parameter vector is  $\theta$  rather than  $\beta$ . The reason that the instruments serve to identify  $\theta$  is that the orthogonality condition (21) holds at the true value  $\theta_0$ , but it does not hold at other values of  $\theta_0$ . If the instrument is irrelevant, so that  $Y_{2t}$  is uncorrelated with  $Z_t$ , then  $E[(Y_{1t} - \theta Y_{2t})Z_t] = 0$  for all values of  $\theta$  and (21) no longer identifies  $\theta_0$  uniquely. If the instruments are weak, then  $E[(Y_{1t} - \theta Y_{2t})Z_t]$  is nearly zero for all values of  $\theta$ , and in this sense  $\theta$  can be said to be weakly identified. Said differently, weak instruments imply that the correlation between the model error term  $Y_{1t} - \theta Y_{2t}$  and the instruments is nearly zero, even at false values of  $\theta$ .

This intuition of weak instruments implying weak identification carries over to the nonlinear GMM setting: if the correlation between the model error term,  $h(Y_t, \theta)$ , and  $Z_t$  is low even for false values of  $\theta$ , then  $\theta$  is weakly identified.

Because there is no exact sampling theory for GMM estimators, a formal treatment of the implications of weak identification for GMM must entail asymptotics. As in the linear case, there are two approaches. One is to use stochastic expansions in orders of  $T^{-1/2}$ , an approach that has been pursued by Newey and Smith (2001). This

approach, however, seems likely to produce poor approximations when identification is very weak, as it does in the linear case.

A second approach is to use an asymptotic nesting lets  $T \rightarrow \infty$  but, loosely speaking, keeps the GMM version of the concentration parameter constant. This approach is developed in Stock and Wright (2000), who develop a stochastic process representation of the limiting objective function (the limit of  $S_T$ , where  $S_T$  is treated as a stochastic process indexed by  $\theta$ ) that holds formally in weakly identified, partially identified, and non-identified cases. Stock and Wright's numerical work suggests that weak identification can explain many of the Monte Carlo results in Tauchen (1986), Kocherlakota (1990) and Hansen, Heaton, and Yaron (1996).

## **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 have yield quite different confidence sets. If identification is weak, GMM estimates can be sensitive to the addition of instruments or changes in the sample. To the extent that these features are present in an empirical application, they could suggest the presence of weak identification.

The only formal test for weak identification that we are aware of in nonlinear GMM is that proposed by Wright (2001). In the conventional asymptotic theory of GMM, the identification condition requires that the gradient of  $\phi_i(\theta_0)$  has full column rank. Wright proposes 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, strictly is a test for non-identification or under-identification, not for weak instruments.

### 7.3 Procedures that are Fully Robust to Weak Identification

We are aware of only two fully robust methods for hypothesis tests and confidence sets in nonlinear GMM: one based on a nonlinear extension of the Anderson-Rubin (1949) statistic, the other based on an extension of Kleibergen’s (2001) statistic.

*Nonlinear AR statistic.* Because the numerator and denominator of *AR* statistic (16) 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 basing tests of  $\theta = \theta_0$  on the nonlinear analog of the *AR* statistic, that is, the GMM objective function in which the weight matrix is evaluated at the same parameter value as the numerator. This special case of (19) is,

$$S_T^{CU}(\theta) = \left[ \sqrt{\frac{1}{T}} \sum_{i=1}^T \phi_i(\theta) \right]' \hat{W}(\theta)^{-1} \left[ \sqrt{\frac{1}{T}} \sum_{i=1}^T \phi_i(\theta) \right]. \quad (22)$$

Hansen, Heaton and Yaron (1996) suggested minimizing  $S_T^{CU}(\theta)$  to obtain what they call the continuous-updating estimator (discussed below), but for now we focus on tests based on  $S_T^{CU}(\theta_0)$ .

Under the null hypothesis  $\theta = \theta_0$ ,  $S_T^{CU}(\theta_0)$  is asymptotically distributed  $\chi_r^2$  whether identification is weak or strong (Stock and Wright (2000)). If the instruments

are relevant, under the alternative that  $\theta \neq \theta_0$ , the “numerator moments” of  $S_T^{CU}(\theta_0)$  have nonzero expectation. Because a test based on  $S_T^{CU}(\theta_0)$  has the correct asymptotic size whether or not identification is weak, it is fully robust to weak identification.

A 95% confidence set for  $\theta$  can be computed as the set of values of  $\theta$  for which  $S_T^{CU}(\theta)$  does not reject at the 5% significance level. When identification is strong, these sets are ellipsoids and are similar to the conventional GMM confidence regions, but when identification is weak, the nonlinearities in  $S_T^{CU}(\theta)$  produce irregularly shaped, sometimes nonconvex sets; examples of such sets are given in Stock and Wright (2000) and Ma (2001). Stock and Wright and Ma both use the irregular shape of these “S-sets” as evidence of weak identification in the applications they examine.

***Kleibergen’s GMM statistic.*** Kleibergen (2002) proposed an alternative test statistic, which extends the  $K(\beta)$  statistic (17) to nonlinear GMM. Let

$$\kappa(\theta) = \left[ \sqrt{\frac{1}{T}} \sum_{i=1}^T \phi_i(\theta) \right] \hat{W}(\theta)^{-1/2} P_{\hat{W}(\theta)^{-1/2} D(\theta)} \hat{W}(\theta)^{-1/2}' \left[ \sqrt{\frac{1}{T}} \sum_{i=1}^T \phi_i(\theta) \right], \quad (23)$$

where  $D(\theta) = \sum_{i=1}^T \left[ \frac{d\phi_i(\theta)}{d\theta} - \frac{1}{2} \phi_i(\theta)' \hat{W}(\theta)^{-1} \frac{d\text{vec}(\hat{W}(\theta))}{d\theta} \right]$ . Kleibergen proposes testing

the hypothesis  $\theta = \theta_0$  using the statistic  $\kappa(\theta_0)$  and shows that this statistic has a  $\chi_n^2$  distribution under both conventional asymptotics and the weak identification asymptotics of Stock and Wright (2000). In the just identified case ( $n = r$ ),  $S_T^{CU}(\theta_0) = \kappa(\theta_0)$ , but otherwise these two test statistics are different. An appealing feature of this statistic is

that, under conventional asymptotics, the confidence set formed by inverting the acceptance region of  $\kappa(\theta_0)$  is the usual GMM confidence ellipse; however, under weak identification  $\kappa$ -sets in general have irregular shapes. In the special case of the linear IV model, the statistic  $\kappa(\theta_0)$  simplifies to  $K(\theta_0)$  in (17). Kleibergen finds in Monte-Carlo simulations that  $\kappa(\theta_0)$  generally gives a more powerful test than  $S_T^{CU}(\theta_0)$ , consistent with the findings reported in Section 5 for the linear IV regression model.

#### 7.4 Procedures that are Partially Robust to Weak Identification

Because there are estimators that improve upon TSLS when instruments are weak in the linear case, it stands to reason that there should be estimators that improve upon two-step GMM in the nonlinear case, and the limited work in this area to date has some promising results. Two GMM estimators that appear to be partially robust to weak instruments are the continuous updating estimator and empirical likelihood estimators.

*The Continuous Updating Estimator.* Hansen, Heaton, and Yaron (1996) proposed minimizing  $S_T^{CU}(\theta)$  in (22), and they termed the resulting estimator the continuous updating estimator (CUE). In the linear model, the CUE is asymptotically equivalent to LIML under weak-instrument or conventional asymptotics if the errors are homoskedastic (Stock and Wright (2000); this obtains from the asymptotic equivalence of (16) and (22) when  $\phi_t(\theta) = (y_t - \theta Y_t)Z_t$ ). Donald and Newey (1999) show that the first-order conditions for the continuous updating estimator have a jackknife interpretation.

This asymptotic equivalence of the CUE and LIML in the homoskedastic linear model suggests that the CUE could bring some of the desirable properties of LIML (less bias and better coverage rate than TSLS) to nonlinear GMM. Monte Carlo simulations in

Hansen, Heaton, and Yaron (1996) and Stock and Wright (2000) suggest that this is the case, but that (like LIML) CUE has heavy tails and can produce extreme estimates under weak identification.

***Empirical Likelihood and Generalized Empirical Likelihood.*** Empirical likelihood and generalized empirical likelihood estimation is discussed in detail in the article by Imbens (2002) in this issue, so our treatment here is brief. The generalized empirical likelihood (GEL) estimator (Smith (1997)) solves the minimization problem,

$$\hat{\theta}^{\text{GEL}} = \operatorname{argmax}_{\theta} \inf_d \sum_{t=1}^T g[d'\phi_t(\theta)], \quad (24)$$

where  $g$  is some function such that  $g'(0) = g''(0) = 1$  (for technical details see Newey and Smith (2001)). The GEL estimator generalizes the empirical likelihood estimator (Owen (1988), DiCiccio, Hall and Romano (1991)) for which  $g(x) = \log(1 + x)$ . The GEL family includes other interesting special cases. In particular, if  $g(x) = \frac{1}{2}(1 + x^2)$ , then the GEL estimator is the CUE.

Newey and Smith (2001), building on work of Rothenberg (1999) for the linear case, derive expressions for the second order asymptotic bias of the usual two-step GMM estimator and of the GEL estimator. They find that, like the second-order bias of TSLS in the linear case, the second-order bias of two-step GMM increases linearly with the number of moment conditions ( $r$ ), but the second-order bias of GEL estimators does not depend on  $r$ . Thus 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 estimator. Work on GEL estimators in the context of weak instruments is promising but young, and 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 that appears to be commonplace is the presence of weak instruments/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 homoskedastic errors and one endogenous regressor, applied researchers at least should 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 less than 10, and if the errors appear to be homoskedastic and serially uncorrelated errors, then either a fully robust method (our preference) from Section 4 or a partially robust method from Section 5 can be used. Even if  $F > 10$ , it is prudent to check your results using LIML, BTSLs, JIVE, or the Fuller- $k$  estimator, especially when the number of instruments is large. In the GMM case (the moments are nonlinear in the parameters, the errors are heteroskedastic, and/or the errors are serially uncorrelated), 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 some 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), and Donald and Newey (2001)). A second topic we have not broached is estimation of linear panel data models with a lagged dependent variable, in which instruments (lags) are weak if the coefficient is almost one; recent work includes Kiviet (1995), Alonso-Borrego and Arellano (1996), and Hahn, Hausman and Kuersteiner (2001b). A third topic not covered here is what happens when there are weak instruments *and* a failure of the exogeneity restrictions, a situation emphasized by Bound, Jaeger, and Baker (1995). In this situation, the conventional  $J$  statistic can have poor power (as well as size). One approach is to use the  $AR$  confidence sets (or  $S$ -sets in the nonlinear case), which will be null if the violation of the orthogonality conditions is sufficiently strong. The vast majority of theoretical papers in this literature, however, work under the maintained hypothesis that the orthogonality conditions hold, and much work remains.

## References

- Alonso-Borrego, C. and Arellano, M. (1996), “Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data,” *Journal of Business and 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, Series 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.
- Angrist, J.D. and Krueger, A.B. (1992), “The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples,” *Journal of the American Statistical Association*, 87, 328-336.
- Angrist, J.D. and Krueger, A.B. (1995), “Split Sample Instrumental Variables Estimates of the Return to Schooling,” *Journal of Business and Economic Statistics*, 13, 225-235.

- 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, Dept. of Economics, University of Groningen.
- 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," manuscript, Purdue University.
- Chao, J. and Swanson, N.R. (2002), "Consistent Estimation with a Large Number of Weak Instruments," 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.
- Donald, S.G. and Newey, W.K. (1999), "A Jackknife Interpretation of the Continuous Updating Estimator," manuscript, Department of Economics, MIT.
- 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. (1997), "The (Un)Importance of Forward-Looking Behavior in Price Specifications," *Journal of Money, Credit and Banking*, 29, 338-350.
- Fuhrer, J.C. and Moore, G.R. (1995), "Inflation Persistence," *Quarterly Journal of Economics*, 110, 127-159.
- 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," manuscript, Dept. of Economics, MIT.
- Hahn, J., Hausman, J. and Kuersteiner, G. (2001b), "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects," manuscript, Dept. of Economics, MIT.
- Hall, A.R. and Inoue, A. (2001), "A Canonical Correlations Interpretation of Generalized Method of Moments Estimation with Applications to Moment Selection," manuscript, North Carolina State University.
- Hall, A.R. and Peixe, F.P.M. (2001), "A Consistent Method for the Selection of Relevant Instruments," 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. and Singleton, K.J. (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269-1286.

- Hansen, L.P. and Singleton, K.J. (1983), "Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249-265.
- Hansen, L.P., Heaton, J. and Yaron, A. (1996), "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262-280.
- Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton.
- Imbens, G.W. (2002), "Generalized Method of Moments and Empirical Likelihood", *Journal of Business and 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.
- Kleibergen, F. (2002), "Testing Parameters in GMM Without Assuming That They are Identified," manuscript, Department of Economics, University of Amsterdam.
- 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. (2001), "GMM Estimation of the New Phillips Curve," *Economics Letters*, forthcoming.

- Mavroeidis, S. (2001), "Identification and Misspecification Issues in Forward Looking Monetary Models," manuscript, Department of Economics, Oxford University.
- Moreira, M.J. (2001), "Tests with Correct Size when Instruments can be Arbitrarily Weak," manuscript, Department of Economics, UC Berkeley.
- Moreira, M.J. (2002), "A Conditional Likelihood Ratio Test for Structural Models," manuscript, Department of Economics, UC Berkeley.
- 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 and 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.
- Nelson, C.R. and Startz, R. (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," manuscript, Department of Economics, MIT.



- Owen, A. (1988), "Empirical Likelihood Ratios Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.
- DiCiccio, T., Hall, P., and Romano, J. (1991), "Empirical Likelihood is Bartlett-Correctable," *Annals of Statistics*, 19, 1053–1061.
- Phillips, P.C.B. (1984), "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, Vol. 1, eds. Griliches, Z. and Intriligator, M.D., North-Holland, Amsterdam.
- Richardson, D.H. (1968), "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214-1226.
- Roberts, J.M. (1997), "Is Inflation Sticky?," *Journal of Monetary Economics*, 39, 173-196.
- Rothenberg, T.J. (1983), "Asymptotic Properties of Some Estimators in Structural Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, eds. Karlin, S., Amemiya, T. and Goodman, L.A., Academic Press, New York.
- Rothenberg, T.J. (1984), "Approximating the Distribution of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, Vol. 2, eds. Griliches, Z. and Intriligator, M.D., North-Holland, Amsterdam.
- Rothenberg, T.J. (1999), "Higher Order Properties of Empirical Likelihood for Simultaneous Equations," manuscript, Department of Economics, UC Berkeley.
- 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," *Review of Economics and Statistics*, 79, 348-352.
- Sims, C.A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1-48.
- Smith, R. (1997), "Alternative Semiparametric Likelihood Approaches to Generalized method of Moments Estimation," *Economic Journal*, 107, 503-519.
- Smith, D.C. (1999), "Finite Sample Properties of Tests of the Epstein-Zin Asset Pricing Model," *Journal of Econometrics*, 93, 113-148.
- Staiger, D. and Stock, J.H. (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.
- Stock, J.H. and Watson, M.W. (1999), "Forecasting Inflation," *Journal of Monetary Economics*, 44, 293-335.
- Stock, J.H. and Wright, J.H. (2000), "GMM with Weak Identification," *Econometrica*, 68, 1055-1096.
- Stock, J.H. and Yogo, M. (2001), "Testing for Weak Instruments in Linear IV Regression," 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 and 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," International Finance Discussion Paper, Federal Reserve Board.

Yogo, M. (2002), “Estimating the Elasticity of Intertemporal Substitution when Instruments are Weak,” manuscript, Department of Economics, Harvard University.

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.

**Table 1**  
**Selected Critical Values for Weak Instrument Tests for TSLS**  
**Based on the First-stage  $F$ -statistics**

Number of instruments ( $K$ )	Relative bias > 10%		Nominal size of 5% test > 15%	
	Threshold $\mu^2/K$	$F$ -statistic 5% critical value	Threshold $\mu^2/K$	$F$ -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  $\mu^2/K$  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  $\mu^2/K$  is less than or equal to the value in the second column, against the alternative that  $\mu^2/K$  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).

## Figure headings

**Figure 1.** pdf of TOLS estimator (1a) and  $t$ -statistic (1b)  $\mu^2 = 0, .25, 10, 100$ , one instrument ( $K = 1$ ), and  $\rho = .99$ , computed by Monte Carlo simulation.

**Figure 2.** Weak-instrument asymptotic power of Gaussian similar tests (Anderson-Rubin (1949), Kleibergen (2001), and Moreira (2002)) and the infeasible Gaussian power envelope with  $K = 5$  instruments for  $\mu^2/K = 1, 5$  and  $\rho = .5, .99$ .

**Figure 3.** Weak-instrument asymptotic power of Gaussian similar tests with  $K = 50$  instruments.

**Figure 4.** Weak-instrument threshold values of  $\mu^2/K$  for the TOLS, LIML, BTOLS, JIVE, and Fuller- $k$  estimators as a function of the number of instruments ( $K$ ). The threshold is the value of  $\mu^2/K$  at which:

- (a) the weak-instrument asymptotic relative bias of the estimator is 10%;
- (b) the weak-instrument asymptotic size of the 5% Wald test is at most 15%.

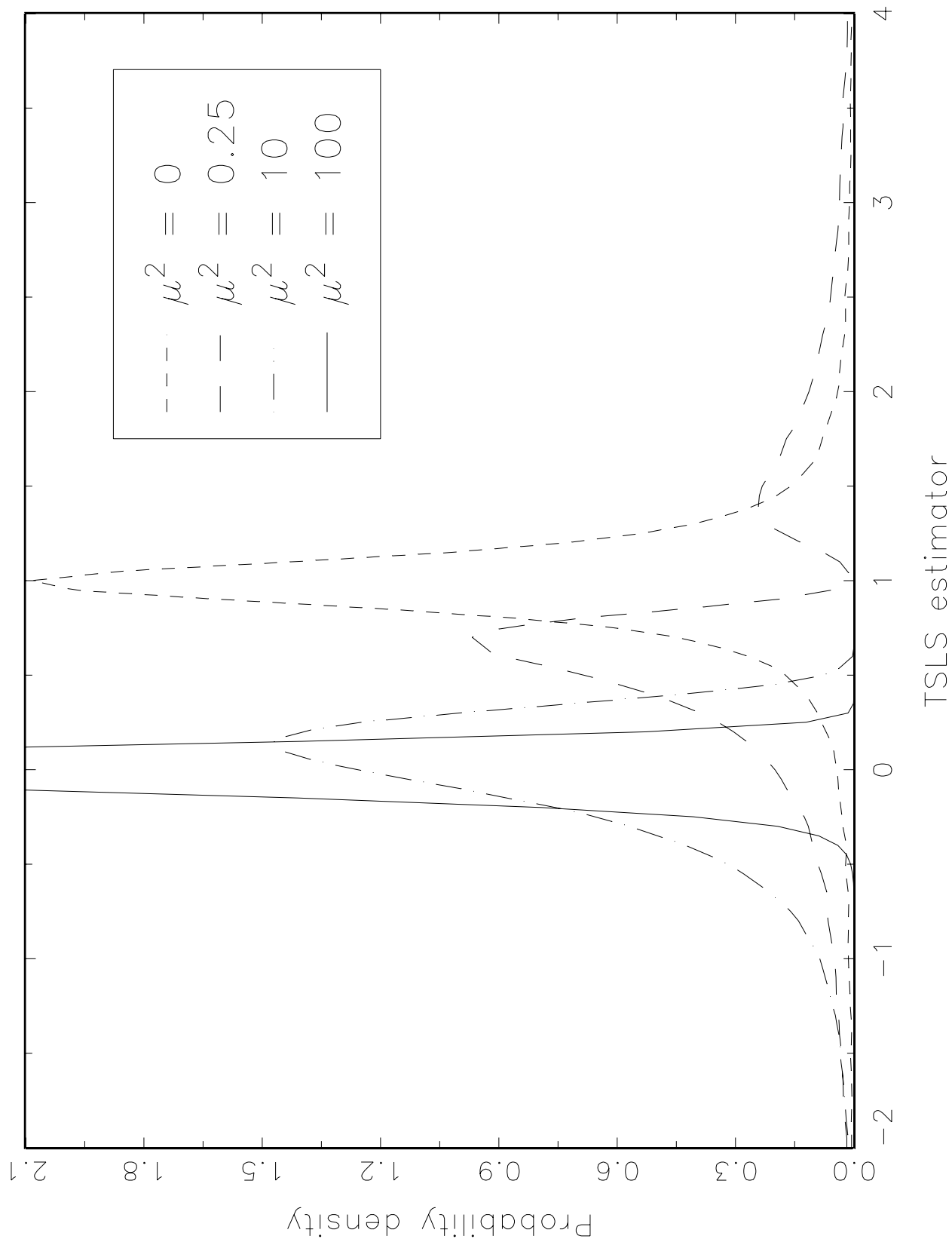


Figure 1a

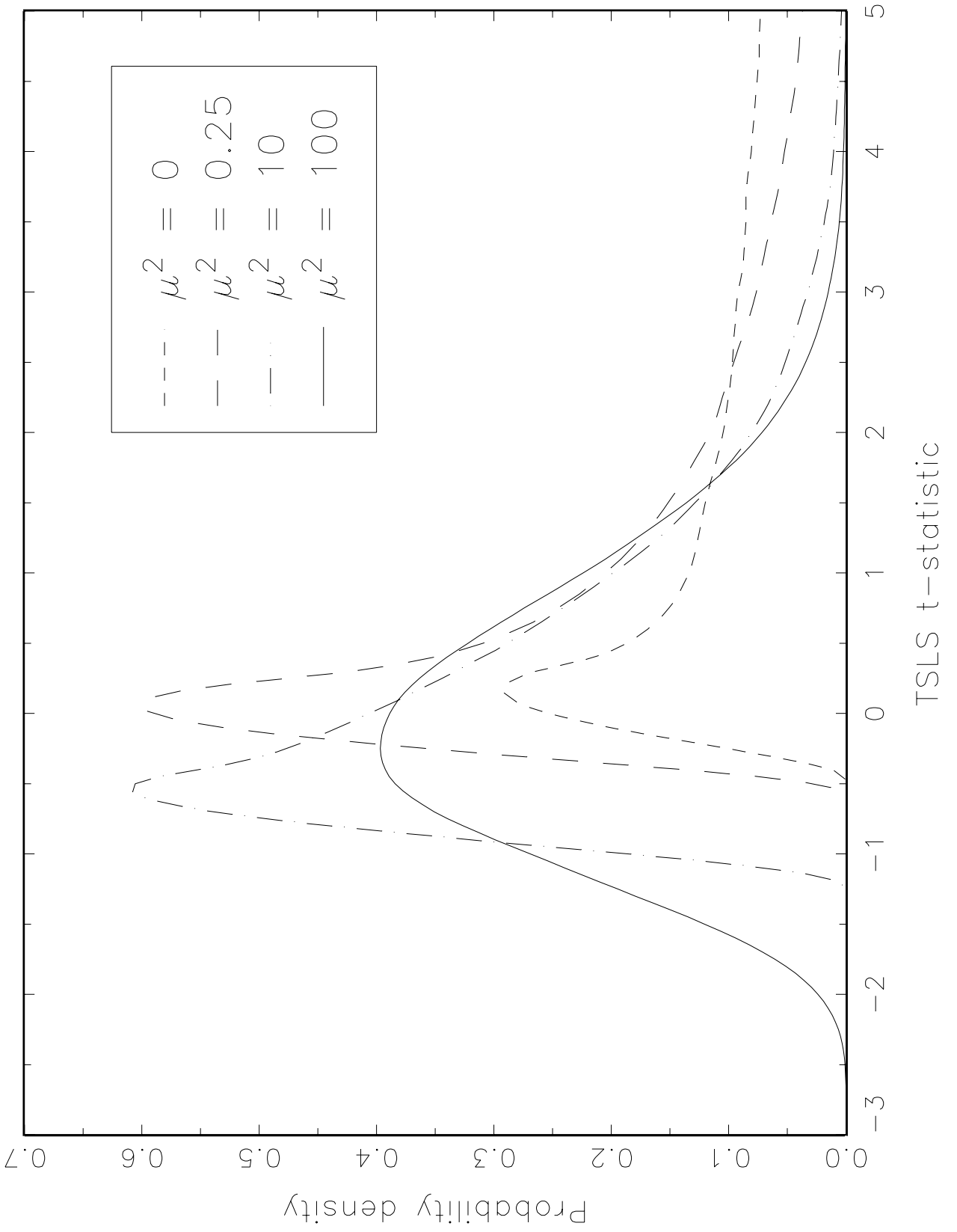


Figure 1b

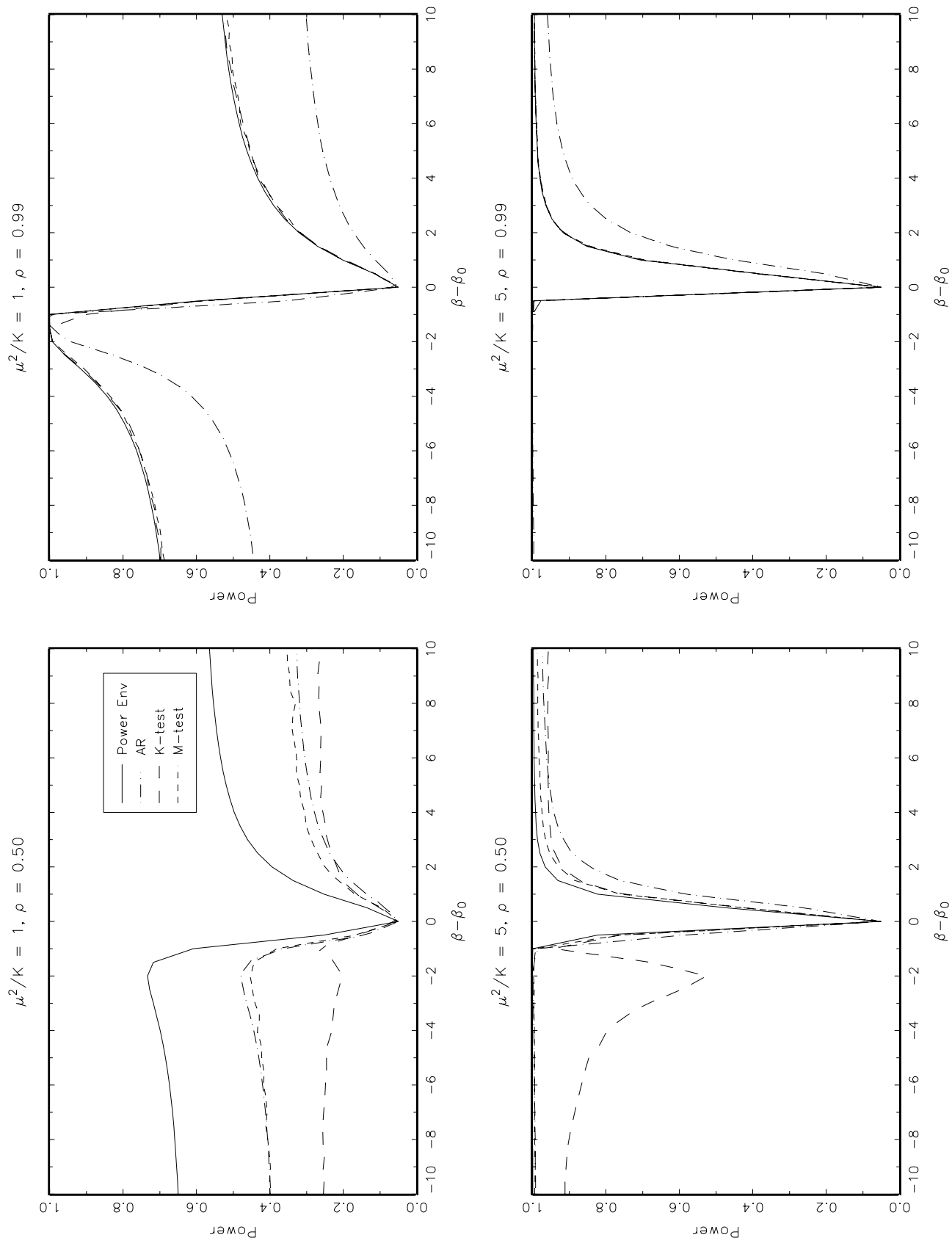


Figure 2



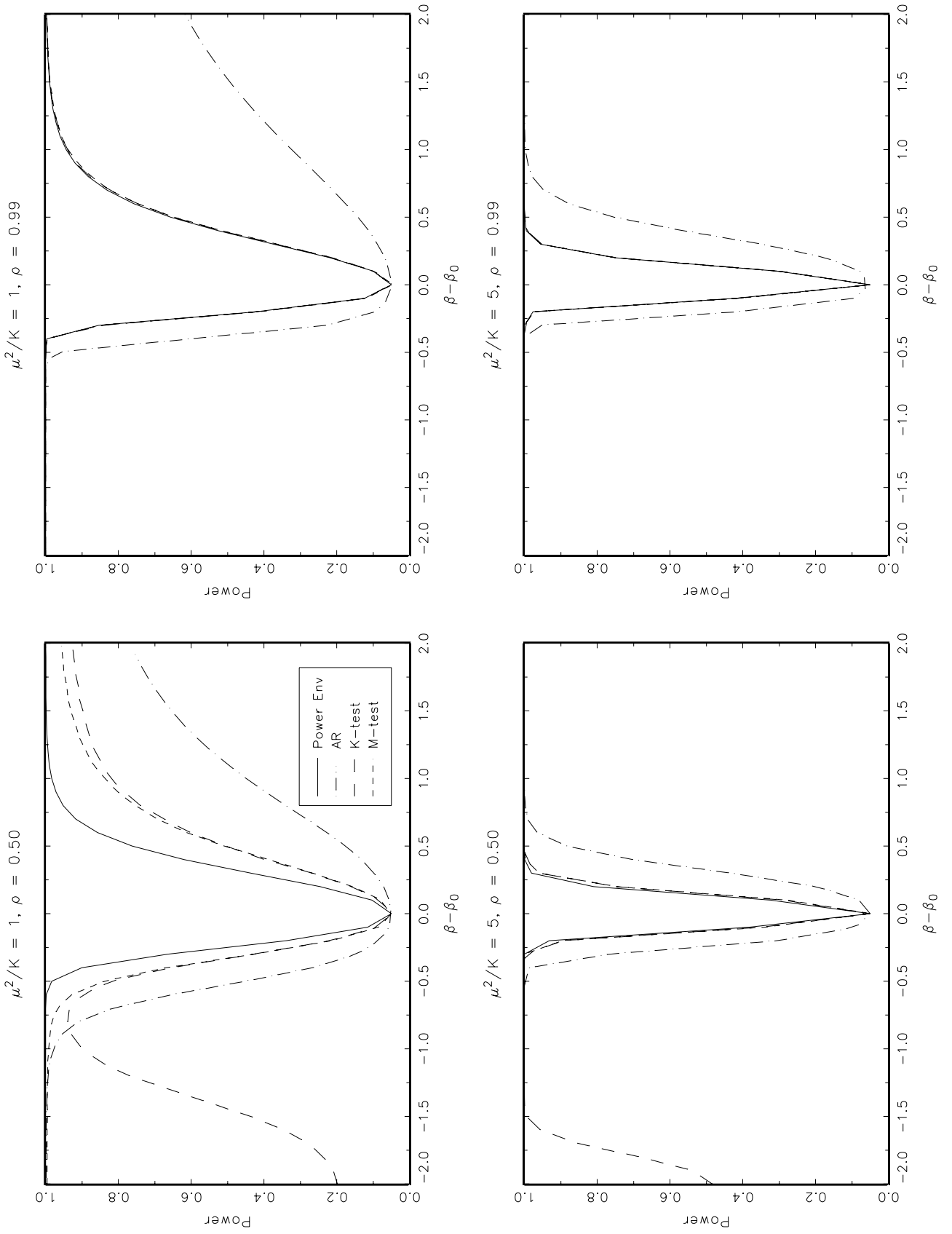


Figure 3

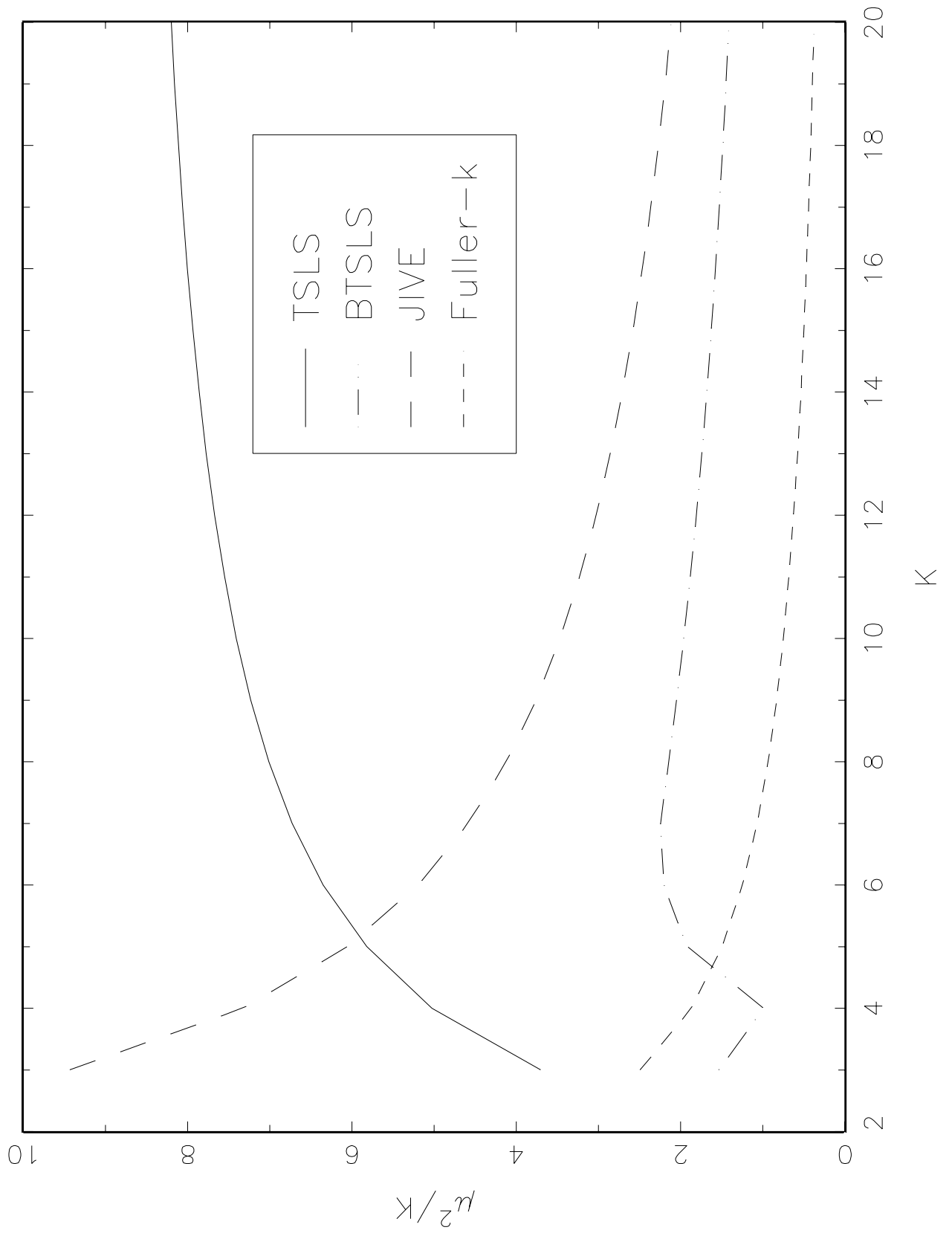


Figure 4a

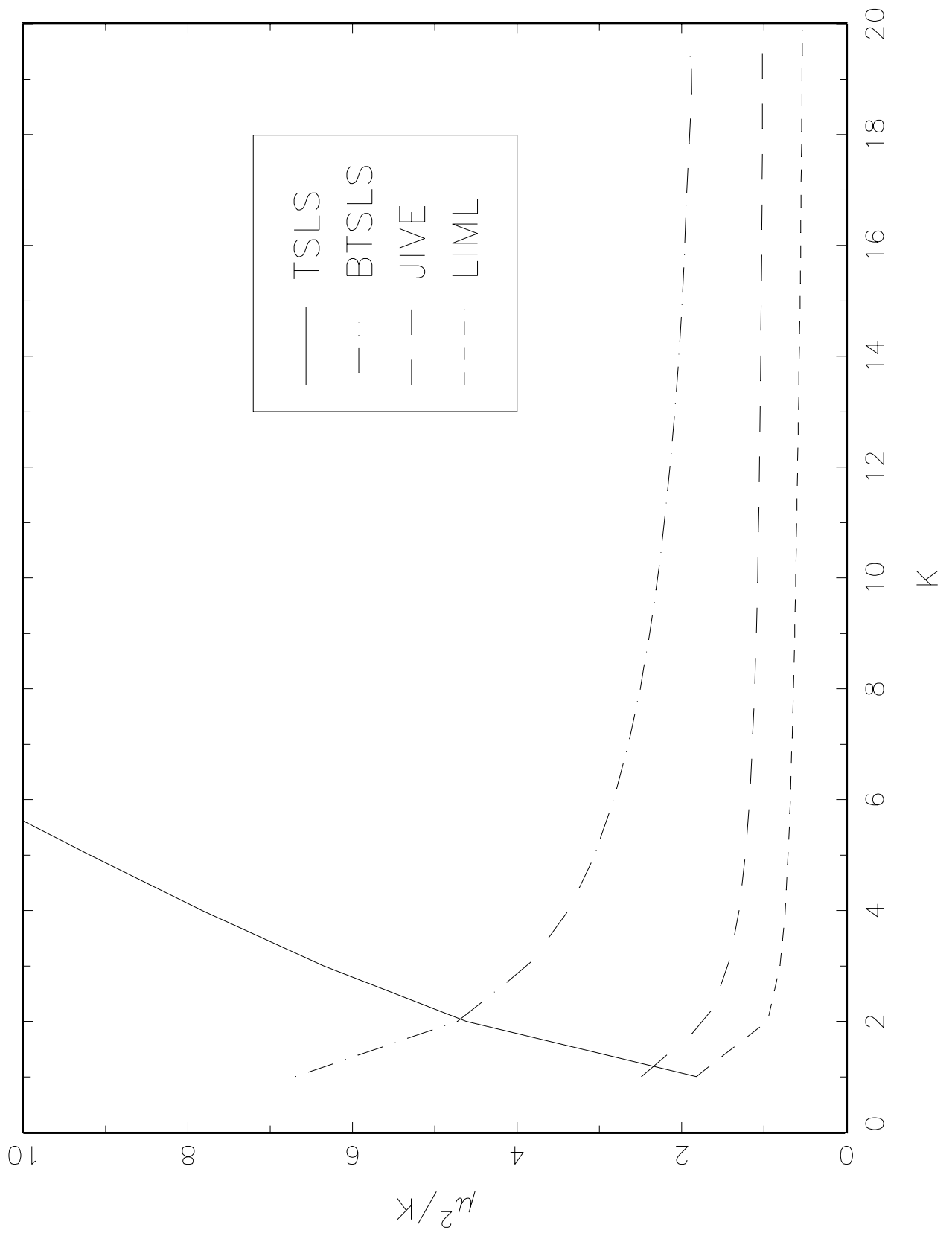


Figure 4b