Consistent Estimation with a Large Number of Weak Instruments*

John C. Chao — University of Maryland
Norman R. Swanson — Rutgers University

first draft: June 2001
this revision: February 2005

Abstract

This paper analyzes the conditions under which consistent estimation can be achieved in instrumental variables (IV) regression when the available instruments are weak, and the number of instruments, K_n, goes to infinity with the sample size. We show that consistent estimation depends importantly on the strength of the instruments as measured by r_n, the rate of growth of the so-called concentration parameter, and also on K_n. In particular, when $K_n \to \infty$, the concentration parameter can grow, even if each individual instrument is only weakly correlated with the endogenous explanatory variables, and consistency of certain estimators can be established under weaker conditions than have previously been assumed in the literature. Hence, the use of many weak instruments may actually improve the performance of certain point estimators. More specifically, we find that $LIML$ and $B2SLS$ are consistent when $\sqrt{K_n}/r_n \to 0$, while $2SLS$ is consistent only if $K_n/r_n \to 0$, as $n \to \infty$. These consistency results suggest that $LIML$ and $B2SLS$ are more robust to instrument weakness than $2SLS$.

JEL classification: C13, C31.
Keywords: concentration parameter, instrumental variables, k-class estimator, many weak instruments

*John C. Chao: Department of Economics, University of Maryland, College Park, MD 20742, chao@econ.umd.edu. Norman R. Swanson: Department of Economics, Rutgers University, New Brunswick, NJ 08901, nswanson@econ.rutgers.edu. The authors thank Don Andrews, Alastair Hall, Atsushi Inoue, Harry Kelejian, Ingmar Prucha, Graham Elliott, Peter Phillips, Tao Zha, Eric Zivot, two anonymous referees, and participants at the 2001 summer econometric society meeting, and workshops at North Carolina State University and Yale University for useful comments and suggestions. Special additional thanks are owed to Whitney Newey, the Co-Editor, who provided extensive and detailed comments and suggestions (including valuable guidance on how to simplify some of our proofs), all of which have led to what we believe is a much improved version of the paper. Swanson gratefully acknowledges financial support in the form of a Rutgers University Research Council grant.

1 Introduction

In recent years, the problem of weak instruments or weak identification has received considerable attention in both theoretical and applied econometrics.[1] This is the problem where the excluded instruments are only weakly correlated with the endogenous explanatory variables of an instrumental variables (IV) regression. As has been pointed out by, amongst others, Phillips (1983), Rothenberg (1984), and Stock and Yogo (2003a), a natural measure of instrument weakness (or strength) in a linear IV framework is the so-called concentration parameter. In the conventional asymptotic setup, which assumes strong instruments, the concentration parameter is taken to grow at the rate of the sample size, say n, as the latter approaches infinity. However, such an asymptotic sequence is not likely to provide good approximations in the case of weak instruments, since with weak instruments the concentration parameter is typically small relative to the sample size. To provide better asymptotic approximations in this case, Staiger and Stock, in an important 1997 paper, propose an alternative asymptotic framework which takes a "local-to-zero" parameterization of the coefficients of the instruments in the first-stage regression; that is, these coefficients are specified to be in a $n^{-1/2}$ shrinking neighborhood of zero. One feature of the Staiger-Stock framework is that the concentration parameter no longer diverges but rather stays roughly constant in expectation as n grows. Staiger and Stock (1997) show that, under this local-to-zero framework with the number of instruments fixed, the two-stage least squares ($2SLS$) and limited information maximum likelihood ($LIML$) estimators are not consistent and instead converge to nonstandard distributions.

Another important direction that research in IV regression has taken involves the study of situations where the number of available instruments is large, using an asymptotic framework that takes the number of instruments, K_n, to infinity as a function of n. This approach was first taken by Morimune (1983) and later generalized by Bekker (1994). Other influential papers in this area include Angrist and Krueger (1995), Donald and Newey (2001), Hahn, Hausman, and Kuersteiner (2001), Hahn (2002), Hahn and Inoue (2002), and Chamberlain and Imbens (2004). Amongst papers adopting the "many instruments" approach, it has until recently been standard practice to assume, as in conventional asymptotics, that the concentration parameter grows at the same rate as the sample size. Hence, these papers study scenarios where the instruments are stronger than assumed in the weak instruments literature.

In this paper, we amalgamate these two strands of the IV literature. In particular, we develop a unified framework for studying the asymptotic behavior of various k-class IV estimators in the presence of weak and/or many instruments. We observe that having many instruments in a weakly-identified situation can help to improve

[1] A small subset of the growing literature on weak instruments includes papers by Nelson and Startz (1990), Bound, Jaeger and Baker (1995), Staiger and Stock (1997), Wang and Zivot (1998), Kleibergen (2002), Bekker and Kleibergen (2003), Guggenberger and Smith (2003), Moreira (2003) and the references cited therein.

estimation accuracy, since using a large number of instruments can enhance the growth of the concentration parameter even if each individual instrument is only weakly correlated with the endogenous explanatory variables. In addition, we find that when there are many weak instruments, consistency depends importantly on the choice of estimator, as some estimators are more robust to instrument weakness than others. More specifically, we show that for certain well-centered IV estimators, such as $LIML$ and the bias-corrected two-stage least squares ($B2SLS$) estimator, consistency can be established even when instrument weakness is such that the rate of growth of the concentration parameter, r_n, is slower than K_n, and possibly much slower than the sample size n, provided that $\sqrt{K_n}/r_n \to 0$, as $n \to \infty$.[2] On the other hand, $2SLS$ turns out to be less robust to instrument weakness and is shown to be consistent only if $K_n/r_n \to 0$ as $n \to \infty$.

Our results complement those obtained in more recent papers by Han and Phillips (2003), Stock and Yogo (2003b), Chao and Swanson (2004), Hansen, Hausman, and Newey (2004), and Newey (2004), which also examine frameworks with many weak instruments (MWI) or, more generally, with many weak moment conditions. As we will discuss in more detail in the next section, relative to the other MWI papers cited above, the consistency results reported here are obtained under conditions which allow for weaker instruments, less restriction on the growth rate of K_n vis-à-vis n, and greater generality in the stochastic properties of the instruments. On the other hand, our paper has a narrower focus in the sense that we consider only consistency here, whereas other MWI papers have derived asymptotic distributional results, and, in the case of Hansen, Hausman, and Newey (2004) and Newey (2004), results on asymptotic covariance matrix estimation and for implementing Wald inference have also been obtained.

The remainder of the paper describes our setup and presents our main results. All proofs are gathered in an appendix. In the sequel, $Tr(\cdot)$ denotes the trace of a matrix; $\|A\| = \sqrt{Tr\,(A'A)}$ denotes the usual Euclidean or Frobenius norm for a matrix A; A^+ denotes the Moore-Penrose generalized inverse of a (possibly singular) matrix A; $P_Z = Z\,(Z'Z)^+\,Z'$; and " >0" and " ≥ 0" denote, respectively, positive definiteness and positive semi-definiteness when applied to matrices.

2 Setup and Main Results

Consider the simultaneous equations model (SEM):

$$y_{1n} = Y_{2n}\beta + u_n, \tag{1}$$

$$Y_{2n} = Z_n\Pi_n + V_n, \tag{2}$$

[2]We shall define more precisely what "well-centered" means in Section 2 (see Remark 2.2(v)).

where y_{1n} and Y_{2n} are, respectively, an $n \times 1$ vector and an $n \times G$ matrix of observations on the $G+1$ endogenous variables of the system, Z_n is an $n \times K_n$ matrix of observations on the K_n instrumental variables, and u_n, V_n are, respectively, an $n \times 1$ vector and an $n \times G$ matrix of random disturbances. Also, let $\eta_i = (u_i, v_i')'$ where u_i and v_i' are, respectively, the i^{th} component of u_n and the i^{th} row of V_n. Note that equations (1) and (2) really define a sequence of models, since we specify Π_n $(K_n \times G)$ to depend on n in order to model the effects of having many weak instruments. In particular, within this setup, the effect of many instruments can be modelled by letting $K_n \to \infty$ as $n \to \infty$, while the effect of weak instruments can be accounted for by shrinking Π_n toward a zero matrix as n grows.[3] The following assumptions are used in the sequel.

Assumption 1: Let $\{Z_{n,i} : i = 1, ..., n;\ n \geq 1\}$ be a triangular array of R^{K_n}-valued random variables, where $Z_{n,i}'$ denotes the i^{th} row of the matrix Z_n. Suppose that: (a) $K_n/n < 1$ for all n, and $K_n \to \infty$ as $n \to \infty$ such that $K_n/n \to \tau$ for some constant τ satisfying $0 \leq \tau < 1$; (b) there exists a positive integer N such that $\forall n \geq N$, Z_n is of full column rank K_n almost surely; and (c) there exists a non-decreasing sequence of positive real numbers $\{r_n\}$ such that, as $n \to \infty$, $r_n/n \to \kappa$, for some constant κ, with $0 \leq \kappa < \infty$, and such that $\Psi_n = \Pi_n' Z_n' Z_n \Pi_n / r_n \to \Psi$ almost surely, for some nonrandom positive definite matrix Ψ.[4]

Assumption 2: Suppose that: (a) Z_n and η_i are independent for all i and n; (b) $\{\eta_i\} \equiv i.i.d.(0, \Sigma)$, where $\Sigma > 0$, and Σ can be partitioned conformably with $(u_i, v_i')'$ as $\Sigma = \begin{pmatrix} \sigma_{uu} & \sigma_{Vu}' \\ \sigma_{Vu} & \Sigma_{VV} \end{pmatrix}$, with $\sigma_{Vu}^{(g)}$ and $\Sigma_{VV}^{(g,h)}$ denoting the g^{th} element of σ_{Vu} and the $(g,h)^{th}$ element of Σ_{VV}; respectively; and (c) there exists some positive constant $D_\eta < \infty$ such that $E\left(u_i^4\right) \leq D_\eta$ and $\max_{1 \leq\ g\ \leq G} E\left(v_{i1}^4\right) \leq D_\eta$.

The estimators that we consider in this paper are all special cases of a general class of estimators, which can be written in the following generic form:

$$\widehat{\beta}_{\alpha,n} = \left(Y_{2n}' P_{Z_n} Y_{2n} - \widehat{\alpha}_n Y_{2n}' Y_{2n}\right)^{-1} \left(Y_{2n}' P_{Z_n} y_{1n} - \widehat{\alpha}_n Y_{2n}' y_{1n}\right). \tag{3}$$

As discussed in Hansen, Hausman, and Newey (2004), the class of estimators defined by (3) includes all estimators of the well-known k-class (i.e., set $\widehat{\alpha}_n$ =k/(1+k)). In particular, note that: (a) letting $\widehat{\alpha}_n = 0$, we obtain the $2SLS$ estimator, denoted by $\widehat{\beta}_{2SLS,n}$; (b) letting $\widehat{\alpha}_n = \min_{\|\gamma\|=1} \gamma' Y_n' P_{Z_n} Y_n \gamma / \gamma' Y_n' Y_n \gamma$, where $Y_n = [y_{1n}, Y_{2n}]$, we obtain the $LIML$ estimator, denoted by $\widehat{\beta}_{LIML,n}$; and (c) letting $\widehat{\alpha}_n = (K_n - 2)/n$, we obtain Nagar's $B2SLS$ estimator, denoted by $\widehat{\beta}_{B2SLS,n}$.[5] In general, not all members of the k-class of estimators are equally robust to

[3]An example of a parameterization where Π_n is shrunken toward the origin is the local-to-zero framework of Staiger and Stock (1997), which takes $\Pi_n = C/\sqrt{n}$, for some (fixed-dimensional) $K \times G$ parameter matrix C.

[4]More primitive conditions that imply Assumption 1 are given in the extended version of this paper (see Chao and Swanson (2002)).

[5]In an earlier version of this paper, Chao and Swanson (2002) discuss a broader class of estimators (called the ω-class), which includes the Jackknife IV estimator discussed in Phillips and Hale (1977) and Angrist, Imbens and Kreuger (1999) in addition to all estimators of the k-class. In that paper, Chao and Swanson extend the results of this paper to the ω-class.

instrument weakness, as we will explain in greater detail in Remarks 2.2(iii)-(v) below. Estimators which are more robust to instrument weakness turn out to satisfy the following condition:

Assumption 3: Suppose that, as $n \to \infty$, $\widehat{\alpha}_n$ satisfies the following condition: $(n/K_n)\,\widehat{\alpha}_n = 1 + o_p(r_n/K_n)$.

Theorem 2.1: *Let $\widehat{\beta}_{\alpha,n}$ be defined as in equation (3). Suppose that Assumptions 1-3 hold and suppose that $r_n \to \infty$ as $n \to \infty$, such that $\frac{\sqrt{K_n}}{r_n} \to 0$. Then, $\widehat{\beta}_{\alpha,n} \xrightarrow{p} \beta_0$ as $n \to \infty$.*

Remark 2.2: (i) Note that, under Assumption 1(c), r_n can be interpreted as the rate at which the concentration parameter, $\Sigma_{VV}^{-\frac{1}{2}}\Pi_n Z_n' Z_n \Pi_n \Sigma_{VV}^{-\frac{1}{2}}$, grows as n increases. Given that the concentration parameter is a natural measure of instrument strength, we shall in this paper characterize the quality of instruments by the order of magnitude of r_n, so that the slower is the divergence of r_n, the weaker are the instruments. Because our focus is on the case of weak instruments, Assumption 1(c) takes r_n to diverge no faster than n. In fact, we will primarily be interested in cases where r_n diverges much more slowly than n.

(ii) Because we focus only on consistency in this paper, we are able to make assumptions that are weaker in three respects than conditions assumed in other more recent papers on many weak instruments. First, we allow for weaker instruments than Stock and Yogo (2003b) and Hansen, Hausman, and Newey (2004), who also study linear IV setups where instrument weakness is measurable in terms of the order of magnitude of the concentration parameter. More precisely, our results cover the case where instrument weakness is such that the concentration parameter grows at a rate slower than K_n, and possibly much slower than n, provided that $\sqrt{K_n}/r_n \to 0$, as $n \to \infty$, whereas Stock and Yogo (2003b) and Hansen, Hausman, and Newey (2004) consider scenarios where the concentration parameter grows at the same rate or at a faster rate than K_n. Second, we place less restriction on the rate of growth of K_n vis-à-vis n. More specifically, we permit K_n to grow either at the same rate or at a slower rate than n, whereas Stock and Yogo (2003b) and Newey (2004) both require that $K_n^2/n \to 0$; the many weak instruments results of Hansen, Hausman, and Newey (2004) require that $K_n/n \to 0$; and Han and Phillips (2003) require that $K_n/(nc_n) \to \alpha \in (0, \infty)$, for some nonrandom sequence c_n which measures the order of magnitude of the main signal component of their GMM model.[6] Finally, we allow for more general conditions on the stochastic properties of the instruments. In particular, we do not require the instruments, Z_i, to be *i.i.d.*, as in Stock and Yogo (2003b), Hansen, Hausman, and Newey (2004), and Newey (2004); nor do we impose restrictions on the amount of correlation across the components of Z_i (i.e. across the index k in Z_{ik}), as in Han and Phillips (2003). Our results show that, within a linear IV setup, consistency can be achieved under very mild conditions on the instruments and on the growth rate of K_n vis-à-vis n. On the other hand, it should be emphasized that to obtain

[6] Note that, since Han and Phillips (2003) and Newey (2004) study nonlinear GMM frameworks, the asymptotic sequence they consider is, strictly speaking, one which takes the number of moment conditions to infinity, as $n \to \infty$. Hence, when referring to their papers, K_n here should be interpreted more generally as the number of moment conditions.

asymptotic distributional results requires further, more restrictive assumptions, as has been shown by Stock and Yogo (2003b), Chao and Swanson (2004), and Hansen, Hausman, and Newey (2004) in the linear IV setup and by Han and Phillips (2003) and Newey (2004) in the nonlinear GMM setup.

(iii) As shown in the appendix, both $LIML$ and $B2SLS$ satisfy Assumption 3, but $2SLS$ does not. This leads to the following results.

Corollary 2.3: *Suppose that* $r_n \to \infty$ *as* $n \to \infty$, *such that* $\sqrt{K_n}/r_n \to 0$. *Then, under Assumptions 1-2, (a)* $\widehat{\beta}_{LIML,n} \xrightarrow{p} \beta_0$ *as* $n \to \infty$ and (b) $\widehat{\beta}_{B2SLS,n} \xrightarrow{p} \beta_0$ *as* $n \to \infty$.

Theorem 2.4: *Under Assumptions 1-2, as* $n \to \infty$: *(a) for* $r_n/K_n \to 0$, $\widehat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0 + \Sigma_{VV}^{-1}\sigma_{Vu}$; *(b) for* $r_n/K_n \to \delta$ $(0 < \delta < \infty)$, $\widehat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0 + (\delta\Psi + \Sigma_{VV})^{-1}\sigma_{Vu}$; *and (c) for* $K_n/r_n \to 0$, $\widehat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0$.

Note that, unlike $LIML$ and $B2SLS$, $2SLS$ is inconsistent when the concentration parameter grows at the same or slower rate than the number of instruments and is consistent only when the instruments are strong enough so that r_n grows faster than K_n. Interestingly, Theorem 2.4(a) shows that for $r_n/K_n \to 0$, the $2SLS$ estimator (while inconsistent) does not converge to a random limit, in contrast to the case when the number of instruments is held fixed (cf. Staiger and Stock (1997)). Rather, the $2SLS$ converges in probability to a nonrandom limit equaling the sum of β_0 and the OLS bias term, $\Sigma_{VV}^{-1}\sigma_{Vu}$. This result can also be derived using a sequential asymptotic scheme, as shown in Chao and Swanson (2001).

(iv) Further light can be shed upon the importance of the rate condition, $\sqrt{K_n}/r_n \to 0$, as $n \to \infty$, by considering a special case of the model given by (1) and (2), where there is only one endogenous regressor, where the instruments are nonrandom and orthonormal (i.e. $Z_n'Z_n = n \cdot I_n$), and where we have the local-to-zero parameterization $\pi_n = n^{-\zeta}\iota_K$, with $\iota_K = (1, ..., 1)'$ denoting a $K_n \times 1$ vector of ones. (Note that we write π_n instead of Π_n here because in the case of one endogenous regressor, this is a parameter vector instead of a matrix.) In this case: $\pi_n' Z_n' Z_n \pi_n = n^{1-2\zeta} K_n = r_n$. Hence, the condition $\sqrt{K_n}/r_n \to 0$, as $n \to \infty$ is equivalent to the condition $n^{1-2\zeta}\sqrt{K_n} \to \infty$. Interestingly, when instruments are weak in the Staiger and Stock (1997) sense (i.e. when $\zeta = 1/2$), consistency of estimators which satisfy Assumption 3 (A3 estimators, henceforth) requires only that $K_n \to \infty$. On the other hand, for the case $\zeta > 1/2$, which corresponds to what could be termed "weaker than weak" instruments, consistency of A3 estimators requires that K_n must not only diverge but grow with some power of n. In particular, note that the maximum allowed growth rate for K_n is n. In this case, the condition becomes: $n^{1/2}n^{1-2\zeta} \to \infty$, which would require that $3/2 - 2\zeta > 0$ (i.e. $\zeta < 3/4$). Thus, when $K_n \sim n$, consistency of A3 estimators, such as $LIML$ and $B2SLS$, can be attained even if the individual coefficients in π_n decays to zero more rapidly than $n^{-1/2}$, so long as they do not approach zero faster than $n^{-3/4}$.

Note that this example illustrates the potential benefit of using many instruments in situations where each individual instrument is only weakly correlated with the endogenous regressors. As illustrated above, even if each component of π_n is small and can be appropriately modelled as being local-to-zero, the combined effect of using a large number of instruments might nevertheless allow the concentration parameter to grow sufficiently fast, so that consistent estimation can be achieved as $K_n, n \to \infty$, at least with respect to A3 estimators. This asymptotic analysis suggests that using a large number of instruments could improve the precision of point estimation in situations with many weak instruments. Moreover, the choice of estimator becomes perhaps more critical in these situations than in situations where conventional asymptotics applies, since not all estimators are equally robust to instrument weakness. In particular, as shown in Corollary 2.3 and Theorem 2.4 above, the $2SLS$ estimator is less able to withstand instrument weakness than either $LIML$ or $B2SLS$. In the context of our illustrative example, the $2SLS$ is consistent only if $r_n/K_n = n^{1-2\zeta} \to \infty$ as $n \to \infty$.[7] Hence, in contrast to $LIML$ and $B2SLS$, consistency of $2SLS$ requires the more stringent condition that $\zeta < 1/2$ or that the instruments must be less weak than that assumed in the Staiger and Stock (1997) setup.[8,9]

(v) To better understand why the $2SLS$ estimator is less robust to instrument weakness than estimators which satisfy Assumption 3 (such as $LIML$ and $B2SLS$), it is useful to view k-class estimators as IV estimators where the matrix of observations on the instrumental variables is given by $W_{\alpha,n} = [P_{Z_n} - \widehat{\alpha}_n I_n] Y_{2n}$. Now, interpret $W'_{\alpha,n} u_n = W'_{\alpha,n} (y_{1n} - Y_{2n}\beta_0)$ as the first-order condition of the estimator $\widehat{\beta}_{\alpha,n}$, evaluated at the true parameter value β_0; and note that, under conventional asymptotic theory with strongly identified models, consistency of IV estimation involves an asymptotic orthogonality condition of the form $W'_{\alpha,n} u_n/n \xrightarrow{p} 0$, as $n \to \infty$. When the instruments are weak, however, $Y'_{2n} [P_{Z_n} - \widehat{\alpha}_n I_n] Y_{2n}$, the "denominator" of the k-class estimator, will grow at a rate r_n, which may be substantially slower than n. As a result, a stronger (asymptotic) orthogonality condition (OC, henceforth) of the form $W'_{\alpha,n} u_n/r_n \xrightarrow{p} 0$, or $W'_{\alpha,n} u_n/n = o_p(r_n/n)$, is required for consistency. Moreover, an additional condition must be imposed on the choice of $\widehat{\alpha}_n$ (i.e. Assumption 3) in order to ensure that OC is satisfied. To see this, note that for $2SLS$, $\widehat{\alpha}_n = 0$, and this choice of $\widehat{\alpha}_n$ does not satisfy OC. In particular, the $2SLS$ first-order condition, evaluated at β_0 and standardized by r_n, can be written as: $W'_{0,n} u_n/r_n = Y'_{2n} P_{Z_n} u_n/r_n \Rightarrow E\left(W'_{0,n} u_n/r_n\right) = E\left(Y'_{2n} P_{Z_n} u_n/r_n\right) = E\left(\Pi'_n Z'_n u_n/r_n\right) + E\left(V'_n P_{Z_n} u_n/r_n\right) = (K_n/r_n)\,\sigma_{Vu}$, which grows without bound when instrument weakness is such that r_n diverges more slowly than K_n. This nonzero expectation induces a $2SLS$ bias, which is more severe the greater the degree of overidentification and the weaker the available instruments, as measured by a slower rate of growth of the concentration parameter. Hence, as shown in Theorem

[7] Note also that, interestingly, the consistency of the $2SLS$ estimator is not helped at all by the growth of K_n, in contrast with A3 estimators such as $LIML$ and $B2SLS$.

[8] We thank the Co-Editor, Whitney Newey, for providing us with the illustrative example used in this remark and for giving us detailed suggestions on how to better discuss the rate condition $\sqrt{K_n}/r_n \to 0$.

[9] See also Hahn and Kuersteiner (2002) for a related discussion with respect to the $2SLS$ estimator.

2.4 above, when r_n diverges more slowly than K_n, the $2SLS$ estimator is inconsistent and has asymptotic bias equal to that of the OLS estimator. On the other hand, it is shown in the proof of Theorem 2.1 that OC holds for A3 estimators. This is because A3 estimators have built into them an "automatic" bias correction. More specifically, note that the first-order condition for A3 estimators (when evaluated at β_0 and standardized by r_n) can be written as:

$$\frac{W'_{\alpha,n}u_n}{r_n} = \frac{Y'_{2n}\left[P_{Z_n} - \widehat{\alpha}_n I_n\right]u_n}{r_n} = \frac{Y'_{2n}P_{Z_n}u_n}{r_n} - \frac{K_n}{n}\frac{Y'_{2n}u_n}{r_n} + o_p(1). \tag{4}$$

If we ignore the $o_p(1)$ term, then (4) differs from the $2SLS$ first-order condition only by the presence of the term $-(K_n/n)(Y'_{2n}u_n/r_n)$, which can be viewed as a bias-correction term. Indeed, by taking expectations of the first two terms of the right-hand side of (4), we obtain $E(Y'_{2n}P_{Z_n}u_n/r_n) - (K_n/n)E(Y'_{2n}u_n/r_n) = (K_n/r_n)\sigma_{Vu} - (K_n/r_n)\sigma_{Vu} = 0$, so that, unlike $2SLS$, A3 estimators have first-order condition centered at zero, asymptotically. As has been pointed out previously by a number of authors, including Donald and Newey (2000, 2001), estimators whose first-order conditions are well-centered, in the sense of having expectation close to zero, tend to exhibit less small sample bias than estimators whose first-order condition is not well-centered. Our results suggest that, in the context of many weak instruments, it is even more important to use estimators which are well-centered, since estimators which are not properly centered (such as $2SLS$) are asymptotically deficient relative to well-centered estimators (such as $LIML$ and $B2SLS$), in the sense that the latter estimators are consistent in certain weakly-identified settings where the former is not.[10]

(vi) It is also of interest to explicitly compare our setup with that of Bekker (1994). Bekker (1994) considers an asymptotic framework which, in our notations, takes $(n-G)^{-1}\Pi' Z'_n Z_n \Pi$ to be fixed, as both K_n and n go to infinity, such that $K_n/n \to \tau$, for some constant τ with $0 \leq \tau < 1$. Hence, the Bekker framework assumes that the concentration parameter grows at the same rate as n. Our setup, on the other hand, allows for weaker instruments, as measured by the order of magnitude of the concentration parameter, since we allow the concentration parameter to possibly grow at a much slower rate than n. Note that, within the Bekker framework, both $2SLS$ and $LIML$ are consistent if $K_n/n \to 0$ as $n \to \infty$, but $2SLS$ is inconsistent whereas $LIML$ is consistent if $K_n/n \to \tau \neq 0$. More generally, however, we show that consistency depends on the relative magnitude of r_n vis-a-vis K_n, as $n \to \infty$, and not so much on the relative orders of magnitude of n and K_n, unless $r_n = n$ as in Bekker (1994). Thus, for example, if instruments are sufficiently weak so that r_n is of a lower order relative to both n and K_n, then $2SLS$ will be inconsistent, even if $K_n/n \to 0$ (see Theorem 2.4).

[10] Our result is consistent with Monte Carlo results reported in Staiger and Stock (1997), that show that LIML is better centered than 2SLS in finite samples, in the sense of having a smaller median bias. Our results are also consistent with results from the finite sample literature on single equation IV estimators, which show that the exact distribution of LIML under Gaussian error assumptions is better centered than that of 2SLS (see e.g. Phillips (1983), and the papers cited therein).

3 Appendix

In the proofs below, we let C denote a generic constant which may be different in different uses. Before proving the main results of this paper, we first state a lemma which is used in subsequent proofs.

Lemma A1: *Under Assumptions 1-2, suppose that $r_n \to \infty$ as $n \to \infty$ such that $\sqrt{K_n}/r_n \to 0$. Then, the following statements are true, as $n \to \infty$: (a) $V_n' P_{Z_n} u_n / K_n = \sigma_{Vu} + O_p\left(1/\sqrt{K_n}\right)$; (b) $V_n' P_{Z_n} V_n / K_n = \Sigma_{VV} + O_p\left(1/\sqrt{K_n}\right)$; (c) $u_n' P_{Z_n} u_n / K_n = \sigma_{uu} + O_p\left(1/\sqrt{K_n}\right)$; (d) $\Pi_n' Z_n' V_n / r_n = O_p\left(1/\sqrt{r_n}\right)$; (e) $\Pi_n' Z_n' u_n / r_n = O_p\left(1/\sqrt{r_n}\right)$.*

Proof of Lemma A1: To prove part (a), note that it suffices to prove that the $V_{\cdot g,n}' P_{Z_n} u_n / K_n = \sigma_{Vu}^{(g)} + O_p\left(1/\sqrt{K_n}\right)$ as $n \to \infty$, where $V_{\cdot g,n}$ denotes the g^{th} column of V_n, so that $V_{\cdot g,n}' P_{Z_n} u_n / K_n$ is the g^{th} element of $V_n' P_{Z_n} u_n / K_n$, and where $\sigma_{Vu}^{(g)}$ denotes the g^{th} element of σ_{Vu}. In fact, it suffices to show that $E\left(V_{\cdot g,n}' P_{Z_n} u_n / K_n - \sigma_{Vu}^{(g)}\right)^2 = O\left(1/K_n\right)$. To proceed, note that in light of Assumption 1(b), $Z_n' Z_n$ is non-singular with probability 1 for n sufficiently large, so that $\left(Z_n' Z_n\right)^{-1}$ is well-defined and $P_{Z_n} = Z_n \left(Z_n' Z_n\right)^{+} Z_n' = Z_n \left(Z_n' Z_n\right)^{-1} Z_n'$ with probability 1 for n sufficiently large. Now, let $p_{ij,n}$ denote the $(i,j)^{th}$ element of P_{Z_n} and we can write for n sufficiently large:

$$
\begin{aligned}
E\left(\frac{V_{\cdot g,n}' P_{Z_n} u_n}{K_n} - \sigma_{Vu}^{(g)}\right)^2 &= \frac{1}{K_n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E\left(p_{ij,n} p_{kl,n}\right) E\left(v_{ig} u_j v_{kg} u_l\right) - \frac{2\sigma_{Vu}^{(g)}}{K_n} \sum_{i=1}^{n} \sum_{j=1}^{n} E\left(p_{ij,n}\right) E\left(v_{ig} u_j\right) \\
&\quad + \left(\sigma_{Vu}^{(g)}\right)^2 \\
&= \frac{1}{K_n^2} E\left(v_{ig}^2 u_i^2\right) \left[\sum_{i=1}^{n} E\left(p_{ii,n}^2\right)\right] + 2 \frac{1}{K_n^2} \Sigma_{VV}^{(g,g)} \sigma_{uu} \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} E\left(p_{ij,n}^2\right)\right] \\
&\quad + \frac{2\left(\sigma_{Vu}^{(g)}\right)^2}{K_n^2} \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} E\left(p_{ii,n} p_{jj,n} + p_{ij,n}^2\right)\right] - \frac{2\left(\sigma_{Vu}^{(g)}\right)^2}{K_n} \sum_{i=1}^{n} E\left(p_{ii,n}\right) + \left(\sigma_{Vu}^{(g)}\right)^2 \\
&= \frac{1}{K_n^2} E\left(v_{ig}^2 u_i^2\right) \left[\sum_{i=1}^{n} E\left(p_{ii,n}^2\right)\right] + 2 \frac{1}{K_n^2} \Sigma_{VV}^{(g,g)} \sigma_{uu} \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} E\left(p_{ij,n}^2\right)\right] \\
&\quad + \left\{2 \frac{1}{K_n^2} \left(\sigma_{Vu}^{(g)}\right)^2 \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} E\left(p_{ii,n} p_{jj,n} + p_{ij,n}^2\right)\right] - \left(\sigma_{Vu}^{(g)}\right)^2\right\} \\
&= \mathcal{A}_n + \mathcal{B}_n + \mathcal{C}_n, \quad \text{say,}
\end{aligned}
$$

where first equality above follows from Assumption 2(a) which implies that $E\left(p_{ij,n} p_{kl,n} v_{ig} u_j v_{kg} u_l\right) = E\left(p_{ij,n} p_{kl,n}\right) E\left(v_{ig} u_j v_{kg} u_l\right)$ and $E\left(p_{ij,n} v_{ig} u_j\right) = E\left(p_{ij,n}\right) E\left(v_{ig} u_j\right)$, where the second equality follows from noting that, under Assumption 2(b), $E\left(v_{ig} u_j v_{kg} u_l\right)$ equals zero except in the cases where either $(i = j = k = l)$ or $(i = k$ and $j = l)$ or $(i = j$ and $k = l)$ or $(i = l$ and $j = k)$ and from noting that $E\left(v_{ig} u_j\right)$ equals $\sigma_{Vu}^{(g)}$ if $i = j$ and equals zero if $i \neq j$, and where the third equality above follows from the fact that $\sum_{i=1}^{n} E\left(p_{ii,n}\right) = E\left(Tr\left[P_{Z_n}\right]\right) = K_n$ for n

sufficiently large such that $P_{Z_n} = Z_n \left(Z_n' Z_n\right)^{-1} Z_n'$ with probability 1. Focusing on $\mathcal{A}_n$ first, note that:

$$\frac{E\left(v_{ig}^2 u_i^2\right)}{K_n^2}\left[\sum_{i=1}^n E\left(p_{ii,n}^2\right)\right] \leq \frac{\sqrt{E\left(v_{ig}^4\right)}\sqrt{E\left(u_i^4\right)}}{K_n^2}\left[\sum_{i=1}^n E\left(p_{ii,n}^2\right)\right] \leq \frac{\sqrt{E\left(v_{ig}^4\right)}\sqrt{E\left(u_i^4\right)}}{K_n} \leq \frac{C}{K_n} = O\left(\frac{1}{K_n}\right). \tag{5}$$

Note that the first inequality in expression (5) follows from the Cauchy-Schwartz inequality, while the second inequality follows from the fact that $\sum_{i=1}^n E\left(p_{ii,n}^2\right) \leq \sum_{i=1}^n E\left(p_{ii,n}\right) = E\left(Tr\left[P_{Z_n}\right]\right) = K_n$ for n sufficiently large since, by properties of the projection matrix P_{Z_n}, $0 \leq p_{ii,n} \leq 1$ for all i. Finally, the last inequality in (5) follows from Assumption 2(c). Next, turning our attention $\mathcal{B}_n$, we note that:

$$\frac{2\Sigma_{VV}^{(g,g)}\sigma_{uu}}{K_n^2}\left[\sum_{i=2}^n\sum_{j=1}^{i-1} E\left(p_{ij,n}^2\right)\right] \leq \frac{\Sigma_{VV}^{(g,g)}\sigma_{uu}}{K_n^2}\left[\sum_{i=1}^n E\left(p_{ii,n}^2\right) + 2\sum_{i=2}^n\sum_{j=1}^{i-1} E\left(p_{ij,n}^2\right)\right] = \frac{\Sigma_{VV}^{(g,g)}\sigma_{uu}}{K_n} \leq \frac{C}{K_n} = O\left(\frac{1}{K_n}\right), \tag{6}$$

where the first inequality above follows from the nonnegativity of the term $\left(1/K_n^2\right)\Sigma_{VV}^{(g,g)}\sigma_{uu}\sum_{i=1}^n E\left(p_{ii,n}^2\right)$, where the first equality above follows from the fact that $\left[\sum_{i=1}^n E\left(p_{ii,n}^2\right) + 2\sum_{i=2}^n\sum_{j=1}^{i-1} E\left(p_{ij,n}^2\right)\right] = Tr\left[E\left(P_{Z_n}' P_{Z_n}\right)\right] = E\left(Tr\left[P_{Z_n}\right]\right) = K_n$ given that P_{Z_n} is symmetric and idempotent, and where the second inequality follows from Liapunov's inequality and Assumption 2(c). Finally, turning our attention to $\mathcal{C}_n$, we note that:

$$|\mathcal{C}_n| = \left|\frac{\left(\sigma_{Vu}^{(g)}\right)^2}{K_n^2}\left\{K_n - 2\sum_{i=1}^n E\left(p_{ii,n}^2\right)\right\}\right| \leq \frac{\left(\sigma_{Vu}^{(g)}\right)^2}{K_n} + \frac{2\left(\sigma_{Vu}^{(g)}\right)^2 \sum_{i=1}^n E\left(p_{ii,n}\right)}{K_n^2} \leq \frac{3\Sigma_{VV}^{(g,g)}\sigma_{uu}}{K_n} \leq \frac{C}{K_n} = O\left(\frac{1}{K_n}\right), \tag{7}$$

Note that the first equality in (7) above follows from the fact that $|\mathcal{C}_n| = \left|\frac{2\left(\sigma_{Vu}^{(g)}\right)^2}{K_n^2}\left[\sum_{i=2}^n\sum_{j=1}^{i-1} E\left(p_{ii,n}p_{jj,n} + p_{ij,n}^2\right)\right] - \left(\sigma_{Vu}^{(g)}\right)^2\right|$ $= \left|\left(1/K_n^2\right)\left(\sigma_{Vu}^{(g)}\right)^2\left\{\left(Tr\left[E\left(P_{Z_n}\right)\right]\right)^2 + Tr\left[E\left(P_{Z_n}' P_{Z_n}\right)\right] - 2\sum_{i=1}^n E\left(p_{ii,n}^2\right)\right\} - \left(\sigma_{Vu}^{(g)}\right)^2\right| =$ $\left|\left(1/K_n^2\right)\left(\sigma_{Vu}^{(g)}\right)^2\left\{K_n^2 + K_n - 2\sum_{i=1}^n E\left(p_{ii,n}^2\right)\right\} - \left(\sigma_{Vu}^{(g)}\right)^2\right| = \left|\left(1/K_n^2\right)\left(\sigma_{Vu}^{(g)}\right)^2\left\{K_n - 2\sum_{i=1}^n E\left(p_{ii,n}^2\right)\right\}\right|$; the first inequality in (7) follows from the triangle inequality and from the fact that $\sum_{i=1}^n E\left(p_{ii,n}^2\right) \leq \sum_{i=1}^n E\left(p_{ii,n}\right)$ given that $0 \leq p_{ii,n} \leq 1$ for all i; the second inequality above follows from applying the Cauchy-Schwartz inequality and from the fact that $\sum_{i=1}^n E\left(p_{ii,n}\right) = E\left(Tr\left[P_{Z_n}\right]\right) = K_n$ for n sufficiently large, and the third inequality above follows from Liapunov's inequality and Assumption 2(c). Expressions (5)-(7) together imply that $E\left(V_{\cdot g,n}' P_{Z_n} u_n/K_n - \sigma_{Vu}^{(g)}\right)^2 = O\left(1/K_n\right)$, from which it follows that $V_{\cdot g,n}' P_{Z_n} u_n/K_n = \sigma_{Vu}^{(g)} + O_p\left(1/\sqrt{K_n}\right)$, as required.

Parts (b) and (c) follow from proofs similar to that of part (a). Hence, for brevity, we omit these proofs here.

The proofs for parts (d) and (e) are similar, so we will only prove (d). To proceed, note that $E\left[\left\|V_n' Z_n \Pi_n/r_n\right\|^2\right] = E\left(Tr\left[\Pi_n' Z_n' V_n V_n' Z_n \Pi_n/r_n^2\right]\right) = Tr\left[\Sigma_{VV}\right] E\left(Tr\left[\Pi_n' Z_n' Z_n \Pi_n\right]/r_n^2\right) \leq \left(1/r_n\right) C E\left(Tr\left[\Pi_n' Z_n' Z_n \Pi_n/r_n\right]\right) = O\left(1/r_n\right)$, where the second equality follows from Assumption 2(a) and the law of iterated expectations, the lone inequality follows from Assumption 2(c), and where the last equality follows from Assumption 1(c). $\square$

Proof of Theorem 2.1: As usual, write $\widehat{\beta}_{\alpha,n}-\beta_0=\left(\frac{Y_{2n}'P_{Z_n}Y_{2n}}{r_n}-\frac{\widehat{\alpha}_nY_{2n}'Y_{2n}}{r_n}\right)^{-1}\left(\frac{Y_{2n}'P_{Z_n}u_n}{r_n}-\frac{\widehat{\alpha}_nY_{2n}'u_n}{r_n}\right)$. By the condition $\frac{n}{K_n}\widehat{\alpha}_n=1+o_p\left(\frac{r_n}{K_n}\right)$, we deduce that $\left[\widehat{\alpha}_n-(K_n/n)\right](Y_{2n}'Y_{2n}/r_n)=(K_n/r_n)\left[(n/K_n)\widehat{\alpha}_n-1\right](Y_{2n}'Y_{2n}/n)=o_p(1)\,O_p(1)\xrightarrow{p}0$. Similarly, $\left[\widehat{\alpha}_n-(K_n/n)\right]Y_{2n}'u_n/r_n\xrightarrow{p}0$. Therefore, for n sufficiently large such that $Z_n'Z_n$ is nonsingular and $P_{Z_n}=Z_n(Z_n'Z_n)^{+}Z_n'=Z_n(Z_n'Z_n)^{-1}Z_n'$ with probability 1, we have $\frac{Y_{2n}'P_{Z_n}Y_{2n}}{r_n}-\frac{\widehat{\alpha}_nY_{2n}'Y_{2n}}{r_n}=\frac{Y_{2n}'P_{Z_n}Y_{2n}}{r_n}-\left(\frac{K_n}{n}\right)\frac{Y_{2n}'Y_{2n}}{r_n}+o_p(1)=\frac{\Pi_n'Z_n'Z_n\Pi_n}{r_n}+\frac{\Pi_n'Z_n'V_n}{r_n}+\frac{V_n'Z_n\Pi_n}{r_n}+\frac{V_n'P_{Z_n}V_n}{r_n}-\left(\frac{K_n}{nr_n}\right)(\Pi_n'Z_n'Z_n\Pi_n+\Pi_n'Z_n'V_n+V_n'Z_n\Pi_n$ $+V_n'V_n)+o_p(1)=\left(1-\frac{K_n}{n}\right)\frac{\Pi_n'Z_n'Z_n\Pi_n}{r_n}+\left(1-\frac{K_n}{n}\right)\left(\frac{\Pi_n'Z_n'V_n}{r_n}+\frac{V_n'Z_n\Pi_n}{r_n}\right)+\frac{V_n'P_{Z_n}V_n}{r_n}-\left(\frac{K_n}{r_n}\right)\frac{V_n'V_n}{n}+o_p(1)$.
Note that $\Pi_n'Z_n'V_n/r_n=O_p\left(1/\sqrt{r_n}\right)$ in light of Lemma A1 part (d). Also, note that:

$$\frac{V_n'P_{Z_n}V_n}{r_n}=\left(\frac{K_n}{r_n}\right)\frac{V_n'P_{Z_n}V_n}{K_n}=\frac{K_n}{r_n}\Sigma_{VV}+O_p\left(\frac{\sqrt{K_n}}{r_n}\right)=\frac{K_n}{r_n}\Sigma_{VV}+o_p(1), \tag{8}$$

$$\left(\frac{K_n}{r_n}\right)\frac{V_n'V_n}{n}=\left(\frac{K_n}{r_n}\right)\Sigma_{VV}+O_p\left(\frac{K_n}{r_n\sqrt{n}}\right)=\left(\frac{K_n}{r_n}\right)\Sigma_{VV}+O_p\left(\frac{\sqrt{K_n}}{r_n}\sqrt{\frac{K_n}{n}}\right)=\frac{K_n}{r_n}\Sigma_{VV}+o_p(1), \tag{9}$$

where (8) follows from part (b) of Lemma A1, while (9) can be shown by noting that the $(g,h)^{th}$ element of $V_n'V_n/n$ when appropriately standardized, i.e., $\sqrt{n}\left(V_{\cdot g,n}'V_{\cdot h,n}/n-\Sigma_{VV}^{(g,h)}\right)$, satisfies the Lindeberg-Lévy central limit theorem and is, thus, $O_p(1)$. Hence,

$$\frac{Y_{2n}'P_{Z_n}Y_{2n}}{r_n}-\frac{\widehat{\alpha}_nY_{2n}'Y_{2n}}{r_n}=\left(1-\frac{K_n}{n}\right)\frac{\Pi_n'Z_n'Z_n\Pi_n}{r_n}+\frac{K_n}{r_n}\Sigma_{VV}-\frac{K_n}{r_n}\Sigma_{VV}+o_p(1)$$
$$=\left(1-\frac{K_n}{n}\right)\Psi+\left(1-\frac{K_n}{n}\right)\left(\frac{\Pi_n'Z_n'Z_n\Pi_n}{r_n}-\Psi\right)+o_p(1)=\left(1-\frac{K_n}{n}\right)\Psi+o_p(1)\geq C\Psi,$$

for $C>0$, where the last inequality holds in the positive semi-definite sense with probability approaching one in light of Assumption 1 parts (a) and (c). It then follows that $\left(\frac{Y_{2n}'P_{Z_n}Y_{2n}}{r_n}-\frac{\widehat{\alpha}_nY_{2n}'Y_{2n}}{r_n}\right)^{-1}=O_p(1)$. Similarly,

$$\frac{Y_{2n}'P_{Z_n}u_n}{r_n}-\frac{\widehat{\alpha}_nY_{2n}'u_n}{r_n}=\frac{Y_{2n}'P_{Z_n}u_n}{r_n}-\left(\frac{K_n}{n}\right)\frac{Y_{2n}'u_n}{r_n}+o_p(1)$$
$$=\left(1-\frac{K_n}{n}\right)\frac{\Pi_n'Z_n'u_n}{r_n}+\frac{V_n'P_{Z_n}u_n}{r_n}-\left(\frac{K_n}{n}\right)\frac{V_n'u_n}{r_n}+o_p(1)=o_p(1)+\frac{K_n}{r_n}\sigma_{Vu}-\frac{K_n}{r_n}\sigma_{Vu}+o_p(1)\xrightarrow{p}0,$$

where the third equality above follows from Lemma A1 parts (a) and (e) and from the Lindeberg-Lévy central limit theorem. Therefore, $\widehat{\beta}_{\alpha,n}-\beta_0=O_p(1)\,o_p(1)\xrightarrow{p}0$, as required. □

Proof of Corollary 2.3: To show part (a), we need to verify Assumption 3 for $LIML$. Before proceeding, we first set some notations. Write $y_{1n}=Z_n\Pi_n\beta_0+u_n+V_n\beta_0$ and let $Y_n=[y_{1n},Y_{2n}]$, $U_n=[u_n+V_n\beta_0,V_n]$, $\widetilde{Z}_n=Z_n\Pi_n[\beta_0,I_G]$, and $\Phi=[\beta_0,I_G]'\Psi[\beta_0,I_G]$. Note that $Y_n=\widetilde{Z}_n+U_n$ is the reduced form of the SEM given by (1) and (2). Moreover, let $U_{i,n}'$ be the i^{th} row of U_n, and note that the reduced form error covariance matrix $\Omega=E\left[U_{i,n}U_{i,n}'\right]=\Upsilon'\Sigma\Upsilon$, where $\Upsilon=\begin{pmatrix}1 & 0\\ \beta_0 & I_G\end{pmatrix}$, so that Ω is positive definite in light of Assumption 2(b).

Next, note that:

$$
\begin{aligned}
U_n' U_n / n &= \Omega + O_p\left(1/\sqrt{n}\right), & (10)\\
\frac{U_n' P_{Z_n} U_n}{K_n} &= \Upsilon' H_n \Upsilon = \Upsilon' \Sigma \Upsilon + O_p\left(\frac{1}{\sqrt{K_n}}\right) = \Omega + O_p\left(\frac{1}{\sqrt{K_n}}\right), & (11)\\
\widetilde{Z}_n' \widetilde{Z}_n / r_n &= \left[\beta_0, I_G\right]' \left(\Pi_n' Z_n' Z_n \Pi_n / r_n\right) \left[\beta_0, I_G\right] = \Phi + o_{a.s.}(1), & (12)\\
\widetilde{Z}_n' U_n / r_n &= \left[\beta_0, I_G\right]' \left[\Pi_n' Z_n' u_n / r_n, \Pi_n' Z_n' V_n / r_n\right] \Upsilon = O_p\left(1/\sqrt{r_n}\right), & (13)
\end{aligned}
$$

where $H_n = \begin{pmatrix} u_n' P_{Z_n} u_n / K_n & u_n' P_{Z_n} V_n / K_n \\ V_n' P_{Z_n} u_n / K_n & V_n' P_{Z_n} V_n / K_n \end{pmatrix}$, and where (10) follows from Assumption 2, Khintchine's law of large numbers, and the Slutsky Theorem; (11) follows from Lemma A1 (a)-(c); (12) follows from Assumption 1(c); and (13) follows from Lemma A1 (d) and (e).

Now, let $\widehat{A}_n = Y_n' P_{Z_n} Y_n / K_n$, $A_n = \Omega + (r_n/K_n)\Phi$, $\widehat{B}_n = Y_n' Y_n / n$, $B_n = \Omega + (r_n/n)\Phi$, and $\widehat{R}_n(\gamma) = \gamma' \widehat{A}_n \gamma / \gamma' \widehat{B}_n \gamma$. Note that, for $LIML$, $\widehat{\alpha}_n = (K_n/n) \min_{\|\gamma\|=1} \widehat{R}_n(\gamma)$. It follows that to verify Assumption 3, it suffices to show that $\left|\min_{\|\gamma\|=1} \widehat{R}_n(\gamma) - 1\right| = o_p(r_n/K_n)$. To show this, first note that, by the triangle inequality,

$$
\begin{aligned}
\left\|\widehat{A}_n - A_n\right\| &\leq \left\|\frac{U_n' P_{Z_n} U_n}{K_n} - \Omega\right\| + 2\left\|\frac{\widetilde{Z}_n' U_n}{K_n}\right\| + \frac{r_n}{K_n}\left\|\frac{\widetilde{Z}_n' \widetilde{Z}_n}{r_n} - \Phi\right\| \\
&= O_p\left(1/\sqrt{K_n}\right) + O_p\left(\left(r_n/K_n\right)/\sqrt{r_n}\right) + o_{a.s.}\left(r_n/K_n\right) \\
&= O_p\left(\left(r_n/K_n\right)\left(\sqrt{K_n}/r_n\right)\right) + o_p\left(r_n/K_n\right) = o_p\left(r_n/K_n\right),
\end{aligned}
$$

given that $\sqrt{K_n}/r_n \to 0$ and $r_n \to \infty$ as $n \to \infty$. Moreover, since $r_n/n = O(1)$ under Assumption 1(c),

$$
\begin{aligned}
\left\|\widehat{B}_n - B_n\right\| &\leq \left\|\frac{U_n' U_n}{n} - \Omega\right\| + 2\left\|\frac{\widetilde{Z}_n' U_n}{n}\right\| + \frac{r_n}{n}\left\|\frac{\widetilde{Z}_n' \widetilde{Z}_n}{r_n} - \Phi\right\| \\
&= O_p\left(1/\sqrt{n}\right) + O_p\left(\sqrt{r_n}/n\right) + o_{a.s.}\left(r_n/n\right) \\
&= O_p\left(\frac{r_n}{K_n}\left(\frac{\sqrt{K_n}}{r_n}\right)\sqrt{\frac{K_n}{n}}\right) + O_p\left(\frac{r_n}{K_n}\left(\frac{K_n}{n}\right)\frac{1}{\sqrt{r_n}}\right) + o\left(\frac{r_n}{K_n}\frac{K_n}{n}\right) = o_p\left(\frac{r_n}{K_n}\right).
\end{aligned}
$$

In addition, using the Cauchy-Schwartz inequality, we obtain that, for all γ with $\|\gamma\| = 1$,

$$
\left|\gamma' \widehat{B}_n \gamma - \gamma' B_n \gamma\right| = \left|Tr\left[\left(\widehat{B}_n - B_n\right)\gamma\gamma'\right]\right| \leq \sqrt{Tr\left[\left(\widehat{B}_n - B_n\right)\left(\widehat{B}_n - B_n\right)'\right]}\sqrt{Tr\left(\gamma\gamma'\gamma\gamma'\right)} = \left\|\widehat{B}_n - B_n\right\| \xrightarrow{p} 0, \tag{14}
$$

where the last equality follows, given that $\sqrt{Tr\left(\gamma\gamma'\gamma\gamma'\right)} = \sqrt{\left(\gamma'\gamma\right)^2} = \|\gamma\|^2 = 1$. Furthermore, note that B_n is positive definite for all n given that both Ω and Ψ are positive definite, so that with probability one there exists a constant $c > 0$ such that $\left|\gamma' \widehat{B}_n \gamma - \gamma' B_n \gamma\right| + \gamma' \widehat{B}_n \gamma \geq \gamma' B_n \gamma \geq c$ for all n and γ, with $\|\gamma\| = 1$. Hence,

$$
1 = P\left(\left|\gamma' \widehat{B}_n \gamma - \gamma' B_n \gamma\right| + \gamma' \widehat{B}_n \gamma \geq c\right) \leq P\left(\left|\gamma' \widehat{B}_n \gamma - \gamma' B_n \gamma\right| \geq \frac{c}{2}\right) + P\left(\left|\gamma' \widehat{B}_n \gamma\right| \geq \frac{c}{2}\right), \tag{15}
$$

Moreover, (14) implies that, for all ε and γ, with $0 < \varepsilon < c/2$ and $\|\gamma\| = 1$, there exists a positive integer N such that for all $n \geq N$, $P\left(\left|\gamma' \widehat{B}_n \gamma - \gamma' B_n \gamma\right| \geq c/2\right) \leq P\left(\left|\gamma' \widehat{B}_n \gamma - \gamma' B_n \gamma\right| \geq \varepsilon\right) < \varepsilon$. It then follows from (15) that $1-\varepsilon \leq P\left(\left|\gamma'\widehat{B}_n\gamma\right| \geq c/2\right)$ for all $n \geq N$ and γ with $\|\gamma\| = 1$. Since ε can be made arbitrarily small, we deduce that $\left|\gamma'\widehat{B}_n\gamma\right|^{-1} \leq \frac{2}{c} < \infty$ with probability approaching one for all γ with $\|\gamma\| = 1$. Next, define $R_n(\gamma) = \gamma' A_n\gamma/\gamma' B_n\gamma$. Now, $R_n(\gamma)$ is bounded for all γ with $\|\gamma\| = 1$, so that:

$$\begin{aligned}\left|\widehat{R}_n(\gamma) - R_n(\gamma)\right| &\leq \left|\gamma'\widehat{B}_n\gamma\right|^{-1}\left(\left|\gamma'\widehat{A}_n\gamma - \gamma' A_n\gamma\right| + R_n(\gamma)\left|\gamma'\widehat{B}_n\gamma - \gamma' B_n\gamma\right|\right) \\ &\leq C\left\|\widehat{A}_n - A_n\right\| + C\left\|\widehat{B}_n - B_n\right\| = o_p(r_n/K_n).\end{aligned}$$

Let $\alpha_n = K_n/n < 1$ and $\widetilde{\Phi}_n = (r_n/K_n)\Phi$ and note that $A_n = B_n + (1-\alpha_n)\widetilde{\Phi}_n$, so that $R_n(\gamma) = 1 + \frac{(1-\alpha_n)\gamma'\widetilde{\Phi}_n\gamma}{\gamma' B_n\gamma}. \geq 1$. In addition, define $\gamma^* = (1, -\beta_0')' / \|(1, -\beta_0')\|$; we have $\widetilde{\Phi}_n\gamma^* = 0$ for all n, so that $\min_{\|\gamma\|=1} R_n(\gamma) = R_n(\gamma^*) = 1$ for all n. Furthermore, note that, with probability approaching one, $\widehat{\gamma}_n = \arg\min_{\|\gamma\|=1}\widehat{R}_n(\gamma)$ exists. Now, since $\widehat{R}_n(\widehat{\gamma}_n) - R_n(\widehat{\gamma}_n) \leq \widehat{R}_n(\widehat{\gamma}_n) - R_n(\gamma^*) \leq \widehat{R}_n(\gamma^*) - R_n(\gamma^*)$, the desired result follows from

$$\left|\min_{\|\gamma\|=1}\widehat{R}_n(\gamma) - 1\right| = \left|\widehat{R}_n(\widehat{\gamma}_n) - 1\right| = \left|\widehat{R}_n(\widehat{\gamma}_n) - R_n(\gamma^*)\right| \leq \max_{\|\gamma\|=1}\left|\widehat{R}_n(\gamma) - R_n(\gamma)\right| = o_p(r_n/K_n).$$

To show part (b), we simply verify that $B2SLS$ satisfies Assumption 3. To proceed, note that for $B2SLS$, $\widehat{\alpha}_n = (K_n - 2)/n$, so that $(n/K_n)\widehat{\alpha}_n - 1 = -2/K_n = o_p(r_n/K_n)$, since $r_n \to \infty$ as $n \to \infty$ by assumption. The desired conclusion, thus, follows. □

Proof of Theorem 2.4: To show part (a), note first that, in light of Assumption 1(b) , $Z_n'Z_n$ is nonsingular and, thus, $P_{Z_n} = Z_n(Z_n'Z_n)^+ Z_n' = Z_n(Z_n'Z_n)^{-1} Z_n'$ with probability 1 for n sufficiently large. Hence, we can write $Y_{2n}'P_{Z_n}Y_{2n}/K_n = V_n'P_{Z_n}V_n/K_n + (r_n/K_n)[\Pi_n'Z_n'V_n/r_n + V_n'Z_n\Pi_n/r_n + \Pi_n'Z_n'Z_n\Pi_n/r_n]$. Now, since it is assumed in part (a) that $\frac{r_n}{K_n} \to 0$ as $n \to \infty$, it follows, from Assumption 1(c) and parts (b) and (d) of Lemma A1, that $\frac{Y_{2n}'P_{Z_n}Y_{2n}}{K_n} \xrightarrow{p} \Sigma_{VV}$, where $\Sigma_{VV} > 0$ by Assumption 2(b) and is, thus, nonsingular. Moreover, for n sufficiently large, we can write $\frac{Y_{2n}'P_{Z_n}u_n}{K_n} = \left(\frac{r_n}{K_n}\right)\frac{\Pi_n'Z_n'u_n}{r_n} + \frac{V_n'P_{Z_n}u_n}{K_n}$, so that $\frac{Y_{2n}'P_{Z_n}u_n}{K_n} \xrightarrow{p} \sigma_{Vu}$, as $n \to \infty$, by parts (a) and (e) of Lemma A1. It follows immediately for the Slutsky Theorem that $\widehat{\beta}_{2SLS,n} = \beta_0 + [Y_{2n}'P_{Z_n}Y_{2n}/K_n]^{-1}[Y_{2n}'P_{Z_n}u_n/K_n] \xrightarrow{p} \beta_0 + \Sigma_{VV}^{-1}\sigma_{Vu}$ as required.

To show part (b) note that since in this case $\frac{r_n}{K_n} \to \delta$, for some $\delta \in (0,\infty)$, as $n \to \infty$, it follows directly from Assumption 1(c) and parts (b) and (d) of Lemma A1 that $\frac{Y_{2n}'P_{Z_n}Y_{2n}}{K_n} \xrightarrow{p} \delta\Psi + \Sigma_{VV}$, where $(\delta\Psi + \Sigma_{VV}) > 0$ and, thus, nonsingular in light of Assumptions 1(c) and 2(b). In addition, from parts (a) and (e) of Lemma A1, we deduce that $\frac{Y_{2n}'P_{Z_n}u_n}{K_n} \xrightarrow{p} \sigma_{Vu}$. The desired result, thus, follows again from the Slutsky Theorem.

To prove part (c), write $\widehat{\beta}_{2SLS,n} - \beta_0 = \left(\frac{Y_{2n}'P_{Z_n}Y_{2n}}{r_n}\right)^{-1}\left(\frac{Y_{2n}'P_{Z_n}u_n}{r_n}\right)$. Note that $\frac{Y_{2n}'P_{Z_n}Y_{2n}}{r_n} = \frac{\Pi_n'Z_n'Z_n\Pi_n}{r_n} + \frac{V_n'Z_n\Pi_n}{r_n} + \frac{\Pi_n'Z_n'V_n}{r_n} + \left(\frac{K_n}{r_n}\right)\frac{V_n'P_{Z_n}V_n}{K_n} \xrightarrow{p} \Psi > 0$, given Assumption 1(c), parts (b) and (d) of Lemma A1, and the assumption that $K_n/r_n \to 0$ as $n \to \infty$. Moreover, $\frac{Y_{2n}'P_{Z_n}u_n}{r_n} = \frac{\Pi_n'Z_n'u_n}{r_n} + \left(\frac{K_n}{r_n}\right)\frac{V_n'P_{Z_n}u_n}{K_n} \xrightarrow{p} 0$, given Lemma A1 parts (a) and (e) and given the assumption that $K_n/r_n \to 0$ as $n \to \infty$. Weak consistency of $\widehat{\beta}_{2SLS,n}$ then follows as a consequence of the Slutsky Theorem.□

4 References

Angrist, J. D. and A. B. Krueger (1995). "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225-235.

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

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

Bekker, P.A. and F. Kleibergen (2003). "Finite Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic," *Econometric Theory*, 19, 744-753.

Bound, J., D.A. Jaeger and R. Baker (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.

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

Chao, J.C. and N.R. Swanson (2001). "Bias and MSE Analysis of the IV Estimator Under Weak Identification with Application to Bias Correction," Working Paper, Rutgers University.

Chao, J.C. and N.R. Swanson (2002). "Consistent Estimation When the Number of Instruments Is Large," Working Paper, Rutgers University.

Chao, J.C. and N.R. Swanson (2004). "Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments," in D. Corbae, S. Durlauf, and B. Hansen eds. *Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, Cambridge University Press, forthcoming.

Donald, S. G. and W. K. Newey (2000). "A Jackknife Interpretation of the Continuous Updating Estimator," *Economics Letters*, 67, 239-243.

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

Guggenberger, P. and R.J. Smith (2003). "Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification," *Econometric Theory,* forthcoming.

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

Hahn, J., J. Hausman, and G. Kuersteiner (2001). "Accuracy of Higher Order Bias and MSE Approximations, *Econometrics Journal*, 7, 272-306.

Hahn, J. and A. Inoue (2002). "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators," *Econometric Reviews*, 21, 309-336.

Hahn, J. and G. Kuersteiner (2002). "Discontinuities of Weak Instrument Limiting Distributions," *Economic Letters*, 75, 325-331.

Hahn, J. and W. Newey (2004). "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295-1319.

Han, C. and P. C. B. Phillips (2003). "GMM with Many Moment Conditions," Working Paper, Yale University.

Hansen, C., J. Hausman, and W.K. Newey (2004). "Many Instruments, Weak Instruments, and Microeconometric Practice," Working Paper, M.I.T.

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

Moreira, M. (2003). "A Conditional Likelihood Ratio Test for Structural Model," *Econometrica,* 71, 1027-1048

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.

Newey, W.K. (2004). "Many Weak Moment Asymptotics for the Continuously Updated GMM Estimator," Working Paper, M.I.T.

Phillips, G.D.A. and C. Hale (1977). "The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems," *International Economic Review*, 18, 219-228.

Phillips, P.C.B. (1983). "Small Sample Distribution Theory in Econometric Models of Simultaneous Equations," in Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Vol I*, North Holland, Amsterdam.

Rothenberg, T. (1984). "Approximating the Distributions of Econometric Estimators and Test Statistics," in Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Vol II*, North Holland, Amsterdam.

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

Stock, J.H. and M. Yogo (2003a). "Testing for Weak Instruments in Linear IV Regression," in D.W.K. Andrews and J.H. Stock eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge: Cambridge University Press, forthcoming.

Stock, J.H. and M. Yogo (2003b). "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," in D.W.K. Andrews and J.H. Stock eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge: Cambridge University Press, forthcoming.

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

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