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: June 2004

Abstract

This paper analyzes the conditions under which consistent estimation can be achieved in instrumental variables (IV) regression when the available instruments are weak, in the local-to-zero sense of Staiger and Stock (1997) and using the many-instrument framework of Morimune (1983) and Bekker (1994). Our analysis of an extended k-class of estimators that includes Jackknife IV ($JIVE$) establishes that consistent estimation depends importantly on the relative magnitudes of r_n, the growth rate of the concentration parameter, and K_n, the number of instruments. In particular, $LIML$ and $JIVE$ are consistent when $\frac{\sqrt{K_n}}{r_n} \to 0$, while two-stage least squares is consistent only if $\frac{K_n}{r_n} \to 0$, as $n \to \infty$. We argue that the use of many instruments may be beneficial for estimation, as the resulting concentration parameter growth may allow consistent estimation, in certain cases.

JEL classification: C13, C31.
Keywords: instrumental variables, k-class estimator, local-to-zero framework, pathwise asymptotics, 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, Whitney Newey, Peter Phillips, Tao Zha, Eric Zivot, and participants at the 2001 summer econometric society meeting, and workshops at North Carolina State University and Yale University for useful comments and suggestions. Swanson gratefully acknowledges financial support in the form of a Rutgers University Research Council grant.

1 Introduction

In the weak instruments literature, it has become standard in recent years to analyze the properties of estimators and test statistics using the local-to-zero framework pioneered by Staiger and Stock (1997), which takes the coefficients of the instruments in the first-stage regression to be in a $n^{-\frac{1}{2}}$ shrinking neighborhood of zero, where n is the sample size. An interesting feature of the Staiger-Stock framework is that unlike the conventional asymptotic setup, the concentration parameter does not diverge but rather, roughly speaking, stays constant in expectation as the sample size grows. Since the concentration parameter is a natural measure of the strength of identification in an IV regression model, the local-to-zero device allows the Staiger-Stock framework to better mimic the weak instrument situation than the conventional setup with fixed coefficients, and Staiger and Stock show that the two-stage least squares ($2SLS$) and the limited information maximum likelihood ($LIML$) estimators are no longer consistent and instead converge to nonstandard distributions in the limit under this framework.[1,2]

Another important direction that IV regression research 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 to infinity as a function of the sample size. This approach was first taken by Morimune (1983) and later generalized by Bekker (1994) (see also Angrist and Krueger (1995), Bekker and van der Ploeg (1999), van Hasselt (2000), Donald and Newey (2001), Hahn, Hausman, and Kuersteiner (2001), Hahn (2002), and Hahn and Inoue (2002)). In contrast to the papers using the local-to-zero setup, authors taking the many instruments approach typically assume that the concentration parameter grows at the same rate as the sample size. Hence, these papers tend to study scenarios where the instruments are not as weak as those assumed in papers employing a local-to-zero setup.

The purpose of this paper is to provide a unified framework under which the asymptotic behavior of different estimators can be studied in the presence of weak and/or many instruments. More precisely, the setup adopted here combines key features of both the local-to-zero and the many instruments asymptotic frameworks. This combined framework, in turn, allows us to analyze the consistency of various single-equation estimators under a single coherent set of conditions. In particular, we show that when the number of instruments is taken to infinity, 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 has previously been assumed in the literature. Our results, thus, complement those of Stock

[1] The observation that the concentration parameter is a natural measure of the strength of instruments has been made by Phillips (1983), Rothenberg (1983), and Stock and Yogo (2003a), amongst others.

[2] Related work by Sargan (1988), Phillips (1989) and Choi and Phillips (1992) addresses the implications for statistical inference when the underlying simultaneous equations model is underidentified or is only partially identified.

and Yogo (2003b), who derive the limiting distributions of various k-class estimators when the concentration parameter grows at the same rate as the number of instruments. Although we do not provide distributional results in this paper, our setup is more general than that of Stock and Yogo (2003b) as we assume more general conditions on the stochastic properties of the instruments and we allow for weaker instruments, as measured by the order of magnitude of the concentration parameter. In addition, we examine a broad class of IV estimators which extends the well-known k-class by allowing the value of k to vary across observations. We refer to this class of estimators as the ω-class, and we show that members of this class which satisfy certain general conditions are consistent even if r_n, the rate of growth of the concentration parameter, grows slower than K_n, the number of instruments, and possibly much slower than the sample size n, provided that $\frac{\sqrt{K_n}}{r_n} \to 0$, as $n \to \infty$. Specializing our results to specific estimators, we show that $LIML$ and $JIVE$ both satisfy our conditions for consistency, whereas the $2SLS$ estimator does not. Indeed, the $2SLS$ estimator is shown to be consistent only if $\frac{K_n}{r_n} \to 0$, as $n \to \infty$. Our analysis, thus, also provides a precise characterization of the sense in which the $2SLS$ estimator is asymptotically deficient relative to $LIML$ and $JIVE$, in the case of weak identification.[3]

The remainder of the paper describes our model and assumptions as well as presents and discusses our main results. All proofs are gathered in an appendix. In the sequel, $Tr(\cdot)$ denotes the trace of a matrix, A^+ denotes the Moore-Penrose inverse of a (possibly singular) matrix, " > 0" denotes positive definiteness when applied to matrices, $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote, respectively, the maximal eigenvalue and the minimal eigenvalue of the real, symmetric matrix A, $\underline{\lim}_{n\to\infty} a_n$ denotes the limit inferior of the sequence $\{a_n\}$, and $\overline{\lim}_{n\to\infty} a_n$ denotes the limit superior of the sequence $\{a_n\}$. In addition, $P_X = X(X'X)^{-1}X'$ denotes the matrix which projects orthogonally onto the range space of X and $Q_X = I - P_X$.

2 Model, Assumptions, and Main Results

Consider the simultaneous equations model (SEM)

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

$$Y_{2n} = Z_n\Pi + X_n\Phi + V_n, \tag{2}$$

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, X_n is an $n \times J$ matrix of observations on the J exogenous variables included in the

[3]A very interesting recent paper on the subject discussed herein, which was written subsequent to this one, is that of Han and Phillips (2003). One of the main differences between our paper and theirs is the following. Our paper focuses explicitly on a linear IV regression setup, while Han and Phillips (2003) consider a (possibly) nonlinear GMM framework. On the other hand, our paper studies estimators such as $LIML$ and $JIVE$, which lie outside of the class of GMM estimators considered by Han and Phillips (2003), and we show that these estimators have some desirable properties under weak identification.

structural equation (1), Z_n is an $n \times K_n$ matrix of observations on the K_n instrumental variables, or exogenous variables excluded from the structural equation (1), and u_n, V_n are, respectively, an $n \times 1$ vector and an $n \times G$ matrix of random disturbances. Also, define $\overline{Z}_n = (Z_n, X_n)$. Furthermore, let $\eta_i = (u_i, v_i')'$ where u_i and v_i' are, respectively, the $i\underline{th}$ component of the random vector u_n and the $i\underline{th}$ row of the random matrix V_n. The following assumptions are used in the sequel.

Assumption 1: $\Pi = \Pi_n = \frac{C_n}{b_n}$ for some sequence of positive real numbers $\{b_n\}$, non-decreasing in n, and for some sequence of nonrandom, $K_n \times G$ parameter matrices $\{C_n\}$.

Assumption 2: Let $\{\overline{Z}_{n,i} : i = 1, ..., n;\ n \geq 1\}$ be a triangular array of R^{K_n+J}-valued random variables, where $\overline{Z}_{n,i} = (Z_{n,i}',\ X_i')'$ with $Z_{n,i}'$ and X_i' denoting the $i\underline{th}$ row of the matrices Z_n and X_n, respectively. Moreover, suppose that: (a) $K_n \to \infty$ as $n \to \infty$ such that $\frac{K_n}{n} \to \alpha$ for some constant α satisfying $0 \leq \alpha < 1$; (b) there exists a positive integer N such that $\forall n \geq N$, $\overline{Z}_n$ is of full column rank $K_n + J$ almost surely; and (c) $\{r_n\}$ is a non-decreasing sequence of positive real numbers such that, as $n \to \infty$, $\frac{r_n}{n} \to \kappa$ for some constant κ, with $0 \leq \kappa < \infty$, and $\Psi_n = \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2 r_n}$, where $Q_{X_n} = I_n - X_n(X_n' X_n)^{-1} X_n'$, and there exist constants $\underline{D}$ and $\overline{D}$, with $0 < \underline{D} \leq \overline{D} < \infty$, such that $\underline{D} \leq \underline{\lim}_{n\to\infty} \lambda_{\min}(\Psi_n)$ and $\overline{\lim}_{n\to\infty} \lambda_{\max}(\Psi_n) \leq \overline{D}$ almost surely.[4]

Assumption 3: Assume that: (a) $\overline{Z}_n$ 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 σ_{Vu}^g and 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 $\max\{E(u_i^4), E(v_{i1}^4), ..., E(v_{iG}^4)\} \leq D_\eta$.

The estimators we consider can be written in the form:

$$\widehat{\beta}_{\omega,n} = \left(Y_{2n}' \left[I_n - Q_{\overline{Z}_n}\Omega_n\right] Q_{X_n} Y_{2n}\right)^+ \left(Y_{2n}' \left[I_n - Q_{\overline{Z}_n}\Omega_n\right] Q_{X_n} y_{1n}\right), \tag{3}$$

where $\Omega_n = diag(\omega_{1,n},\ \omega_{2,n}, ..., \omega_{n,n})$. It is easily seen that this class of estimators is broader than the well-known k-class in the sense that each k-class estimator can be obtained from the formula above by setting $\omega_{1,n} = \omega_{2,n} = ... = \omega = k$ for a particular k. We shall refer to the class of estimators given by expression (3) above as the $\omega-$class. One reason for defining this broader class of estimators is that an interesting estimator which is not a member of the k-class is $JIVE$; but $JIVE$ is a member of the $\omega-$class. It is often convenient to rewrite (3) as:

$$\widehat{\beta}_{\omega,n} = \left(Y_{2n}' \left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\widetilde{\Omega}_n Q_{X_n}\right] Y_{2n}\right)^+ \left(Y_{2n}' \left[P_{\overline{Z}_n} - P_{X_n} - Q_{\overline{Z}_n}\widetilde{\Omega}_n Q_{X_n}\right] y_{1n}\right), \tag{4}$$

[4]More primitive conditions that imply Assumption 2 are given in the extended version of this paper, Chao and Swanson (2002).

where $\widetilde{\Omega}_n = \Omega_n - I_n$, $\widetilde{\Omega}_n = diag\left(\widetilde{\omega}_{1,n},\ \widetilde{\omega}_{2,n}, ..., \widetilde{\omega}_{n,n}\right)$, and $\widetilde{\omega}_{i,n} = \omega_{i,n} - 1$, for $i = 1, ..., n$. Note that, without further restrictions on $\omega_{i,n}$ $(i = 1, ..., n)$, (or $\widetilde{\omega}_{i,n}$ $(i = 1, ..., n)$), $\widehat{\beta}_{\omega,n}$ is not a consistent estimator for β. Consistent estimation can be obtained, however, given the following restriction:

Assumption 4: Suppose that for each i and n, $\widetilde{\omega}_{i,n}$ can be decomposed into the sum of two components as $\widetilde{\omega}_{i,n} = \overline{\omega}_{i,n} + \xi_{i,n}$, such that $\overline{\omega}_{i,n}$ is either non-random or depends only on the exogenous variables $\overline{Z}_n$, so that $\overline{\omega}_{i,n} = f_{n,i}(\overline{Z}_n)$. Also, assume that $\overline{\omega}_{i,n}$ and $\xi_{i,n}$ satisfy the following conditions: (a) $\overline{\lim_{n\to\infty}}\, \bar{l}_n < \infty$ *a.s.*, where $\bar{l}_n = \sup_{1\leq\ i\ \leq n} |\overline{\omega}_{i,n}|$; (b) $\sum_{i=1}^{n} \overline{\omega}_{i,n} (1 - h_{i,n}) = K_n$ *a.s.* $\forall n$, where $h_{i,n}$ is the ith diagonal element of $P_{\overline{Z}_n}$; (c) $\sum_{i=1}^{n} E\left(\overline{\omega}_{i,n}^2\right) = O(K_n)$; and (d) $\sup_{1\leq\ i\ \leq n} |\xi_{i,n}| = o_p\left(\frac{r_n}{n}\right)$.

Theorem 2.1: *Under Assumptions 1-4, let* $\widehat{\beta}_{\omega,n}$ *be defined as in equation (4) above. Suppose that* $r_n \to \infty$ *as* $n \to \infty$, *such that* $\frac{\sqrt{K_n}}{r_n} \to 0$. *Then,* $\widehat{\beta}_{\omega,n} \xrightarrow{p} \beta_0$ *as* $n \to \infty$.

Remark 2.2: (i) Note that under Assumption 2(c), r_n can be interpreted as the rate at which the concentration parameter $\Sigma_{VV}^{-\frac{1}{2}}\Pi_n Z_n' Q_{X_n} Z_n \Pi_n \Sigma_{VV}^{-\frac{1}{2}}$ grows as n increases. An assumption on the rate of growth of the concentration parameter seems natural here since the concentration parameter is a measure of instrumental strength. Because we are interested in the case of weak instruments, Assumption 2(c) stipulates that r_n must grow no faster than n. In fact, we will be interested primarily in the case where r_n grows much slower than n. In addition, Assumption 3 requires the instrument matrix $\overline{Z}_n$ to be independent of the disturbance vector η_i for all i and n, and also requires the disturbances to have finite absolute fourth moments. Note that these assumptions are weaker than the corresponding assumptions in Morimune (1983) and Bekker (1994), where fixed instruments and *i.i.d.* Gaussian errors are assumed.

(ii) To see the relationship between our framework and that of Staiger and Stock (1997), note that the Staiger-Stock setup takes $b_n = \sqrt{n}$, $Z_n' Q_{X_n} Z_n = O_p(n)$, and the number of instruments to be fixed as $n \to \infty$, so that C is a (fixed-dimensional) $K \times G$ matrix such that $C'C = O(1)$. It is easily shown in their case that $r_n = O(1)$, so that the concentration parameter does not diverge but is bounded in probability. Numerical results reported in Staiger and Stock (1997) show that asymptotic distributions of estimators and test statistics derived under this setup give very good approximations to their finite sample distributions, particularly if the number of instruments is not large relative to the sample size. Our paper here builds upon their work by looking at the case where one uses a large number of weak instruments, so that K_n is allowed to approach infinity as a function of n. In this case, it turns out that the concentration parameter may grow even when the coefficient matrix Π in the first-stage equation is modeled as being small in the local-to-zero sense because, while each individual instruments may be weakly correlated with the endogenous regressors, the combined effect of using a lot of instruments may nevertheless lead to a very large concentration parameter.[5]

(iii) It is also of interest to compare our setup with that of Bekker (1994).[6] Focusing on the case $J = 0$, the alternative asymptotics considered by Bekker (1994), in our notations, boils down to one where the quantity $(n-G)^{-1}\Pi' Z_n' Z_n \Pi$ is kept fixed, as both K_n and n go to infinity, such that $\frac{K_n}{n} \to \alpha$, for some constant α satisfying $0 \leq \alpha < 1$. It follows that the Bekker approach assumes that the concentration parameter grows at the rate of the sample size n. Our framework, thus, allows for weaker instruments, as measured by the order of magnitude of the concentration parameter, than that of Bekker, since we allow r_n to grow at possibly a much slower rate than n.

Our setup is also closely related to others in the literature. For example, in an important paper, Donald and Newey (2001) use a setup which is similar to ours, although we do not require the exogenous regressors $\overline{Z}_i$ to be $i.i.d$, and we allow K_n to diverge at the same rate as n.

(iv) Many commonly-used IV estimators are members of the $\omega-$class, including:

a. Limited Information Maximum Likelihood (LIML) Estimator:

$$\widehat{\beta}_{LIML,n} = \left(Y_{2n}' Q_{X_n} Y_{2n} - \widehat{\lambda}_{LIML,n} Y_{2n}' Q_{\overline{Z}_n} Y_{2n}\right)^+ \left(Y_{2n}' Q_{X_n} y_{1n} - \widehat{\lambda}_{LIML,n} Y_{2n}' Q_{\overline{Z}_n} y_{1n}\right), \tag{5}$$

where $\widehat{\lambda}_{LIML,n}$ is the smallest root of the determinantal equation:

$$\det\left\{\begin{pmatrix} y_{1n}' Q_{X_n} y_{1n} & y_{1n}' Q_{X_n} Y_{2n} \\ Y_{2n}' Q_{X_n} y_{1n} & Y_{2n}' Q_{X_n} Y_{2n} \end{pmatrix} - \lambda_n \begin{pmatrix} y_{1n}' Q_{\overline{Z}_n} y_{1n} & y_{1n}' Q_{\overline{Z}_n} Y_{2n} \\ Y_{2n}' Q_{\overline{Z}_n} y_{1n} & Y_{2n}' Q_{\overline{Z}_n} Y_{2n} \end{pmatrix}\right\} = 0 \tag{6}$$

b. Two-Stage Least Squares (2SLS) Estimator:

$$\widehat{\beta}_{2SLS,n} = \left(Y_{2n}'(P_{\overline{Z}_n} - P_{X_n}) Y_{2n}\right)^+ \left(Y_{2n}'(P_{\overline{Z}_n} - P_{X_n}) y_{1n}\right). \tag{7}$$

c. Jackknife Instrumental Variables Estimator (JIVE):

$$\widehat{\beta}_{JIVE,n} = \left(Y_{2n}'\left[I_n - Q_{\overline{Z}_n} H_n\right] Q_{X_n} Y_{2n}\right)^+ \left(Y_{2n}'\left[I_n - Q_{\overline{Z}_n} H_n\right] Q_{X_n} y_{1n}\right), \tag{8}$$

where $H_n = diag\left(\frac{1}{1-h_{1,n}}, ..., \frac{1}{1-h_{n,n}}\right)$, with $h_{i,n}$ being the ith diagonal element of $P_{\overline{Z}_n}$.

With respect to the last estimator, $JIVE$, it should be noted this estimator was first proposed by Phillips and Hale (1977) but was further studied and given its jackknife interpretation by Angrist, Imbens, and Krueger (1999)[7]. In addition, we note that for $JIVE$ to be well-defined, we need the following assumption:

[5] Technically, this can occur because, in our setup, K_n, the number of rows of the sequence of matices C_n, is now allowed to go to infinity as $n \to \infty$. Hence, the concentration parameter $\Sigma_{VV}^{-\frac{1}{2}} \Pi_n' Z_n' Q_{X_n} Z_n \Pi_n \Sigma_{VV}^{-\frac{1}{2}} = \frac{\Sigma_{VV}^{-\frac{1}{2}} C_n' Z_n' Q_{X_n} Z_n C_n \Sigma_{VV}^{-\frac{1}{2}}}{n}$ may diverge even if $\frac{Z_n' Q_{X_n} Z_n}{n} = O_p(1)$.

[6] The type of asymptotic approximation used by Bekker (1994) dates back to the work of Anderson (1976), Kunitomo (1980), and Morimune (1983), as is pointed out by Bekker in his paper.

[7] Still another paper that examined $JIVE$ is Blomquist and Dahlberg (1999), which provided a Monte Carlo study comparing $JIVE$ with a number of other IV estimators.

Assumption J: There exists a constant $\overline{h}$, with $0 < \overline{h} < 1$, such that $0 \leq h_{i,n} \leq \overline{h}$ $a.s.$ for $1 \leq i \leq n$ and for all n sufficiently large such that $P_{\overline{Z}_n}$ is well-defined almost surely[8],[9].

As shown in the appendix, both $LIML$ and $JIVE$ satisfy Assumption 4, and hence Theorem 2.1 holds. However, this is not the case for the $2SLS$ estimator.[10]

Corollary 2.3: *Suppose that* $r_n \to \infty$ *as* $n \to \infty$*, such that* $\frac{\sqrt{K_n}}{r_n} \to 0$*. Then, (a), under Assumptions 1-3,* $\widehat{\beta}_{LIML,n} \xrightarrow{p} \beta_0$ *as* $n \to \infty$ and (b)*, under Assumptions 1-3 and J,* $\widehat{\beta}_{JIVE,n} \xrightarrow{p} \beta_0$ *as* $n \to \infty$.

Theorem 2.4: *Under Assumptions 1-3, let* $\widehat{\beta}_{2SLS,n}$ *be defined as in equation (7) above. As* $n \to \infty$ *: (a) for* $\frac{r_n}{K_n} \to 0$, $\widehat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0 + \Sigma_{VV}^{-1}\sigma_{Vu}$*; (b) for* $\frac{r_n}{K_n} \to \delta$ $(0 < \delta < \infty)$, $\widehat{\beta}_{2SLS,n} - \beta_0 = (\delta\Psi_n + \Sigma_{VV})^{-1}\sigma_{Vu} + o_p(1)$*, where* $\Psi_n = \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2 r_n}$*; and (c) for* $\frac{K_n}{r_n} \to 0$, $\widehat{\beta}_{2SLS,n} \xrightarrow{p} \beta_0$.

Thus, in contrast to $LIML$ and $JIVE$, $2SLS$ 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.[11] Interestingly, Theorem 2.4(a) shows that for $\frac{r_n}{K_n} \to 0$, 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 (i.e. see 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 is consistent with the result given in Chao and Swanson (2001) based on sequential asymptotics, where it is shown that the variance of the $2SLS$ estimator tends to zero as the number of instruments goes to infinity, but a non-zero bias remains.

(v) To better understand the discrepancy between asymptotic behavior of the $2SLS$ estimator vis-à-vis those estimators which satisfy Assumption 4, it is helpful to focus discussion on the special case where $J = 0$ and $G = 1$ (i.e., the case where there are no included exogenous regressors and only one endogenous regressor in the structural equation). As mentioned above, an $\omega-$class estimator can be viewed as an IV estimator where the

[8]Note that Assumption J does rule out exogenous regressors of the form $e_i = (0, ..., 0, 1, 0, ..., 0)$, where e_i denotes the i^{th} column of I_n, but it does not rule out dummy variable regressors in general.

[9]Note also that Assumption 2(b) implies that, for n sufficiently large, the matrix $\overline{Z}_n'\overline{Z}_n$ is positive definite almost surely, so that $\left(\overline{Z}_n'\overline{Z}_n\right)^{-1}$ exists. It is in this sense that we say that the projection matrix $P_{\overline{Z}_n} = \overline{Z}_n\left(\overline{Z}_n'\overline{Z}_n\right)^{-1}\overline{Z}_n'$ is well-defined almost surely for n sufficiently large.

[10]There are many other recent important papers that examine the JIVE estimator, and which are not discussed here. For example, the reader is referred to Hahn and Newey (2003), and the references cited therein.

[11]Theorem 2.4 above is consistent with the results of Hahn and Kuersteiner (2002), who examine the asymptotic properties of the $2SLS$ estimator in the case where the number of instruments is fixed, $b_n = n^{-\delta}$,and (in our notation) $Z_n'Z_n = O_p(n)$. In their case, $r_n = n^{1-2\delta}$; and hence, as long as $\delta < 1/2$, the concentration parameter will grow and the $2SLS$ estimator will be weakly consistent. Our results extend theirs and show that, more generally, it is the relative magnitudes of r_n and K_n which determine whether the $2SLS$ estimator is consistent, and the rate at which the first-stage coefficient is allowed to shrink toward zero is important only in so much that it affects r_n.

vector of observations on the instrumental variable is given by $w_n\left(\widetilde{\Omega}_n\right) = \left[P_{Z_n} - \widetilde{\Omega}_n Q_{Z_n}\right] y_{2n}$, in the case where $J = 0$ and $G = 1$. Now, under conventional asymptotic theory with strongly identified models, consistency of IV estimation involves an asymptotic orthogonality condition of the form $\frac{1}{n} w_n\left(\widetilde{\Omega}_n\right)' u_n \xrightarrow{p} 0$ as $n \to \infty$. When the instruments are weak, however, $y'_{2n}\left[P_{Z_n} - Q_{Z_n}\widetilde{\Omega}_n\right] y_{2n}$, the "denominator" of the $\omega-$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 $\frac{1}{r_n} w_n\left(\widetilde{\Omega}_n\right)' u_n = \left(\frac{n}{r_n}\right)\left[\frac{1}{n} w_n\left(\widetilde{\Omega}_n\right)' u_n\right] \xrightarrow{p} 0$ (or $\frac{1}{n} w_n\left(\widetilde{\Omega}_n\right)' u_n = o_p\left(\frac{r_n}{n}\right)$) is required for consistency (i.e., $\frac{1}{n} w_n\left(\widetilde{\Omega}_n\right)' u_n$ must not only converge in probability to zero, but must be of an order lower than $\frac{r_n}{n}$ in probability). Moreover, it is clear that restrictions are needed on the choice of the diagonal matrix $\widetilde{\Omega}_n$ in order to ensure that OC is satisfied. Given the other assumptions above, OC holds for $\omega-$class estimators which satisfy Assumption 4. To see this, note first that, under Assumption 4, $\widetilde{\Omega}_n$ can be decomposed as $\widetilde{\Omega}_n = \overline{\Omega}_n + \Xi_n$, where $\overline{\Omega}_n = diag\left(\overline{\omega}_{1,n}, ..., \overline{\omega}_{n,n}\right)$ and $\Xi_n = diag\left(\xi_{1,n},\ \xi_{2,n}, ..., \xi_{n,n}\right)$. It follows that, after some algebra, we can write

$$\frac{1}{r_n} w_n\left(\widetilde{\Omega}_n\right)' u_n = \frac{1}{r_n} y'_{2n}\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n - \frac{1}{r_n} v'_n Q_{Z_n}\Xi_n u_n. \tag{9}$$

Now, part (d) of Assumption 4 and the Cauchy-Schwartz inequality immediately imply that $\frac{1}{r_n} v'_n Q_{Z_n}\Xi_n u_n \xrightarrow{p} 0$ as $n \to \infty$. Furthermore, the first term of expression (9) has expectation zero since $E\left[y'_{2n}\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n\right] = E\left(b_n^{-1} c'_n Z'_n u_n\right) + E\left(v'_n\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n\right) = E\left(b_n^{-1} c'_n Z'_n u_n\right) + \sigma_{vu} E_{Z_n}\left[K_n - \sum_{i=1}^{n} \overline{\omega}_{i,n}(1 - h_{i,n})\right] = 0$, where the second equality follows from Assumption 4 (b). Part (c) of Assumption 4 then helps to ensure that $Var\left(\frac{1}{r_n} y'_{2n}\left[P_{Z_n} - Q_{Z_n}\overline{\Omega}_n\right] u_n\right) \to 0$ as $n \to \infty$, so that the first term of expression (9) also converges to zero asymptotically.

In contrast to $\omega-$class estimators which satisfy Assumption 4, the $2SLS$ estimator is not as well-centered in that it does not satisfy OC. Indeed, the $2SLS$ estimator takes $\widetilde{\Omega}_n$ to be the zero matrix; so that, for this estimator, $\frac{1}{r_n} E\left[w_n\left(\widetilde{\Omega}_n\right)' u_n\right] = \frac{1}{r_n} E\left[y'_{2n} P_{Z_n} u_n\right] = \frac{1}{r_n} E\left[v'_n P_{Z_n} u_n\right] = \left(\frac{K_n}{r_n}\right)\sigma_{vu}$, which grows without bound when instrument weakness is such that r_n grows slower than K_n. This lack of centering, in turn, causes the bias of the $2SLS$ estimator to increase as the number of instruments increases; and this bias problem becomes more severe when the instruments are weak as measured by a slower rate of growth of the concentration parameter. In consequence, as shown in theorem 2.4 above, the $2SLS$ estimator is inconsistent and has asymptotic bias equals to that of the OLS estimator unless r_n grows faster than K_n. This analysis is consistent with results reported in some recent Monte Carlo studies, such as those of Staiger and Stock (1997), Chao and Swanson (2001), and Hahn and Inoue (2002).

(vi) Within his framework, Bekker (1994) finds that, if $\frac{K_n}{n} \to \alpha \neq 0$ as $n \to \infty$, then $2SLS$ is inconsistent whereas $LIML$ is consistent. On the other hand, if $\frac{K_n}{n} \to 0$, then both $2SLS$ and $LIML$ are consistent. These results are special cases of the results provided in Theorem 2.1, Corollary 2.3, and Theorem 2.4 if we assume $r_n = n$, but analysis from our more general framework shows that some of these results need not hold more generally in the presence of weak instruments. In particular, the Bekker framework suggests that the $2SLS$ estimator is consistent whenever K_n grows at a slower rate than n, but more generally the consistency of the $2SLS$ estimator 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 of course $r_n = n$. Hence, instruments may be sufficiently weak (so that r_n is of a lower order relative to both n and K_n) in which case $2SLS$ is inconsistent, even though $\frac{K_n}{n} \to 0$ (see Theorem 2.4).

(vii) Interestingly, our analysis shows that ω-class estimators satisfying Assumption 4 may be consistent even if weakness in the instruments is such that the concentration parameter grows at a rate slower than K_n, so long as r_n grows faster than $\sqrt{K_n}$ as $n \to \infty$. Thus, the consistency result given in Theorem 2.1 allows for weaker instruments, as measured by the order of magnitude of the concentration parameter, than any previous results establishing consistent estimation of IV estimators. All previous consistency results assume that the concentration parameter grows at the same rate or at a faster rate than K_n.

(viii) For models that are weakly identified, our results suggest that it might be beneficial to use a lot of instruments, since even if each individual instrument is only weakly correlated with the endogenous regressors, the combined effect of using a lot of them might nevertheless allow the concentration parameter to be sufficiently large so that the precision with which we estimate is improved.[12] However, this advice must be qualified in two ways. First, it is well-known that the bias of the $2SLS$ estimator increases as the number of instruments increases, so that the use of the $2SLS$ estimator with a large number of instruments is not recommended as a way of dealing with weak identification. On the other hand, $\omega-$class estimators which satisfy our Assumption 4 are sufficiently well-centered, so that their bias does not increase appreciably with the number of instruments. Hence, when the instruments are weak, using an estimator such as $LIML$ with a large number of instruments may represent an empirical researcher's best chance at reliable point estimation. Secondly, our analysis focuses only on point estimation, whereas questions of set (or interval) estimation and hypothesis testing using a large number of weak instruments require further study and are left for future research. With regard to set estimation and hypothesis testing, a number of procedures have recently been shown by Staiger and Stock (1997), Wang and Zivot (1998), Kleibergen (2002), and Moreira (2002) to give asymptotically valid confidence region under both conventional and local-to-zero asymptotics. It would be of interest to also analyze the properties of these procedures in a many, weak instruments framework such as the one studied here.

[12] Phillips and Han (2003) give a remarkable result showing that, within a many-instrument framework, totally irrelevant instruments can be used to provide consistent estimation of the mean of a location model. However, their result does not generalize completely to the case of an instrumental variables regression model since there is in general no way to consistently estimate the entire structural coefficient vector, β, with totally irrelevant instruments, even if the number of such instruments is allowed to approach infinity.

3 Appendix

For the sake of brevity, some proofs are not included, and the reader is referred to the working paper version of this note (Chao and Swanson (2002)). We begin by providing four lemmas which are used to prove the main results of the paper.

Lemma A1: *Under Assumptions 1-4, suppose that $r_n \to \infty$ as $n \to \infty$ such that $\frac{\sqrt{K_n}}{r_n} \to 0$. Then, the following statements are true, for $n \to \infty$: (a) $\frac{V_n' M_n u_n}{r_n} \xrightarrow{p} 0$; (b) $\frac{V_n' M_n V_n}{r_n} \xrightarrow{p} 0$; (c) $\frac{u_n' M_n u_n}{r_n} \xrightarrow{p} 0$; (d) $\frac{C_n' Z_n' Q_{X_n} u_n}{b_n r_n} \xrightarrow{p} 0$; (e) $\frac{C_n' Z_n' Q_{X_n} V_n}{b_n r_n} \xrightarrow{p} 0$; and (f) $\frac{V_n' M_n Z_n C_n}{b_n r_n} \xrightarrow{p} 0$, where $M_n = \left[\left(P_{\overline{Z}_n} - P_{X_n} \right) - Q_{\overline{Z}_n} \widetilde{\Omega}_n Q_{X_n} \right]$.*

Proof of Lemma A1: Since each part of this lemma can be demonstrated by showing mean square convergence, we will only prove part (a) to give a flavor of the mean square calculations involved. Proofs for parts (b)-(f) are omitted to avoid redundancy. Details of these proofs can be found in an earlier version of our paper, Chao and Swanson (2002). To show part (a), it suffices to show that, under the assumptions of the lemma, the g^{th} element of $\frac{V_n' M_n u}{r_n}$ converges in probability to zero, i.e. $\frac{V_n^{(g)\prime} M_n u_n}{r_n} \xrightarrow{p} 0$, where $V_n^{(g)}$, $g \in \{1, ..., G\}$, denotes an arbitrary gth column of V_n. Note first that, given Assumption 4 and n sufficiently large, we can write

$$\frac{V_n^{(g)\prime} M_n u}{r_n} = \frac{V_n^{(g)\prime} \overline{M}_n u_n}{r_n} - \frac{V_n^{(g)\prime} Q_{\overline{Z}_n} \Xi_n Q_{X_n} u_n}{r_n}, \tag{10}$$

where $\overline{M}_n = \left(P_{\overline{Z}_n} - P_{X_n} \right) - Q_{\overline{Z}_n} \overline{\Omega}_n Q_{X_n}$ and where $\overline{\Omega}_n = diag\left(\overline{\omega}_{1,n}, \overline{\omega}_{2,n}, ..., \overline{\omega}_{n,n}\right)$ and $\Xi_n = diag\left(\xi_{1,n},\ \xi_{2,n}, ..., \xi_{n,n}\right)$. We will show that both of the terms on the right-hand side of (10) converge in probability to zero. To proceed, note that to show that $\frac{V_n^{(g)\prime} \overline{M}_n u_n}{r_n} \xrightarrow{p} 0$ as $n \to \infty$, it suffices to show that $\frac{1}{r_n^2} E\left(V_n^{(g)\prime} \overline{M}_n u_n \right)^2 \to 0$. Next, let $\overline{m}_{ij,n}$ denote the $(i,j)^{th}$ element of $\overline{M}_n$ and let v_{ig} denote the $(i,g)^{th}$ element of V_n, and we can calculate the second moment of $\frac{V_n^{(g)\prime} \overline{M}_n u_n}{r_n}$ as follows:

$$\begin{aligned}
\frac{1}{r_n^2} E\left(V_n^{(g)\prime} \overline{M}_n u_n \right)^2 &= \left(\frac{1}{r_n^2} \right) E(v_{ig}^2 u_i^2) E_{\overline{Z}_n} \left[\sum_{i=1}^{n} \overline{m}_{ii,n}^2 \right] + \left(\frac{1}{r_n^2} \right) \Sigma_{VV}^{(g,g)} \sigma_{uu} E_{\overline{Z}_n} \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} \overline{m}_{ij,n}^2 + \right. \\
&\quad \left. \sum_{j=2}^{n} \sum_{i=1}^{j-1} \overline{m}_{ij,n}^2 \right] + \left(\frac{2\left(\sigma_{Vu}^{g}\right)^2}{r_n^2} \right) E_{\overline{Z}_n} \left[\sum_{i=2}^{n} \sum_{j=1}^{i-1} \overline{m}_{ii,n}\, \overline{m}_{jj,n} + \sum_{i=2}^{n} \sum_{j=1}^{i-1} \overline{m}_{ij,n}\, \overline{m}_{ji,n} \right] \\
&= \mathcal{A}_n + \mathcal{B}_n + \mathcal{C}_n, \text{ say,}
\end{aligned} \tag{11}$$

where $E_{\overline{Z}_n}(\cdot)$ denotes the expectation taken with respect to the probability measure of $\overline{Z}_n$ and where the first equality above follows from Assumption 3, parts (a) and (b). Dealing first with $\mathcal{A}_n$, observe that with probability one

$$
\begin{aligned}
\left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[\sum_{i=1}^{n} \overline{m}_{ii,n}^2\right] &\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2) Tr\left[\overline{M}_n' \overline{M}_n\right] \\
&\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + \left|Tr\left(\overline{\Omega}_n Q_{\overline{Z}_n} \overline{\Omega}_n Q_{X_n}\right)\right|\right] \\
&\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + \sqrt{Tr\left(\overline{\Omega}_n Q_{\overline{Z}_n} \overline{\Omega}_n\right)}\sqrt{Tr\left(\overline{\Omega}_n Q_{X_n} \overline{\Omega}_n\right)}\right] \\
&\leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + \sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right], \qquad (12)
\end{aligned}
$$

where the third inequality above follows from the Cauchy-Schwarz inequality and where the fourth inequality above follows from the fact that $Q_{\overline{Z}_n}$ and Q_{X_n} are symmetric, idempotent matrices. To see the argument behind the fourth inequality, take $Q_{\overline{Z}_n}$ as an example, and note that we can write $Q_{\overline{Z}_n} = B_n \Lambda_n B_n'$, where B_n is an orthogonal matrix (i.e., $B_n B_n' = I_n = B_n' B_n$) whose columns are the orthonormal eigenvectors of $Q_{\overline{Z}_n}$ and Λ_n is a diagonal matrix with $n - K_n - J$ one's and $K_n + J$ zero's along the main diagonal; hence, it follows that $Tr\left(\overline{\Omega}_n Q_{\overline{Z}_n} \overline{\Omega}_n\right) = Tr\left(\overline{\Omega}_n B_n \Lambda_n B_n' \overline{\Omega}_n\right) \leq Tr\left(\overline{\Omega}_n B_n B_n' \overline{\Omega}_n\right) = Tr\left(\overline{\Omega}_n^2\right) = \sum_{i=1}^{n} \overline{\omega}_{i,n}^2$. By a similar argument, $Tr\left(\overline{\Omega}_n Q_{X_n} \overline{\Omega}_n\right) \leq \sum_{i=1}^{n} \overline{\omega}_{i,n}^2$. Now, note that since the bound given by (12) holds with probability one, we deduce that

$$
\mathcal{A}_n = \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2) E_{\overline{Z}_n}\left[\sum_{i=1}^{n} \overline{m}_{ii,n}^2\right] \leq \left(\frac{1}{r_n^2}\right) E(v_{ig}^2 u_i^2)\left[K_n + E_{\overline{Z}_n}\left(\sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right)\right] = O\left(K_n / r_n^2\right); \qquad (13)
$$

where the last equality above follows from Assumption 4(c). Hence, $\mathcal{A}_n \to 0$ as $n \to \infty$ if $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$.

Turning our attention next to the term $\mathcal{B}_n$, we note that similar to the argument given for $\mathcal{A}_n$ above, we have that with probability one

$$
\begin{aligned}
&\left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} \overline{m}_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1} \overline{m}_{ij,n}^2\right] \\
\leq\ &\left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu} Tr\left[\overline{M}_n' \overline{M}_n\right] \leq \left(\frac{1}{r_n^2}\right) \sigma_{VV}^{(g,g)} \sigma_{uu}\left[K_n + \sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right]. \qquad (14)
\end{aligned}
$$

It follows again that since the bound given in (14) holds with probability one, we deduce that

$$
\begin{aligned}
\mathcal{B}_n &= \left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu} E_{\overline{Z}_n}\left[\sum_{i=2}^{n}\sum_{j=1}^{i-1} \overline{m}_{ij,n}^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1} \overline{m}_{ij,n}^2\right] \\
&\leq \left(\frac{1}{r_n^2}\right) \Sigma_{VV}^{(g,g)} \sigma_{uu}\left[K_n + E_{\overline{Z}_n}\left(\sum_{i=1}^{n} \overline{\omega}_{i,n}^2\right)\right] = O\left(K_n / r_n^2\right), \qquad (15)
\end{aligned}
$$

so that $\mathcal{B}_n \to 0$ as $n \to \infty$ if $\frac{\sqrt{K_n}}{r_n} \to 0$ as $n \to \infty$.

Finally, turning to the term $\mathcal{C}_n$, we note that

$$\begin{aligned} & 2\left(\frac{1}{r_n^2}\right)\left(\sigma_{Vu}^g\right)^2\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}\;+\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right] \\ = & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left\{Tr\left[\overline{M}_n^2\right]+\left(Tr\left[\overline{M}_n\right]\right)^2-2\sum_{i=1}^n\overline{m}_{ii}^2\right\}, \end{aligned} \tag{16}$$

so that with probability one

$$\begin{aligned} & 2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}\;+\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right]\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left[K_n+\left|Tr\left(Q_{\overline{Z}_n}\overline{\Omega}_nQ_{\overline{Z}_n}\overline{\Omega}_nQ_{X_n}\right)\right|\right]+2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|K_n+\sum_{i=1}^n\overline{\omega}_{i,n}^2\right| \\ \leq & \left(\frac{\sigma_{Vu}^g}{r_n}\right)^2K_n+\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2Tr\left(\overline{\Omega}_n^2\right)+2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|K_n+\sum_{i=1}^n\overline{\omega}_{i,n}^2\right| \\ = & 3\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left[K_n+\sum_{i=1}^n\overline{\omega}_{i,n}^2\right], \end{aligned} \tag{17}$$

where the first equality above makes use of the fact that under Assumption 4(b) $Tr\left(P_{\overline{Z}_n}-P_{X_n}-Q_{\overline{Z}_n}\overline{\Omega}_nQ_{X_n}\right)=K_n-Tr\left(Q_{\overline{Z}_n}\overline{\Omega}_n\right)=0$ $a.s.$In addition, the second inequality in (17) follows in part from the Cauchy-Schwarz inequality and in part from argument similar to that given subsequent to expression (12) above. Since the upper bound given in (17) above holds almost surely, it follows that

$$\begin{aligned} |\mathcal{C}_n| = & 2\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left|E_{\overline{Z}_n}\left[\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ii,n}\,\overline{m}_{jj,n}\;+\sum_{i=2}^n\sum_{j=1}^{i-1}\overline{m}_{ij,n}\,\overline{m}_{ji,n}\right]\right| \\ \leq & 3\left(\frac{\sigma_{Vu}^g}{r_n}\right)^2\left[K_n+\sum_{i=1}^nE\left(\overline{\omega}_{i,n}^2\right)\right]=O\left(K_n/r_n^2\right), \end{aligned} \tag{18}$$

so that $\mathcal{C}_n\to 0$ as $n\to\infty$ if $\frac{\sqrt{K_n}}{r_n}\to 0$ as $n\to\infty$. It follows from (13), (15), and (18) that $\frac{1}{r_n^2}E\left(V_n^{(g)\prime}\overline{M}_nu_n\right)^2\to 0$, as $n\to\infty$ under the condition that $\frac{\sqrt{K_n}}{r_n}\to 0$ as $n\to\infty$,from which it follows immediately as a direct consequence of Chebyshev's inequality that $\frac{V_n^{(g)\prime}\overline{M}_nu_n}{r_n}\xrightarrow{p}0$. Next, we show that, under the assumptions of the lemma, $\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}\Xi_nQ_{X_n}u_n}{r_n}\xrightarrow{p}0$. To show this, note that

$$\begin{aligned} \left|\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}\Xi_nQ_{X_n}u_n}{r_n}\right| \leq & \sqrt{\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}V_n^{(g)}}{r_n}}\sqrt{\frac{u_n'Q_{X_n}\Xi_n^2Q_{X_n}u_n}{r_n}} \\ \leq & \left[\left(\frac{n}{r_n}\right)\sup_i|\xi_{i,n}|\right]\sqrt{\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}V_n^{(g)}}{n}}\sqrt{\frac{u_n'Q_{X_n}u_n}{n}}, \end{aligned} \tag{19}$$

where the first inequality above follows from Cauchy-Schwarz. Next, note that standard arguments yield $\frac{u_n'Q_{X_n}u_n}{n}\xrightarrow{p}\sigma_{uu}<\infty$, and, from part (e) of lemma A2 given below, we obtain $\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}V_n^{(g)}}{n}\xrightarrow{p}\Sigma_{VV}^{(g,g)}(1-\alpha)<\infty$, where $\Sigma_{VV}^{(g,g)}$

denotes the $(g,g)^{th}$ element of Σ_{VV}. Moreover, it follows from assumption 4(d) that $\left(\frac{n}{r_n}\right)\sup_i|\xi_{i,n}| \xrightarrow{p} 0$. The Slutsky Theorem then implies that $\frac{V_n^{(g)\prime}Q_{\overline{Z}_n}\Xi_n Q_{X_n}u_n}{r_n} \xrightarrow{p} 0$, as n, K_n, $r_n \to \infty$ such that $\frac{K_n}{n} \to \alpha$ and $\frac{\sqrt{K_n}}{r_n} \to 0$. The desired result follows directly in light of equation (10). □

Lemma A2: *Under Assumptions 2-3, suppose that* $r_n \to \infty$ *as* $n \to \infty$ *such that* $\frac{\sqrt{K_n}}{r_n} \to 0$. *Define* $M_n^* = \left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - \left(\frac{K_n}{n-K_n-J}\right)Q_{\overline{Z}_n}\right]$; *then, the following statements are true, as* $n \to \infty$: *(a)* $\frac{V_n' M_n^* u_n}{r_n} \xrightarrow{p} 0$; *(b)* $\frac{V_n' M_n^* V_n}{r_n} \xrightarrow{p} 0$; *(c)* $\frac{u_n' M_n^* u_n}{r_n} \xrightarrow{p} 0$; *(d)* $\frac{V_n' Q_{\overline{Z}_n} u_n}{n} \xrightarrow{p} \sigma_{Vu}(1-\alpha)$; *(e)* $\frac{V_n' Q_{\overline{Z}_n} V_n}{n} \xrightarrow{p} \Sigma_{VV}(1-\alpha)$; and *(f)* $\frac{u_n' Q_{\overline{Z}_n} u_n}{n} \xrightarrow{p} \sigma_{uu}(1-\alpha)$.

Proof of Lemma A2: For the sake of brevity, the proof of this lemma is omitted. Interested readers are referred to Chao and Swanson (2002) for a proof of this result.

Lemma A3: *Under Assumptions 2-3, the following statements are true, as* $n \to \infty$: *(a)* $\frac{V_n'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n} \xrightarrow{p} \sigma_{Vu}$; *and (b)* $\frac{V_n'\left(P_{\overline{Z}_n} - P_{X_n}\right)V_n}{K_n} \xrightarrow{p} \Sigma_{VV}$.

Proof of Lemma A3: For the sake of brevity, the proof of this lemma is omitted. Interested readers are referred to Chao and Swanson (2002) for a proof of this result.

Lemma A4 : *Under Assumptions 1-3, let* $\widehat{\lambda}_{LIML,n}$ *be the smallest root of the determinantal equation given by (6). Suppose that* $r_n \to \infty$, *as* $n \to \infty$, *such that* $\frac{\sqrt{K_n}}{r_n} \to 0$. *Then,* $\widehat{\lambda}_{LIML,n} = \frac{n-J}{n-K_n-J} + \xi_n$, *where* $\xi_n = o_p\left(\frac{r_n}{n}\right)$.

Proof of Lemma A4: To proceed, define $Y_n = [y_{1n}, Y_{2n}]$ and $\Upsilon = \begin{pmatrix} 1 & 0 \\ -\beta_0 & I_G \end{pmatrix}$, and note that the smallest root of the determinantal equation (6) is the same as the smallest root of the equation

$$\det\left\{\Upsilon' Y_n' Q_{X_n} Y_n \Upsilon - \lambda_n \Upsilon' Y_n' Q_{\overline{Z}_n} Y_n \Upsilon\right\} = 0, \tag{20}$$

where the equivalence follows from the fact that

$$\begin{aligned} \det\left\{\Upsilon' Y_n' Q_{X_n} Y_n \Upsilon - \lambda_n \Upsilon' Y_n' Q_{\overline{Z}_n} Y_n \Upsilon\right\} &= \\ &= \det\{\Upsilon\}\det\left\{Y_n' Q_{X_n} Y_n - \lambda_n Y_n' Q_{\overline{Z}_n} Y_n\right\}\det\{\Upsilon\} \\ &= \det\left\{Y_n' Q_{X_n} Y_n - \lambda_n Y_n' Q_{\overline{Z}_n} Y_n\right\}, \end{aligned}$$

given that $\det\{\Upsilon\} = 1$. Note also that the inverses, which appear in the projection matrices Q_{X_n} and $Q_{\overline{Z}_n}$ in equation (20), are all well-defined with probability one for n sufficiently large in light of Assumption 2. Moreover, it can be shown, by straightforward but tedious calculations, that $\widehat{\lambda}_{LIML,n}$, the smallest root of the determinantal equation (20), can be given the representation $\widehat{\lambda}_{LIML,n} = \frac{n-J}{n-K_n-J} + \widehat{\tau}_{LIML,n}\left(\frac{r_n}{n}\right)$, where $\widehat{\tau}_{LIML,n}$ is the smallest root of the determinantal equation

$$\det\left\{\left(\begin{array}{cc}\frac{u_n' M_n^* u_n}{r_n} & \frac{u_n' Q_{X_n} Z_n C_n}{b_n r_n} + \frac{u_n' M_n^* V_n}{r_n} \\ \frac{C_n' Z_n' Q_{X_n} u_n}{b_n r_n} + \frac{V_n' M_n^* u_n}{r_n} & \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2 r_n} + \frac{C_n' Z_n' Q_{X_n} V_n}{b_n r_n} + \frac{V_n' Q_{X_n} Z_n C_n}{b_n r_n} + \frac{V_n' M_n^* V_n}{r_n}\end{array}\right)\right.$$
$$\left. - \tau_n \left(\begin{array}{cc}\frac{u_n' Q_{\overline{Z}_n} u_n}{n} & \frac{u_n' Q_{\overline{Z}_n} V_n}{n} \\ \frac{V_n' Q_{\overline{Z}_n} u_n}{n} & \frac{V_n' Q_{\overline{Z}_n} V_n}{n}\end{array}\right)\right\} = 0, \qquad (21)$$

with $M_n^* = \left[\left(P_{\overline{Z}_n} - P_{X_n}\right) - \left(\frac{K_n}{n-K_n-J}\right) Q_{\overline{Z}_n}\right]$. It then follows from Assumption 2(c), parts (d) and (e) of lemma A1 and parts (a)-(f) of Lemma A2 and by continuity that as $n \to \infty$, the difference between $\widehat{\tau}_{LIML,n}$ and the smallest root of

$$\det\left\{\left(\begin{array}{cc} 0 & 0 \\ 0 & \Psi_n \end{array}\right) - \tau_n \left(\begin{array}{cc} \sigma_{uu}(1-\alpha) & \sigma_{Vu}'(1-\alpha) \\ \sigma_{Vu}(1-\alpha) & \Sigma_{VV}(1-\alpha) \end{array}\right)\right\} = 0 \qquad (22)$$

goes to zero in probability as $n \to \infty$. Since the smallest root of (22) is obviously zero, we deduce immediately that $\widehat{\tau}_{LIML,n} = o_p(1)$, from which it follows that $\widehat{\lambda}_{LIML,n} = \frac{n-J}{n-K_n-J} + o_p\left(\frac{r_n}{n}\right)$, as required. □

Proof of Theorem 2.1: To proceed, note first that, given Assumption 2(b), the inverses which appear in the projection matrices P_{X_n}, $P_{\overline{Z}_n}$, Q_{X_n}, and $Q_{\overline{Z}_n}$ in the expression $M_n = \left(P_{\overline{Z}_n} - P_{X_n}\right) - Q_{\overline{Z}_n} \widetilde{\Omega}_n Q_{X_n}$ are all well-defined with probability one for n sufficiently large. Hence, for n sufficiently large, we can write

$$\frac{Y_{2n}' M_n Y_{2n}}{r_n} = \frac{C_n' Z_n' Q_{X_n} Z_n C_n}{b_n^2 r_n} + \frac{C_n' Z_n' Q_{X_n} V_n}{b_n r_n} + \frac{V_n' M_n Z_n C_n}{b_n r_n} + \frac{V_n' M_n V_n}{r_n}, \qquad (23)$$

Now, it follows from Assumption 2(c) and parts (b), (e), and (f) of lemma A1 that $\frac{Y_{2n}' M_n Y_{2n}}{r_n} = \Psi_n + o_p(1)$,where $\Psi_n = b_n^{-2} r_n^{-1} C_n' Z_n' Q_{X_n} Z_n C_n$ is positive definite almost surely for n sufficiently large given Assumption 2(c). Moreover, for n sufficiently large, we can write $\frac{Y_2' M_n u_n}{r_n} = \frac{C_n' Z_n' Q_{X_n} u_n}{b_n r_n} + \frac{V_n' M_n u_n}{r_n}$, so that $\frac{Y_{2n}' M_n u_n}{r_n} \xrightarrow{p} 0$, as $n \to \infty$,given parts (a) and (d) of lemma A1. Next, note that we can write

$$\widehat{\beta}_{\omega,n} - \beta_0 = \left(\left[\frac{Y_{2n}' M_n Y_{2n}}{r_n}\right]^+ \left[\frac{Y_{2n}' M_n Y_{2n}}{r_n}\right] - I_G\right)\beta_0 + \left[\frac{Y_{2n}' M_n Y_{2n}}{r_n}\right]^+ \left[\frac{Y_{2n}' M_n u_n}{r_n}\right]. \qquad (24)$$

It follows by Proposition 2.30 of White (1999) and the Slutsky's theorem that $\left[\frac{Y_{2n}' M_n Y_{2n}}{r_n}\right]^+ \left[\frac{Y_{2n}' M_n Y_{2n}}{r_n}\right] - I_G \xrightarrow{p} 0$ and that $\left[\frac{Y_{2n}' M_n Y_{2n}}{r_n}\right]^+ \left[\frac{Y_{2n}' M_n u_n}{r_n}\right] \xrightarrow{p} 0$,from which we deduce immediately that $\widehat{\beta}_{\omega,n} \xrightarrow{p} \beta_0$, as required. □

Proof of Corollary 2.3: To show part (a), we verify that $LIML$ satisfies Assumption 4. To proceed, note that, for the case of $LIML$, we have $\widetilde{\omega}_{i,n} = \widetilde{\omega}_n = \widehat{\lambda}_{LIML,n} - 1$ for $i = 1, ..., n$. Now, set $\overline{\omega}_{i,n} = \overline{\omega}_n = \left(\frac{n-J}{n-K_n-J}\right) - 1 = \frac{K_n}{n-K_n-J}$, for all i. Observe that $\overline{\lim_{n\to\infty}}\, \bar{l}_n = \overline{\lim_{n\to\infty}}\left(\sup_{1\le i\le n} |\overline{\omega}_{i,n}|\right) = \overline{\lim_{n\to\infty}}\left(\frac{K_n}{n-K_n-J}\right) < \infty$, since we assume that $\frac{K_n}{n} \to \alpha$ for $0 \le \alpha < 1$. Hence, Assumption 4(a) is satisfied. Next, observe that, in this case, $\sum_{i=1}^n \overline{\omega}_{i,n}(1-h_{i,n}) = \left(\frac{K_n}{n-K_n-J}\right)\sum_{i=1}^n (1-h_{i,n}) = \left(\frac{K_n}{n-K_n-J}\right)(n-K_n-J) = K_n$, so that Assumption 4(b) is satisfied. Furthermore, note that, in this case, $\sum_{i=1}^n E\left(\overline{\omega}_{i,n}^2\right) = K_n\left\{\frac{K_n n}{(n-K_n-J)^2}\right\} = O(K_n)$,since we assume that $K_n = O(n)$. Thus, Assumption 4(c) is also satisfied. Finally, note that, by construction, $\xi_{i,n} = \xi_n = \widetilde{\omega}_n - \overline{\omega}_n$, where we have dropped the subscript i on the right-hand side because, in the case of $LIML$, ω does not vary with i. Given that we have set $\overline{\omega}_n = \frac{K_n}{n-K_n-J}$, Lemma A4 implies that $\sup_i |\xi_{i,n}| = |\xi_n| = |\widetilde{\omega}_n - \overline{\omega}_n| = \left|\widehat{\lambda}_{LIML,n} - 1 - \frac{K_n}{n-K_n-J}\right| = \left|\widehat{\lambda}_{LIML,n} - \frac{n-J}{n-K_n-J}\right| = o_p\left(\frac{r_n}{n}\right)$,so that Assumption 4(d) is satisfied as well.

To show part (b), we first compare expression (8) with expression (3) or expression (4) and observe that $JIVE$ can also be obtained as a special case of the ω-class, by setting $\omega_{i,n} = \left(\frac{1}{1-h_{i,n}}\right)$, for $i = 1, ..., n$ or, alternatively, by setting $\widetilde{\omega}_{i,n} = \left(\frac{h_{i,n}}{1-h_{i,n}}\right)$, for $i = 1, ..., n$. Given Assumption J, we can then verify that $JIVE$ satisfies Assumption 4. To proceed, set $\overline{\omega}_{i,n} = \left[h_{i,n} - \frac{J}{n}\right]\left(\frac{1}{1-h_{i,n}}\right)$ for $i = 1, ..., n$; so that, by construction, $\xi_{i,n} = \frac{J}{n}\left(\frac{1}{1-h_{i,n}}\right)$. Now, observe that, in this case with probability one, $\overline{l}_n = \sup_{1\leq i\leq n} |\overline{\omega}_{i,n}| \leq \left[\overline{h} + \frac{J}{n}\right]\left(\frac{1}{1-\overline{h}}\right)$ for all n sufficiently large, from which it follows that $\overline{\lim}_{n\to\infty} \overline{l}_n \leq \left(\frac{\overline{h}}{1-\overline{h}}\right) < \infty$ *a.s.*, so that Assumption 4(a) is satisfied. Next, observe that in this case, $\sum_{i=1}^{n} \overline{\omega}_{i,n}(1 - h_{i,n}) = \sum_{i=1}^{n}\left[h_{i,n} - \frac{J}{n}\right]\left(\frac{1}{1-h_{i,n}}\right)(1 - h_{i,n}) = \sum_{i=1}^{n}\left[h_{i,n} - \frac{J}{n}\right] = K_n$, so that Assumption 4(b) is satisfied. Moreover:

$$\begin{aligned}
\sum_{i=1}^{n} E\left(\overline{\omega}_{i,n}^2\right) &\leq \left(\frac{1}{1-\overline{h}}\right)^2 E\left\{\sum_{i=1}^{n}\left[h_{i,n}^2 - 2h_{i,n}\frac{J}{n} + \frac{J^2}{n^2}\right]\right\} \\
&\leq \left(\frac{1}{1-\overline{h}}\right)^2\left[K_n + J - 2\frac{(K_n + J)J}{n} + \frac{J^2}{n}\right] = O(K_n), \qquad (25)
\end{aligned}$$

where the second inequality follows from the fact that, even if we ignore Assumption J, it must be that $0 \leq h_{i,n} \leq 1$; and, hence, $\sum_{i=1}^{n} h_{i,n}^2 \leq \sum_{i=1}^{n} h_{i,n} = K_n + J$. It follows that Assumption 4(c) is also satisfied. Finally, note that $\sup_{1\leq i\leq n} |\xi_{i,n}| = \sup_{1\leq i\leq n} \frac{J}{n}\left(\frac{1}{1-h_{i,n}}\right) \leq \frac{J}{n}\left(\frac{1}{1-\overline{h}}\right) = O(n^{-1})$, where the inequality holds by Assumption J, almost surely. Hence, Assumption 4(d) is satisfied as well. □

Proof of Theorem 2.4: To show part (a), note first that, for n sufficiently large, the inverses which appear in the projection matrices $P_{\overline{Z}_n}$, P_{X_n}, and Q_{X_n} are all well-defined with probability one, so we can write

$$\begin{aligned}
\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n} &= \left(\frac{r_n}{K_n}\right)\frac{C_n'Z_n'Q_{X_n}Z_nC_n}{b_n^2 r_n} + \left(\frac{r_n}{K_n}\right)\frac{C_n'Z_n'Q_{X_n}V_n}{b_n r_n} \\
&\quad + \left(\frac{r_n}{K_n}\right)\frac{V_n'Q_{X_n}Z_nC_n}{b_n r_n} + \frac{V_n'\left(P_{\overline{Z}_n} - P_{X_n}\right)V_n}{K_n}. \qquad (26)
\end{aligned}$$

Now, since it is assumed in part (a) that $\frac{r_n}{K_n} \to 0$ as $n \to \infty$, it follows from Assumption 2(c) and Lemma A1 part (e) and from Lemma A3 part (b) that $\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)Y_{2n}}{K_n} \xrightarrow{p} \Sigma_{VV}$,where Σ_{VV} is positive definite by Assumption 3(b) and is, thus, nonsingular. Moreover, for n sufficiently large, we can write $\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n} = \left(\frac{r_n}{K_n}\right)\frac{C_n'Z_n'Q_{X_n}u_n}{b_n r_n} + \frac{V_n'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n}$, so that $\frac{Y_{2n}'\left(P_{\overline{Z}_n} - P_{X_n}\right)u_n}{K_n} \xrightarrow{p} \sigma_{Vu}$, as $n \to \infty$, by Lemma A1 part (d) and

Lemma A3 part (a). Next, write

$$\widehat{\beta}_{2SLS,n}-\beta_0 = \left(\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}\right]^{+}\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}\right]-I_G\right)\beta_0 + \left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}\right]^{+}\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{K_n}\right], \tag{27}$$

Since the Slutsky's theorem implies that

$$\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}\right]^{+}\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}\right]-I_G\xrightarrow{p}0$$

and

$$\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}\right]^{+}\left[\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{K_n}\right]\xrightarrow{p}\Sigma_{VV}^{-1}\sigma_{Vu},$$

it follows immediately by a further application of the Slutsky's theorem that $\widehat{\beta}_{2SLS,n}\xrightarrow{p}\beta_0+\Sigma_{VV}^{-1}\sigma_{Vu}$, 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 2(c) and Lemma A1 part (e) and from Lemma A3 part (b) that $\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{K_n}=(\delta\Psi_n+\Sigma_{VV})+o_p(1)$,where $\Psi_n=b_n^{-2}r_n^{-1}C_n'Z_n'Q_{X_n}Z_nC_n$ is positive definite almost surely for n sufficiently large as a result of Assumption 2(c). In addition, from part (d) of Lemma A1 and part (a) of Lemma A3, we deduce that $\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{K_n}\xrightarrow{p}\sigma_{Vu}$.The desired result, thus, follows from Proposition 2.30 of White (1999).

Part (c) can be shown using arguments similar to parts (a) and (b) above, except that here we standardize both $Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}$ and $Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n$ by r_n instead of K_n. In this case, it is easy to show that $\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)Y_{2n}}{r_n}=\Psi_n+o_p(1)$,where Ψ_n is positive definite almost surely for n sufficiently large, and that $\frac{Y_{2n}'\left(P_{\overline{Z}_n}-P_{X_n}\right)u_n}{r_n}\xrightarrow{p}0$, as $n\to\infty$. Weak consistency of $\widehat{\beta}_{2SLS,n}$ then follows as a consequence of the Slutsky Theorem. Please see Chao and Swanson (2002) for details. □

4 References

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

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

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.

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

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

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

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.

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 and University of Maryland.

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

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

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

Hahn, J. (1997). "Optimal Inference with Many Instruments," Working Paper, University of Pennsylvania.

Hahn, J., J. Hausman, and G. Kuersteiner (2001). "Higher Order MSE of Jackknife 2SLS," Working Paper, MIT.

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 (2003). "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," Working Paper, MIT.

Hall, A.R. and F. P. M. Peixe (2000). "A Consistent Method for the Selection of Relevant Instruments," Working Paper, North Carolina State University.

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

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

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.

Moreira, M. (2002). "A Conditional Likelihood Ratio Test for Structural Model," *Econometrica,* forthcoming.

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.

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).

Phillips, P.C.B. (1989). "Partially Identified Econometric Models," *Econometric Theory*, 5, 181-240.

Phillips, P.C.B. and C. Han (2003). "IV Estimation with Irrelevant Instruments," Working Paper, Yale University.

Rothenberg, T. (1983). "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).

Sargan, J.D. (1988). "The Identification and Estimation of Sets of Simultaneous Stochastic Equations," in E. Maasoumi (Ed.) *Contributions to Econometrics Volume 1.* Cambridge: Cambridge University Press.

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," Working Paper, Harvard University.

Stock, J.H. and M. Yogo (2003b). "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," Working Paper, Harvard University.

Van Hasselt, M. (2000). "The Instrumental Variables Estimator and the Effects of Non-Gaussian Distributions," Masters Thesis, University of Groningen.

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

White, H. (1999). *Asymptotic Theory for Econometricians,* Academic Press, San Diego.