CHAPTER 6

Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments

James H. Stock and Motohiro Yogo

ABSTRACT

This paper extends Staiger and Stock's (1997) weak instrument asymptotic approximations to the case of many weak instruments by modeling the number of instruments as increasing slowly with the number of observations. It is shown that the resulting "many weak instrument" approximations can be calculated sequentially by letting first the sample size, and then the number of instruments, tend to infinity. The resulting distributions are given for k-class estimators and test statistics.

1. INTRODUCTION

Most of the literature on the distribution of statistics in instrumental variables (IV) regression assumes, either implicitly or explicitly, that the number of instruments (K_2) is small relative to the number of observations (T); see Rothenberg's (1984) survey of Edgeworth approximations to the distributions of IV statistics. In some applications, however, the number of instruments can be large; for example, Angrist and Krueger (1991) had 178 instruments in one of their specifications. Sargan (1975), Kunitomo (1980), and Morimune (1983) provided early asymptotic treatments of many instruments. More recently, Bekker (1994) obtained first-order distributions of various IV estimators under the assumptions that $K_2 \to \infty$, $T \to \infty$, and $K_2/T \to c$, $0 \le c < 1$, when the so-called concentration parameter (μ^2) is proportional to the sample size and the errors are Gaussian. Chao and Swanson (2002) have explored the consistency of IV estimators with weak instruments when the number of instruments is large, in the sense that K_2 is also modeled as increasing to infinity, but more slowly than T.

This paper continues this line of research on the asymptotic distribution of IV estimators when there are many instruments. Our focus is on the case of many weak instruments, that is, when there are many instruments that are, on average, only weakly correlated with the included endogenous regressors. Specifically, we extend the weak instrument asymptotics developed in Staiger and Stock (1997) to the case of many instruments. The key technical device of the Staiger–Stock (1997) weak instrument asymptotics is fixing the expected value of the concentration parameter, along with the number of instruments, as the sample size increases. Here, we extend this to the case that the expected value of the concentration parameter is proportional to the number of instruments, and the number of instruments is allowed to increase slowly with the sample size, specifically, as $T \to \infty$, $K_2 \to \infty$, $E(\mu^2)/K_2 \to \mathbf{\Lambda}_\infty$ (a fixed matrix), and $K_2^4/T \to 0$. We refer to asymptotic limits taken under sequences satisfying these conditions as *many weak instrument limits*. (The term "many" should not be overinterpreted because while the number of instruments is allowed to tend to infinity, the condition $K_2^4/T \to 0$ requires it to do so very slowly relative to the sample size.) Under these conditions, and some additional technical conditions stated in Section 2 (including i.i.d. sampling and existence of fourth moments), it is shown that the limits of k-class IV statistics as K_2 and T jointly tend to infinity can in general be computed using sequential asymptotic limits. Under sequential asymptotics, the fixed-K_2 weak instrument limit is obtained first, then the limit of that distribution is taken as $K_2 \to \infty$. The advantage of this "first T then K_2" approach is that the sequential calculations are simpler than the calculations that arise along the joint sequence of (K_2, T). A potential disadvantage of this approach is that this simplicity comes at the cost of a stronger rate condition than might be obtained along the joint sequence.

We begin in Section 2 by specifying the model, the k-class IV statistics of interest, and our assumptions. Section 3 justifies the sequential asymptotics by showing that, under these assumptions, a key uniform convergence condition holds. In Section 4, we derive the many weak instrument limits of k-class estimators and test statistics using sequential asymptotics. These many weak instrument limits are used in Stock and Yogo (2004) to develop tests for weak instruments when the number of instruments is moderate. Some of these results might be of more general interest, however; for example, Chao and Swanson (2002) show that LIML is consistent under these conditions, and in this paper we provide its $\sqrt{K_2}$-limiting distribution. Section 5 provides some concluding remarks.

2. THE MODEL, STATISTICS, AND ASSUMPTIONS

2.1. Model and Notation

We consider the IV regression model with n included endogenous regressors:

$$\mathbf{y} = \mathbf{Y}\beta + \mathbf{u}, \tag{2.1}$$

$$\mathbf{Y} = \mathbf{Z}\mathbf{\Pi} + \mathbf{V}, \tag{2.2}$$

where $\mathbf{y}$ is the $T \times 1$ vector of T observations on the dependent variable, $\mathbf{Y}$ is the $T \times n$ matrix of n included endogenous variables, $\mathbf{Z}$ is the $T \times K_2$ matrix of K_2 excluded exogenous variables to be used as instruments, and $\mathbf{u}$ and $\mathbf{V}$ are a $T \times 1$ vector and $T \times n$ matrix of disturbances, respectively. The $n \times 1$

vector β and $K_2 \times n$ matrix $\mathbf{\Pi}$ are unknown parameters. Throughout this paper we exclusively consider inference about β.

It is useful to introduce some additional notation. Let $\mathbf{Z}_t = (Z_{1t} \cdots Z_{K_2t})'$, $\mathbf{V}_t = (V_{1t} \cdots V_{nt})'$, $\underline{\mathbf{Y}} = [\mathbf{y}\ \mathbf{Y}]$, $\mathbf{Q}_{ZZ} = E(\mathbf{Z}_t\mathbf{Z}_t')$,

$$\Sigma = E\left[\begin{pmatrix} u_t \\ \mathbf{V}_t \end{pmatrix}\begin{pmatrix} u_t & \mathbf{V}_t' \end{pmatrix}\right] = \begin{bmatrix} \sigma_{uu} & \Sigma_{u\mathbf{V}} \\ \Sigma_{\mathbf{V}u} & \Sigma_{\mathbf{VV}} \end{bmatrix}, \tag{2.3}$$

$$\rho = \Sigma_{\mathbf{VV}}^{-1/2'}\Sigma_{\mathbf{V}u}\sigma_{uu}^{-1/2}, \tag{2.4}$$

$$\mathbf{C} = \sqrt{T}\mathbf{\Pi}, \quad \text{and} \tag{2.5}$$

$$\Lambda_{K_2} = T\Sigma_{\mathbf{VV}}^{-1/2}\mathbf{\Pi}'\mathbf{Q}_{ZZ}\mathbf{\Pi}\Sigma_{\mathbf{VV}}^{-1/2'}/K_2 = \Sigma_{\mathbf{VV}}^{-1/2}\mathbf{C}'\mathbf{Q}_{ZZ}\mathbf{C}\Sigma_{\mathbf{VV}}^{-1/2'}/K_2. \tag{2.6}$$

The $n \times n$ matrix Λ_{K_2} is the expected value of the concentration parameter, divided by the number of instruments, K_2. Note that $\rho'\rho \leq 1$.

2.2. *k*-Class Statistics

The k-class estimator of β is

$$\hat{\beta}(k) = [\mathbf{Y}'(\mathbf{I} - k\mathbf{M}_Z)\mathbf{Y}]^{-1}[\mathbf{Y}'(\mathbf{I} - k\mathbf{M}_Z)\mathbf{y}], \tag{2.7}$$

where $\mathbf{M}_Z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and k is a scalar. The Wald statistic, based on the k-class estimator, testing the null hypothesis $\beta = \beta_0$ is

$$W(k) = \frac{[\hat{\beta}(k) - \beta_0]'[\mathbf{Y}'(I - k\mathbf{M}_Z)\mathbf{Y}][\hat{\beta}(k) - \beta_0]}{n\hat{\sigma}_{uu}(k)}, \tag{2.8}$$

where $\hat{\sigma}_{uu}(k) = \hat{\mathbf{u}}(k)'\hat{\mathbf{u}}(k)/(T - n)$ and $\hat{\mathbf{u}}(k) = \mathbf{y} - \mathbf{Y}\hat{\beta}(k)$.

Specific k-class estimators of interest include two-stage least squares (TSLS), the limited information maximum likelihood (LIML) estimator, Fuller's (1977) k-class estimator, and bias-adjusted TSLS (BTSLS; Nagar 1959; Rothenberg 1984). The values of k for these estimators are (cf. Donald and Newey 2001):

TSLS: $k = 1$, (2.9)

LIML: $k = \hat{k}_{\text{LIML}}$ is the smallest root of det $(\underline{\mathbf{Y}}'\underline{\mathbf{Y}} - k\underline{\mathbf{Y}}'\mathbf{M}_Z\underline{\mathbf{Y}}) = 0$, (2.10)

Fuller-k: $k = \hat{k}_{\text{LIML}} - c/(T - K_2)$, where c is a positive constant, (2.11)

BTSLS: $k = T/(T - K_2 + 2)$, (2.12)

where det($\mathbf{A}$) is the determinant of matrix $\mathbf{A}$.

2.3. Assumptions

We assume that the random variables are i.i.d. with four moments, the instruments are not multicollinear, and the errors are homoskedastic; that is, we assume:

Assumption A

(a) *There exists a constant* $D_1 > 0$ *such that* $mineval(\mathbf{Z}'\mathbf{Z}/T) \geq D_1$ *a.s. for all* K_2 *and for all* T *greater than some* T_0.

(b) $\mathbf{Z}_t$ *is i.i.d. with* $E\mathbf{Z}_t\mathbf{Z}_t' = \mathbf{Q}_{ZZ}$, *where* $\mathbf{Q}_{ZZ}$ *is positive definite, and* $EZ_{it}^4 \leq D_2 < \infty$, *where* $i = 1, \ldots, K_2$.

(c) $\eta_t = [u_t \mathbf{V}_t']'$ *is i.i.d. with* $E(\eta_t | \mathbf{Z}_t) = 0$, $E(\eta_t\eta_t' | \mathbf{Z}_t) = \Sigma$, *which is positive definite, and* $E(|\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}| \mid \mathbf{Z}_t) = E(|\eta_{it}\eta_{jt}\eta_{kt}\eta_{lt}|) \leq D_3 < \infty$, *where* $i, j, k, l = 1, \ldots, n+1$.

The next assumption is that the instruments are weak in the sense that the amount of information per instrument does not increase with the sample size, that is, the concentration parameter is proportional to the number of instruments. For fixed K_2, this assumption is achieved by considering the sequence of models in which $\mathbf{C} = \mathbf{\Pi}/\sqrt{T}$ is fixed, so that $\mathbf{\Pi}$ is modeled as local to zero (Staiger and Stock 1997). We adopt this nesting here, specifically:

Assumption B. $max_{i,j}|\mathbf{C}_{i,j}| \leq D_4 < \infty$, *where* D_4 *does not depend on* T *or* K_2, *and* $\mathbf{C}'\mathbf{C}/K_2 \to \mathbf{H}$ *as* $T \to \infty$, *where* $\mathbf{H}$ *is a fixed* $n \times n$ *matrix.*

Assumption B implies that $\mathbf{\Lambda}_{K_2} \to \mathbf{\Lambda}_\infty$ as $T \to \infty$, where $\mathbf{\Lambda}_\infty$ is a fixed matrix with maxeval$(\mathbf{\Lambda}_\infty) < \infty$. When the number of instruments is fixed, this assumption is equivalent to the weak-instrument Assumption L_Π in Staiger and Stock (1997).

Our analysis focuses on sequences of K_2 that, if they increase, do so slower than $\sqrt{T}$. Specifically, we assume:

Assumption C. $K_2^4/T \to 0$ *as* $T \to \infty$.

Note that Assumption C does not require K_2 to increase, but it limits the rate at which it can increase.

3. UNIFORM CONVERGENCE RESULT

This section provides the uniform convergence result (Theorem 3.1) that justifies the use of sequential asymptotics to compute the many weak instrument limiting representations. We adopt Phillips and Moon's (1999) notation in which $(T, K_2 \to \infty)_{seq}$ denotes the sequential limit in which first $T \to \infty$, then $K_2 \to \infty$; the notation $(K_2, T \to \infty)$ denotes the joint limit in which K_2 is implicitly indexed by T.

Lemma 6 of Phillips and Moon (1999) provides general conditions under which sequential convergence implies joint convergence.

Phillips and Moon (1999), Lemma 6

(a) Suppose there exist random vectors X_K and X on the same probability space as $X_{K,T}$ satisfying, for all K, $X_{K,T} \xrightarrow{p} X_K$ as $T \to \infty$ and $X_K \xrightarrow{p} X$ as $K \to \infty$. Then $X_{K,T} \xrightarrow{p} X$ as $(K, T \to \infty)$ if and only if

$$\limsup_{K,T} \Pr[\|X_{K,T} - X_K\| > \varepsilon] = 0 \text{ for all } \varepsilon > 0. \quad (3.1)$$

(b) Suppose there exist random vectors X_K such that, for any fixed K, $X_{K,T} \xrightarrow{d} X_K$ as $T \to \infty$ and $X_K \xrightarrow{d} X$ as $K \to \infty$. Then $X_{K,T} \xrightarrow{d} X$ as $(K, T \to \infty)$ if and only if, for all bounded continuous functions f,

$$\limsup_{K,T} |E[f(X_{K,T})] - E[f(X_K)]| = 0. \quad (3.2)$$

Note that condition (3.2) is equivalent to the requirement

$$\limsup_{K,T} \sup_x |F_{X_{K,T}}(x) - F_{X_K}(x)| = 0, \quad (3.3)$$

where $F_{X_{K,T}}$ is the c.d.f. of $X_{K,T}$ and F_{X_K} is the c.d.f. of X_K.

The rest of this section is devoted to showing that the conditions of this lemma, that is, (3.1) and (3.3), hold under assumptions A, B, and C for the statistics that enter the k-class estimators and test statistics. To do so, we use the following Berry–Esseen bound proven by Bertkus (1986):

Berry–Esseen Bound (Bertkus 1986). Let $\{X_1, \ldots, X_T\}$ be an i.i.d. sequence in $\mathbb{R}^K$ with zero means, a nonsingular second moment matrix, and finite absolute third moments. Let P_T be the probability measure associated with $T^{-1/2} \sum_{t=1}^{T} X_t$, and let P be the limiting Gaussian measure. Then for each T,

$$\begin{aligned} \sup_{A \in C^K} |P_T(A) - P(A)| &\leq \text{const} \times (K/T)^{1/2} E\|X\|^3 \\ &= O\left(\left[K_2^4/T\right]^{1/2}\right) \end{aligned} \quad (3.4)$$

where C^K is the class of all measurable convex sets in $\mathbb{R}^K$.

We now turn to k-class statistics. First note that, for fixed K_2, under Assumptions A and B, the weak law of large numbers and the central limit theorem imply that the following limits hold jointly for fixed K_2:

$$(T^{-1}\mathbf{u}'\mathbf{u}, T^{-1}\mathbf{V}'\mathbf{u}, T^{-1}\mathbf{V}'\mathbf{V}) \xrightarrow{p} (\sigma_{uu}, \Sigma_{\mathbf{V}u}, \Sigma_{\mathbf{VV}}), \quad (3.5)$$

$$\mathbf{\Pi}'\mathbf{Z}'\mathbf{Z}\mathbf{\Pi} \xrightarrow{p} \mathbf{C}'\mathbf{Q}_{\mathbf{ZZ}}\mathbf{C}, \quad (3.6)$$

$$(\mathbf{\Pi}'\mathbf{Z}'\mathbf{u}, \mathbf{\Pi}'\mathbf{Z}'\mathbf{V}) \xrightarrow{d} (\mathbf{C}'\Psi_{\mathbf{Z}u}, \mathbf{C}'\Psi_{\mathbf{ZV}}), \quad (3.7)$$

$$\begin{aligned} (\mathbf{u}'\mathbf{P_Z}\mathbf{u}, \mathbf{V}'\mathbf{P_Z}\mathbf{u}, \mathbf{V}'\mathbf{P_Z}\mathbf{V}) \xrightarrow{d} \big(&\Psi_{\mathbf{Z}u}' Q_{\mathbf{ZZ}}^{-1}\Psi_{\mathbf{Z}u}, \Psi_{\mathbf{ZV}}' Q_{\mathbf{ZZ}}^{-1}\Psi_{\mathbf{Z}u}, \\ &\Psi_{\mathbf{ZV}}' Q_{\mathbf{ZZ}}^{-1}\Psi_{\mathbf{ZV}}\big), \end{aligned} \quad (3.8)$$

where Ψ_{Zu} and Ψ_{ZV} are, respectively, $K_2 \times 1$ and $K_2 \times n$ random variables and $\Psi \equiv [\Psi'_{Zu}, \text{vec}(\Psi_{ZV})']'$ is distributed $N(0, \Sigma \otimes \mathbf{Q}_{ZZ})$.

The following theorem shows that the limits in (3.5)–(3.8) and related limits hold uniformly in K_2 under the sampling assumption (Assumption A), the weak instrument assumption (Assumption B), and the rate condition (Assumption C). Let $\|\mathbf{A}\| = [\text{tr}(\mathbf{A}'\mathbf{A})]^{1/2}$ denote the norm of the matrix $\mathbf{A}$ and, as in (3.3), let F_X denote the c.d.f. of the random variable X (etc.).

Theorem 3.1. *Under Assumptions A, B, and C,*

(a) $\lim\sup_{K_2,T} \Pr[\|(\mathbf{u}'\mathbf{u}/T, \mathbf{V}'\mathbf{u}/T, \mathbf{V}'\mathbf{V}/T) - (\sigma_{uu}, \Sigma_{\mathbf{V}u}, \Sigma_{\mathbf{VV}})\| > \varepsilon] = 0 \;\forall\, \varepsilon > 0,$

(b) $\lim\sup_{K_2,T} \Pr[\|\Pi'\mathbf{Z}'\mathbf{Z}\Pi\|/K_2 - \mathbf{C}'\mathbf{Q}_{ZZ}\mathbf{C}/K_2\| > \varepsilon] = 0 \;\forall\, \varepsilon > 0,$

(c) $\lim\sup_{K_2,T}\sup_x |F_{\Pi'\mathbf{Z}'\mathbf{u}}(\mathbf{x}) - F_{\mathbf{C}'\Psi_{Zu}}(x)| = 0,$

(d) $\lim\sup_{K_2,T}\sup_x |F_{\Pi'\mathbf{Z}'\mathbf{V}}(\mathbf{x}) - F_{\mathbf{C}'\Psi_{ZV}}(x)| = 0,$

(e) $\lim\sup_{K_2,T}\sup_x |F_{\mathbf{u}'P_Z\mathbf{u}}(x) - F_{\Psi'_{Zu}\mathbf{Q}_{ZZ}^{-1}\Psi_{Zu}}(x)| = 0,$

(f) $\lim\sup_{K_2,T}\sup_x |F_{\mathbf{V}'P_Z\mathbf{u}}(\mathbf{x}) - F_{\Psi'_{ZV}\mathbf{Q}_{ZZ}^{-1}\Psi_{Zu}}(\mathbf{x})| = 0,$

(g) $\lim\sup_{K_2,T}\sup_x |F_{\mathbf{V}'P_Z\mathbf{V}}(\mathbf{x}) - F_{\Psi'_{ZV}\mathbf{Q}_{ZZ}^{-1}\Psi_{ZV}}(\mathbf{x})| = 0.$

The proof of Theorem 3.1 is contained in the Appendix.

Theorem 3.1 verifies the conditions (3.1) and (3.3) of Phillips and Moon's (1999) Lemma 6 for statistics that enter the k-class estimator and Wald statistic. Some of these objects converge in probability uniformly under the stated assumptions (parts (a) and (b)), while others converge in distribution uniformly (parts (c)–(g)). It follows from the continuous mapping theorem that continuous functions of these objects also converge in probability (and/or distribution) uniformly under the stated assumptions. Because the k-class estimator $\hat{\beta}(k)$ and Wald statistic $W(k)$ are continuous functions of these statistics (after centering and scaling as needed), it follows that the $(K_2, T \to \infty)$ joint limit of these k-class statistics can be computed as the sequential limit $(T, K_2 \to \infty)_{\text{seq}}$.

4. MANY WEAK INSTRUMENT ASYMPTOTIC LIMITS

This section collects calculations of the many weak instrument asymptotic limits of k-class estimators and Wald statistics. These calculations are done using sequential asymptotics (justified by Theorem 3.1), in which the fixed-K_2 weak instrument asymptotic limits of Staiger and Stock (1997, Theorem 1) are analyzed as $K_2 \to \infty$. The limiting distributions differ depending on the limiting behavior of k. The main results are collected in Theorem 4.1, which is proven in the Appendix.

Theorem 4.1. *Suppose that Assumptions A, B, and C hold, and that $K_2 \to \infty$. Let* $\mathbf{x}$ *be an n-dimensional standard normal random variable. Then the following* *limits hold as* $(K_2, T \to \infty)$*:*

(a) *TSLS: If* $T(k-1)/K_2 \to 0$, *then*

$$\hat{\beta}(k) - \beta \xrightarrow{p} \sigma_{uu}^{1/2}\Sigma_{\mathbf{VV}}^{-1/2}(\Lambda_\infty + \mathbf{I}_n)^{-1}\rho \text{ and} \tag{4.1}$$

$$W(k)/K_2 \xrightarrow{p} \frac{\rho'(\Lambda_\infty + \mathbf{I}_n)^{-1}\rho}{n[1 - 2\rho'(\Lambda_\infty + \mathbf{I}_n)^{-1}\rho + \rho'(\Lambda_\infty + \mathbf{I}_n)^{-2}\rho]}. \tag{4.2}$$

(b) *BTSLS: If* $\sqrt{K_2}[T(k-1)/K_2 - 1] \to 0$ *and mineval* $(\Lambda_\infty) > 0$, *then*

$$\sqrt{K_2}(\hat{\beta}(k) - \beta) \xrightarrow{d} N(0, \sigma_{uu}\Sigma_{\mathbf{VV}}^{-1/2}\Lambda_\infty^{-1}(\Lambda_\infty + \mathbf{I}_n + \rho\rho')\Lambda_\infty^{-1}\Sigma_{\mathbf{VV}}^{-1/2\prime}) \text{ and} \tag{4.3}$$

$$W(k) \xrightarrow{d} \mathbf{x}'(\Lambda_\infty + \mathbf{I}_n + \rho\rho')^{1/2}\Lambda_\infty^{-1}(\Lambda_\infty + \mathbf{I}_n + \rho\rho')^{1/2\prime}\mathbf{x}/n. \tag{4.4}$$

(b) *LIML, Fuller-k: If* $T(k - k_{\text{LIML}})/\sqrt{K_2} \to 0$ *and* mineval$(\Lambda_\infty) > 0$, *then*

$$\sqrt{K_2}[T(k-1)/K_2 - 1] \xrightarrow{d} N(0, 2), \tag{4.5}$$

$$\sqrt{K_2}(\hat{\beta}(k) - \beta) \xrightarrow{d} N(0, \sigma_{uu}\Sigma_{\mathbf{VV}}^{-1/2}\Lambda_\infty^{-1}(\Lambda_\infty + \mathbf{I}_n - \rho\rho')\Lambda_\infty^{-1}\Sigma_{\mathbf{VV}}^{-1/2\prime}) \text{ and} \tag{4.6}$$

$$W(k) \xrightarrow{d} \mathbf{x}'(\Lambda_\infty + \mathbf{I}_n - \rho\rho')^{1/2}\Lambda_\infty^{-1}(\Lambda_\infty + \mathbf{I}_n - \rho\rho')^{1/2\prime}\mathbf{x}/n. \tag{4.7}$$

5. DISCUSSION

To simplify the proofs we have assumed i.i.d. sampling. Götze (1991) provides a Berry–Esseen bound for i.n.i.d. sampling. The bound in the i.n.i.d. case is const $\times (K_1^2/T)E\|X\|^3 = O([K_2^5/T]^{1/2})$, so the rate in Assumption C would be slower, $K_2^5/T \to 0$. With this slower rate, the results in Section 3 would extend to the case where the errors and instruments are independently but not necessarily identically distributed.

The many weak instrument representations in Theorem 4.1 for BTSLS, LIML, and the Fuller-k estimator rule out the partially identified and unidentified cases, for which mineval$(\Lambda_\infty) = 0$. This suggests that the approximations in Theorem 4.1, parts (b) and (c), might become inaccurate as Λ_{K_2} becomes nearly singular. The behavior of the many weak instrument approximations in partially identified and unidentified cases remain to be explored.

ACKNOWLEDGMENTS

We thank an anonymous referee for helpful suggestions that spurred this research, and Whitney Newey for pointing out an error in an earlier draft. This work was supported by NSF grant SBR-0214131.

APPENDIX

This appendix contains the proofs of Theorems 3.1 and 4.1. The proof of Theorem 3.1 uses the following lemma.

Lemma A.1. *Let* $\Delta_T = (\mathbf{Z}'\mathbf{Z}/T)^{-1} - \mathbf{Q}_{\mathbf{ZZ}}^{-1}$. *Under Assumptions A and C,*

(a) $\limsup_{K_2,T} \Pr[|T^{-1}\mathbf{u}'\mathbf{Z}\Delta_T\mathbf{Z}'\mathbf{u}| > \varepsilon] = 0 \; \forall \, \varepsilon > 0$,
(b) $\limsup_{K_2,T} \Pr[\|T^{-1}\mathbf{V}'\mathbf{Z}\Delta_T\mathbf{Z}'\mathbf{u}\| > \varepsilon] = 0 \; \forall \, \varepsilon > 0$,
(c) $\limsup_{K_2,T} \Pr[\|T^{-1}\mathbf{V}'\mathbf{Z}\Delta_T\mathbf{Z}'\mathbf{V}\| > \varepsilon] = 0 \; \forall \, \varepsilon > 0$.

Proof of Lemma A.1. The strategy for proving each part is first to show that the relevant quadratic form (for example, in (a), the quadratic form $T^{-1}\mathbf{u}'\mathbf{Z}\Delta_T\mathbf{Z}'\mathbf{u}$) has expected mean square that is bounded by const $\times$ (K_2^2/T), and then to apply Chebychev's inequality and the condition in Assumption C that $K_2^2/T \to 0$. The details of these calculations are tedious and are omitted; they can be found in an earlier working paper (Stock and Yogo 2002, Lemma A.2).

Proof of Theorem 3.1. (a) This follows from the weak law of large numbers because $(\mathbf{u}'\mathbf{u}/T, \mathbf{V}'\mathbf{u}/T, \mathbf{V}'\mathbf{V}/T)$ do not depend on K_2.

(b) Note that $E[\mathbf{\Pi}'\mathbf{Z}'\mathbf{Z}\mathbf{\Pi}/K_2 - \mathbf{C}'\mathbf{Q}_{\mathbf{ZZ}}\mathbf{C}/K_2] = 0$. The (1,1) element of this matrix is

$$\begin{aligned}&(\mathbf{\Pi}'\mathbf{Z}'\mathbf{Z}\mathbf{\Pi} - \mathbf{C}'\mathbf{Q}_{\mathbf{ZZ}}\mathbf{C})_{1,1}/K_2\\ &\quad = (TK_2)^{-1}\sum_{t=1}^{T}\sum_{i=1}^{K_2}\sum_{j=1}^{K_2} C_{i1}C_{j1}(Z_{it}Z_{jt} - q_{ij}),\end{aligned}$$

where q_{ij} is the (i, j) element of $\mathbf{Q}_{\mathbf{ZZ}}$. Because $\mathbf{Z}_t$ is i.i.d. (Assumption A(b)) and the elements of $\mathbf{C}$ are bounded (Assumption B), the expected value of the square of this element is

$$\begin{aligned}&E\{[(\mathbf{\Pi}'\mathbf{Z}'\mathbf{Z}\mathbf{\Pi} - \mathbf{C}'\mathbf{Q}_{\mathbf{ZZ}}\mathbf{C})_{1,1}/K_2]^2\}\\ &\quad = E\left[\frac{1}{TK_2}\sum_{t=1}^{T}\sum_{i=1}^{K_2}\sum_{j=1}^{K_2} C_{i1}C_{j1}(Z_{it}Z_{jt} - q_{ij})\right]^2\\ &\quad = \frac{1}{TK_2^2}\sum_{i=1}^{K_2}\sum_{j=1}^{K_2}\sum_{k=1}^{K_2}\sum_{l=1}^{K_2} C_{i1}C_{j1}C_{k1}C_{l1}E[(Z_{it}Z_{jt} - q_{ij})(Z_{kt}Z_{lt} - q_{kl})]\\ &\quad \le \text{const} \times \frac{K_2^2}{T} \times \left(\frac{1}{K_2}\sum_{i=1}^{K_2}|C_{i1}|\right)^4 \le \text{const} \times \frac{K_2^2}{T}.\end{aligned}$$

By the same argument applied to the (1,1) element, the remaining elements of $\mathbf{\Pi}'\mathbf{Z}'\mathbf{Z}\mathbf{\Pi}/K_2 - \mathbf{C}'\mathbf{Q}_{\mathbf{ZZ}}\mathbf{C}/K_2$ are also bounded in mean square by const $\times$ (K_2^2/T). The matrix $\mathbf{\Pi}'\mathbf{Z}'\mathbf{Z}\mathbf{\Pi}/K_2$ is $n \times n$ and so the number of elements does not depend on K_2, and the result (b) follows by Chebychev's inequality and noting that, under Assumption C, $K_2^2/T \to 0$.

(c) Under Assumption B, $\mathbf{\Pi}'\mathbf{Z}'\mathbf{u} = T^{-1/2}\mathbf{C}'\mathbf{Z}'\mathbf{u} = \mathbf{C}'(T^{-1/2}\sum_{t=1}^{T} \mathbf{Z}_t u_t)$. Let P_T denote the probability measure associated with $T^{-1/2}\mathbf{Z}'\mathbf{u}$ and let P denote the limiting probability measure associated with Ψ_{Zu}. Define the convex set $A(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^{K_2}\colon \mathbf{C}'\mathbf{y} \leq \mathbf{x}\}$, so that $P_T(A(\mathbf{x})) = F_{\Pi'Z'u}(\mathbf{x})$ and $P(A(\mathbf{x})) = F_{C'\Psi_{Zu}}(x)$. By Assumption A, $\mathbf{Z}_t u_t$ is an i.i.d., mean zero K_2-dimensional random variable with finite third moments, so the Berry–Esseen bound (3.4) applies and $\sup_x |F_{\Pi'Z'u}(\mathbf{x}) - F_{C'\Psi_{Zu}}(x)| \leq \text{const} \times \sqrt{K_2^4/T}$. The result (c) follows from Assumption C. We note that this line of argument is used in Jensen and Mayer (1975).

(d) The proof is the same as for (c).

(e) Write $\mathbf{u}'\mathbf{P_Z}\mathbf{u} = (T^{-1/2}\mathbf{u}'\mathbf{Z})(T^{-1}\mathbf{Z}'\mathbf{Z})(T^{-1/2}\mathbf{Z}'\mathbf{u}) = \xi_1 + \xi_2$, where $\xi_1 = (T^{-1/2}\mathbf{u}'\mathbf{Z})\mathbf{Q}_{\mathbf{ZZ}}^{-1}(T^{-1/2}\mathbf{Z}'\mathbf{u})$ and $\xi_2 = (T^{-1/2}\mathbf{u}'\mathbf{Z})\Delta_T(T^{-1/2}\mathbf{Z}'\mathbf{u})$. As in the proof of (c), let P_T denote the probability measure associated with $T^{-1/2}\mathbf{Z}'\mathbf{u}$ and let P denote the limiting probability measure of Ψ_{Zu}. Let $B(x)$ be the convex set, $B(x) = \{\mathbf{y} \in \mathbb{R}^{K_2}\colon \mathbf{y}'Q_{\mathbf{ZZ}}^{-1}\mathbf{y} \leq x\}$, so that $P_T(B(x)) = F_{\xi_1}(x)$ and $P(B(x)) = F_{\Psi'_{Zu}Q_{ZZ}^{-1}\Psi_{Zu}}(x)$. It follows from (3.4) that $\sup_x |F_{\xi_1}(x) - F_{\Psi'_{Zu}Q_{ZZ}^{-1}\Psi_{Zu}}(x)| \leq \text{const} \times \sqrt{K_2^4/T}$. By Lemma A.1(a), $\xi_2 \xrightarrow{p} 0$ uniformly as $(K_2, T \to \infty)$, and the result (e) follows.

(f) and (g). The dimensions of $\mathbf{V}'\mathbf{P_Z}\mathbf{u}$ and $\mathbf{V}'\mathbf{P_Z}\mathbf{V}$ do not depend on K_2, and the proofs of (f) and (g) are similar to that of (e).

Proof of Theorem 4.1. We first state the fixed-K_2 weak instrument asymptotic representations of the k-class estimators. Define the $K_2 \times 1$ and $K_2 \times n$ random variables $\mathbf{z}_u = \mathbf{Q}_{\mathbf{ZZ}}^{-1/2\prime}\Psi_{Zu}\sigma_{\mathbf{uu}}^{-1/2}$ and $\mathbf{z_V} = \mathbf{Q}_{\mathbf{ZZ}}^{-1/2\prime}\Psi_{ZV}\Sigma_{\mathbf{VV}}^{-1/2}$ (Ψ_{Zu} and Ψ_{ZV} are defined following (3.8)), so that

$$\begin{pmatrix} \mathbf{z}_u \\ \text{vec}(\mathbf{z_V}) \end{pmatrix} \sim N(\mathbf{0}, \bar{\mathbf{\Sigma}} \otimes \mathbf{I}_{K_2}), \text{ where } \bar{\mathbf{\Sigma}} = \begin{bmatrix} 1 & \rho' \\ \rho & \mathbf{I}_n \end{bmatrix}. \tag{A.1}$$

Also let

$$\nu_1 = (\boldsymbol{\lambda} + \mathbf{z_V})'(\boldsymbol{\lambda} + \mathbf{z_V}) \text{ and} \tag{A.2}$$

$$\nu_2 = (\boldsymbol{\lambda} + \mathbf{z_V})'\mathbf{z}_u, \tag{A.3}$$

where $\lambda = \mathbf{Q}_{\mathbf{ZZ}}^{1/2}\mathbf{C}\Sigma_{\mathbf{VV}}^{-1/2}$. Then under Assumptions A and B, with fixed K_2,

$$\hat{\beta}(k) - \beta \xrightarrow{d} \sigma_{uu}^{1/2}\Sigma_{\mathbf{VV}}^{-1/2}(\nu_1 - \kappa\mathbf{I}_n)^{-1}(\nu_2 - \kappa\rho) \text{ and} \tag{A.4}$$

$$W(k) \xrightarrow{d} \frac{(\nu_2 - \kappa\rho)'(\nu_1 - \kappa\mathbf{I}_n)^{-1}(\nu_2 - \kappa\rho)}{n[1 - 2\rho'(\nu_1 - \kappa\mathbf{I}_n)^{-1}(\nu_2 - \kappa\rho) + (\nu_2 - \kappa\rho)'(\nu_1 - \kappa\mathbf{I}_n)^{-2}(\nu_2 - \kappa\rho)]}, \tag{A.5}$$

where (A.5) holds under the null hypothesis $\beta = \beta_0$. The representations (A.4) and (A.5) follow from Staiger and Stock (1997, Theorem 1) because Assumptions A and B imply Staiger and Stock's Assumptions M and L_Π when K_2 is fixed.

The following limits hold jointly as $K_2 \to \infty$:

$$\nu_1/K_2 \xrightarrow{p} \Lambda_\infty + \mathbf{I}_n, \tag{A.6}$$

$$\nu_2/K_2 \xrightarrow{p} \rho, \tag{A.7}$$

$$\begin{pmatrix} \dfrac{\mathbf{z}_u'\mathbf{z}_u - K_2}{\sqrt{K_2}} \\ \dfrac{\lambda'\mathbf{z}_u}{\sqrt{K_2}} \\ \dfrac{\mathbf{z}_\mathbf{V}'\mathbf{z}_u - K_2\rho}{\sqrt{K_2}} \end{pmatrix} \xrightarrow{d} N(0, B), \quad \text{where} B = \begin{bmatrix} 2 & 0 & 2\rho' \\ 0 & \Lambda_\infty & 0 \\ 2\rho & 0 & \mathbf{I}_n + \rho\rho' \end{bmatrix}, \tag{A.8}$$

$$(\nu_2 - K_2\rho)/\sqrt{K_2} \to N(0, \Lambda_\infty + \mathbf{I}_n + \rho\rho'). \tag{A.9}$$

The results (A.6)–(A.9) follow by straightforward calculations using the central limit theorem, the weak law of large numbers, and the joint normal distribution of $\mathbf{z}_u$ and $\mathbf{z}_\mathbf{V}$ in (A.1).

We now turn to the proof of Theorem 4.1.

(a) From (A.4), the fixed-K_2 weak instrument approximation to the distribution of the TSLS estimator is $\hat{\beta}^{\text{TSLS}} - \beta \sim \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}\nu_1^{-1}\nu_2 = \sigma_{uu}^{1/2}\Sigma_{VV}^{-1/2}$ $(\nu_1/K_2)^{-1}(\nu_2/K_2)$. The limit stated in the theorem for the estimator follows by substituting (A.6) and (A.7) into this expression. The many weak instrument limit for the TSLS Wald statistic follows by rewriting (A.5) as

$$W^{\text{TSLS}}/K_2 \sim \frac{(\nu_2/K_2)'(\nu_1/K_2)^{-1}(\nu_2/K_2)}{n[1 - 2\rho'(\nu_1/K_2)^{-1}(\nu_2/K_2) + (\nu_2/K_2)'(\nu_1/K_2)^{-2}(\nu_2/K_2)]}$$

and applying (A.6) and (A.7).

(b) The fixed-K_2 weak instrument approximation to the distribution of a k-class estimator, given in (A.4), in general can be written as

$$\sqrt{K_2}[\hat{\beta}(k) - \beta] \sim \sigma_{uu}^{1/2}\Sigma_{\mathbf{VV}}^{-1/2}\left[\frac{\nu_1 - K_2\mathbf{I}_n}{K_2} - \frac{1}{\sqrt{K_2}}\left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\mathbf{I}_n\right]^{-1} \times \left[\frac{\nu_2 - K_2\rho}{\sqrt{K_2}} - \left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\rho\right], \tag{A.10}$$

where $T(k-1) \xrightarrow{d} \kappa$ for fixed K_2. The assumption $\sqrt{K_2}[T(k-1)/K_2 - 1] \to 0$ implies that $(\kappa - K_2)/\sqrt{K_2} \to 0$, so by (A.6) and (A.9) we have, as $K_2 \to \infty$,

$$\frac{\nu_1 - K_2\mathbf{I}_n}{K_2} - \frac{1}{\sqrt{K_2}}\left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\mathbf{I}_n \xrightarrow{p} \Lambda_\infty \text{ and}$$

$$\frac{\nu_2 - K_2\rho}{\sqrt{K_2}} - \left(\frac{\kappa - K_2}{\sqrt{K_2}}\right)\rho \xrightarrow{d} N(0, \Lambda_\infty + \mathbf{I}_n + \rho\rho'),$$

and the result (4.3) follows. The assumption mineval$(\Lambda_\infty) > 0$ is used to ensure the invertibility of Λ_∞. The distribution of the Wald statistic follows.

(c) For fixed K_2, $T(k_{\text{LIML}} - 1) \xrightarrow{d} \kappa^*$. We show below that, as $K_2 \to \infty$,

$$\frac{\kappa^* - K_2}{\sqrt{K_2}} = \frac{\mathbf{z}_u'\mathbf{z}_u - K_2}{\sqrt{K_2}} + o_p(1). \tag{A.11}$$

The result (4.5) follows from (A.11) and (A.8). Moreover, applying (A.6), (A.8), (A.9), and (A.11) yields

$$\frac{\nu_1 - K_2\mathbf{I}_n}{K_2} - \frac{1}{\sqrt{K_2}}\left(\frac{\kappa^* - K_2}{\sqrt{K_2}}\right)\mathbf{I}_n \xrightarrow{p} \Lambda_\infty \text{ and}$$

$$\frac{\nu_2 - K_2\rho}{\sqrt{K_2}} - \left(\frac{\kappa^* - K_2}{\sqrt{K_2}}\right)\rho = \frac{\lambda'\mathbf{z}_u}{\sqrt{K_2}} + \frac{z_V'\mathbf{z}_u - K_2\rho}{\sqrt{K_2}}$$
$$- \left(\frac{\mathbf{z}_u'\mathbf{z}_u - K_2}{\sqrt{K_2}}\right)\rho + o_p(1) \xrightarrow{d} N(0, \Lambda_\infty + \mathbf{I}_n - \rho\rho'),$$

where Λ_∞ is invertible by the assumption mineval$(\Lambda_\infty) > 0$. The result (4.6) follows, as does the distribution of the Wald statistic.

It remains to show (A.11). From (2.11), κ^* is the smallest root of

$$0 = \det\left[\begin{pmatrix} \mathbf{z}_u'\mathbf{z}_u & \nu_2' \\ \nu_2 & \nu_1 \end{pmatrix} - \kappa^*\begin{pmatrix} 1 & \rho' \\ \rho & \mathbf{I}_n \end{pmatrix}\right]. \tag{A.12}$$

Let $\phi = (\kappa^* - K_2)/\sqrt{K_2}$, $a = (\mathbf{z}_u'\mathbf{z}_u - K_2)/\sqrt{K_2}$, $\mathbf{b} = (\nu_2 - K_2\rho)/\sqrt{K_2}$, and $\mathbf{L} = (\nu_1 - K_2\mathbf{I}_n)/K_2$. Then (A.12) can be rewritten so that ϕ is the smallest root of

$$0 = \det\begin{bmatrix} a - \phi & (\mathbf{b} - \phi\rho)' \\ \mathbf{b} - \phi\rho & \sqrt{K_2}\mathbf{L} - \phi\mathbf{I}_n \end{bmatrix}. \tag{A.13}$$

We first show that $K_2^{-1/4}\phi \xrightarrow{p} 0$. Let $\tilde{\phi} = K_2^{-1/4}\phi$. By (A.6), (A.8), and (A.9), $K_2^{-1/4}a \xrightarrow{p} 0$, $K_2^{-1/4}\mathbf{b} \xrightarrow{p} 0$, and $\mathbf{L} \xrightarrow{p} \Lambda_\infty$. By the continuity of the determinant, it follows that in the limit $K_2 \to \infty$, $\tilde{\phi}$ is the smallest root of the equation

$$0 = \det\begin{bmatrix} \tilde{\phi} & \tilde{\phi}\rho' \\ \tilde{\phi}\rho & \tilde{\phi}\mathbf{I}_n + O_p(K_2^{1/4}) \end{bmatrix}, \tag{A.14}$$

from which it follows that $\tilde{\phi} = K_2^{-1/4}\phi \xrightarrow{p} 0$.

To obtain (A.11), write the determinantal equation (A.13) as

$$\begin{aligned} 0 &= [(a - \phi) - (\mathbf{b} - \phi\rho)'(K_2^{1/2}\mathbf{L} - \phi\mathbf{I}_n)^{-1}(\mathbf{b} - \phi\rho)]\det(K_2^{1/2}\mathbf{L} - \phi\mathbf{I}_n) \\ &= K_2^{n/2}\{(a - \phi) - [K_2^{-1/4}(\mathbf{b} - \phi\rho)]'(\mathbf{L} - K_2^{-1/2}\phi\mathbf{I}_n)^{-1} \\ &\quad \times [K_2^{-1/4}(\mathbf{b} - \phi\rho)]\det(\mathbf{L} - K_2^{-1/2}\phi\mathbf{I}_n) \\ &= K_2^{n/2}\{[(a - \phi)]\det(\Lambda_\infty) + o_p(1)\}, \end{aligned} \tag{A.15}$$

where the final equality follows from $K_2^{-1/4}\mathbf{b} \xrightarrow{p} 0$, $\mathbf{L} \xrightarrow{p} \Lambda_\infty$, $K_2^{-1/4}\phi \xrightarrow{p} 0$, and $\det(\Lambda_\infty) > 0$. By the continuity of the solution to (A.13), it follows that $\phi = a + o_p(1)$, which, in the original notation, is (A.11).

References

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

Bekker, P. A. (1994), "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica*, 62, 657–81.

Bertkus, V. Y. (1986), "Dependence of the Berry–Esseen Estimate on the Dimension," *Litovsk. Mat. Sb.*, 26, 205–10.

Chao, J. C., and N. R. Swanson (2002), "Consistent Estimation with a Large Number of Weak Instruments," unpublished manuscript, University of Maryland.

Donald, S. G., and W. K. Newey (2001), "Choosing the Number of Instruments," *Econometrica*, 69, 1161–91.

Fuller, W. A. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–53.

Götze, F. (1991), "On the Rate of Convergence in the Multivariate CLT," *Annals of Probability*, 19, 724–39.

Jensen, D. R., and L. S. Mayer (1975), "Normal-Theory Approximations to Tests of Linear Hypotheses," *Annals of Statistics*, 3, 429–44.

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.

Morimune, K. (1983), "Approximate Distributions of k-Class Estimators when the Degree of Overidentifiability is Large Compared with the Sample Size," *Econometrica*, 51, 821–41.

Nagar, A. L. (1959), "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575–95.

Phillips, P. C. B., and H. R. Moon (1999), "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica*, 67, 1057–111.

Rothenberg, T. J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," Chapter 15 in *Handbook of Econometrics*, Vol. II (ed. by Z. Griliches and M. D. Intriligator), Amsterdam: North Holland, pp. 881–935.

Sargan, D. (1975), "Asymptotic Theory and Large Models," *International Economic Review*, 16, 75–91.

Staiger, D., and J. H. Stock (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–86.

Stock, J. H., and M. Yogo (2002), "Testing for Weak Instruments in Linear IV Regression," NBER Technical Working Paper 284.

Stock, J. H., and M. Yogo (2004), "Testing for Weak Instruments in Linear IV Regression," Chapter 5 in this volume.