Econometric Theory, **26**, 2010, 633–645.
doi:10.1017/S0266466609100117

MANY INSTRUMENTS ASYMPTOTIC APPROXIMATIONS UNDER NONNORMAL ERROR DISTRIBUTIONS

MARTIJN VAN HASSELT
University of Western Ontario

In this paper we derive an alternative asymptotic approximation to the sampling distribution of the limited information maximum likelihood estimator and a bias-corrected version of the two-stage least squares estimator. The approximation is obtained by allowing the number of instruments and the concentration parameter to grow at the same rate as the sample size. More specifically, we allow for potentially nonnormal error distributions and obtain the conventional asymptotic distribution and the results of Bekker (1994, *Econometrica* 62, 657–681) and Bekker and Van der Ploeg (2005, *Statistica Neerlandica* 59, 139–267) as special cases. The results show that when the error distribution is not normal, in general both the properties of the instruments and the third and fourth moments of the errors affect the asymptotic variance. We compare our findings with those in the recent literature on many and weak instruments.

1. INTRODUCTION

In this paper we derive the asymptotic distributions of the limited information maximum likelihood (LIML) estimator and a bias-corrected version of two-stage least squares (2SLS) when the number of instruments and the concentration parameter grow at the same rate as the sample size. The focus is on these estimators because in this asymptotic sequence both are consistent. We build on the framework introduced by Morimune (1983) and Bekker (1994) but do not assume that the errors in the model are Gaussian.[1] We use a martingale difference representation of the estimators due to Chao and Swanson (2004, 2006) and formulate conditions on the errors and the sequence of instruments to obtain the limit distributions. It turns out that third and fourth moments and functions of the instruments and reduced form coefficients affect the asymptotic variance, unless the error distribution satisfies certain conditions.

I thank Paul Bekker, John Knight, Frank Kleibergen, and Tony Lancaster for their support and suggestions. The co-editor and two anonymous referees provided valuable comments that greatly improved the presentation in this paper. I am responsible for any remaining errors. Address correspondence to Martijn van Hasselt, Department of Economics, University of Western Ontario, Social Science Centre, room 4033, London, Ontario, Canada N6A 5C2; e-mail: mvanhass@uwo.ca.

© Cambridge University Press, 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

Since the influential contribution of Bekker (1994), deriving the properties of instrumental variables (IV) estimators and test statistics under alternative asymptotic schemes has been an area of active research. Our paper is best seen as complementing this literature. The asymptotic sequence we use can be described as many *strong* instruments because the concentration parameter grows at the same rate as the sample size. In contrast, Stock and Yogo (2005) and Chao and Swanson (2004, 2006) derive limit distributions in a many *weak* instruments setting, where this parameter (and potentially the number of instruments) grows at a slower rate. The consistency of LIML and (bias-corrected) 2SLS crucially depends on the growth rates of the number of instruments, the concentration parameter, and the sample size (see Chao and Swanson, 2005). Moreover, the limit distributions with many weak instruments are different from the ones we derive in this paper.

Two recent papers are closely related to our work. Under a many strong instruments sequence Anderson and Kunitomo (2006) and Hansen, Hausman, and Newey (2006) find the same limit distribution for LIML as the one in this paper. The differences lie in the setup and assumptions used. Although Anderson and Kunitomo (2006) allow for stochastic instruments, they require the errors to have bounded sixth moments and directly impose a Lindeberg-type condition. Hansen et al. (2006) allow for a potentially nonlinear reduced form and weaker instruments. Their method of proof relies on boundedness of the eighth moments and a condition on the rate at which the reduced form can be approximated by a linear combination of the instruments.

Though the setup in the aforementioned papers is more general, our paper can be viewed as giving an alternative derivation, under different assumptions, of various limit distributions. Our approach combines the martingale difference construction of Chao and Swanson (2004, 2006) and the finite-sample results of Bekker (1994, Lem. 2.1) and Bekker and Van der Ploeg (2005, Lem. 1). A convenient feature of our formulation is that, under some assumptions, the results apply to any estimator that can be expressed as a combination of quadratic and bilinear forms in symmetric matrices. Unlike Hansen et al. (2006) and Anderson and Kunitomo (2006) our paper also analyzes a bias-corrected 2SLS estimator. We compare the estimator with LIML and consider the effects of nonnormality on the asymptotic variance.

The remainder of this paper is organized as follows. Section 2 presents a result on the limit distribution of a certain quadratic form. This result is then applied in Section 3 to derive the many instruments asymptotic distributions of LIML and bias-corrected 2SLS. Section 4 concludes. Proofs are collected in the Appendix.

2. AN AUXILIARY RESULT

We first discuss a result concerning a quadratic form in a symmetric matrix. Following Bekker (1994), let $U = M + V$ be a $T \times J$ matrix, where M contains constants and the rows v_t' of V are independent and identically distributed (i.i.d.) with zero mean and covariance matrix $\Omega = (\omega_{ij})$. The columns of U, M, and V

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

are denoted by u_j, m_j, and v_j, respectively ($j = 1, \ldots, J$). Let $C = (c_{st})$ be a symmetric $T \times T$ matrix; though not explicit in notation, we allow individual elements c_{st} to depend on T. Finally, a is an arbitrary $J \times 1$ vector and d_C the diagonal of C ($d_{C,t} = c_{tt}$). The following conditions are imposed on the matrices U, M, V, and C.

Assumption 1.

(a) $\{v_t'; 1 \le t \le T\}$ is i.i.d. with mean zero, positive definite covariance matrix Ω, and finite fourth moments;

(b) $\sup_{t \ge 1} |m_{tj}| = D_j < \infty$, $j = 1, \ldots, J$;

(c) the sequence of C and M is such that, as $T \to \infty$,

$$\frac{1}{T}(CM)'(CM) \to Q_{CM} > 0, \qquad \frac{1}{T}M'Cd_C \to \mu_{CM},$$

$$\frac{\operatorname{tr}(C)}{T} = \tau_C + o(T^{-1/2}), \qquad \frac{\operatorname{tr}(C^2)}{T} \to \tau_{C^2}, \qquad \frac{d_C'd_C}{T} \to \delta_C,$$

$$\sup_{T \ge 1} \sup_{1 \le t \le T} \sum_{s=1}^{T} |c_{st}| = D_C < \infty.$$

Because $\mathrm{E}(u_j) = m_j$, part (b) states that the mean is finite. Part (c) imposes regularity conditions on the sequence of matrices C and M. In particular, it ensures uniform integrability of a certain martingale difference array and a well-defined asymptotic variance. Bekker (1994) derives the limit distribution of $U'CUa$ when C is a projection matrix and v_t is normally distributed. The next result shows that asymptotic normality can still be achieved when C is merely symmetric and the errors are nonnormal. The proof is given in the Appendix.

THEOREM 1. *If Assumption 1 holds, then* $T^{-1/2}[U'CUa - \mathrm{E}(U'CUa)] \xrightarrow{d} N(0, W)$, *where*

$$\begin{aligned} W = {} & \Omega a' Q_{CM} a + a' \Omega a Q_{CM} + \Omega a a' Q_{CM} + Q_{CM} a a' \Omega \\ & + \tau_{C^2}(a' \Omega a \Omega + \Omega a a' \Omega) + \delta_C \left\{ \mathrm{E}[(a'v_t)^2 v_t v_t'] - a' \Omega a \Omega - 2 \Omega a a' \Omega \right\} \\ & + 2a' \mu_{CM} \mathrm{E}[(a'v_t) v_t v_t'] + \mu_{CM} \mathrm{E}[(a'v_t)^2 v_t'] + \mathrm{E}[(a'v_t)^2 v_t] \mu_{CM}'. \end{aligned}$$

Under normality of v_t, *the asymptotic variance simplifies to*

$$W = \Omega a' Q_{CM} a + a' \Omega a Q_{CM} + \Omega a a' Q_{CM} + Q_{CM} a a' \Omega + \tau_{C^2} \left\{ a' \Omega a \Omega + \Omega a a' \Omega \right\}.$$

3. LIML AND BIAS-CORRECTED 2SLS

Consider the following model:

$$\begin{aligned} y &= X\beta_0 + u, \\ X &= Z\Pi + V_2, \end{aligned} \tag{1}$$

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

where y and u are $T \times 1$, X is $T \times G$, Z is a $T \times K$ matrix of instruments, and Π is a $K \times G$ matrix of reduced form coefficients. We assume that the instruments are nonstochastic; if the instruments are random, all subsequent results should be read as conditional on Z. Note that the variables in X are endogenous when u and V_2 are correlated. The reduced form for the complete model is $(y, X) = Z\Pi(\beta_0, I_G) + (v_1, V_2)$, where $v_1 = u + V_2\beta_0$ and I_G is the G-dimensional identity matrix. Let (v_{t1}, v_{t2}') denote the tth row of (v_1, V_2). Regarding the instruments and errors we make the following assumptions.

Assumption 2 (Instruments).

(a) $rank\ (Z) = K$;

(b) $rank\ (\Pi) = G$.

Assumption 3 (Errors).

(a) $\left\{\left(v_{t1}, v_{t2}'\right); 1 \le t \le T\right\}$ is i.i.d. with mean zero and positive definite covariance matrix Ω;

(b) the fourth moments of (v_{t1}, v_{t2}') exist and are finite.

Assumption 3 implies that $\mathrm{E}(X)$ is linear in Z; it is possible to allow for nonlinearity along the lines of Bekker (1994, p. 659) or Hansen et al. (2006). Also, the errors are homoskedastic. With heteroskedastic errors and many strong instruments LIML is no longer consistent. In this context Bekker and Van der Ploeg (2005) and Hausman, Newey, Woutersen, Chao, and Swanson (2007) propose consistent alternatives. Part (b) of Assumption 3 is necessary because, in contrast to the standard asymptotic distribution, moments up to order 4 will appear in the many instruments asymptotic variance.

To apply Theorem 1 we impose some restrictions, equivalent to those in Assumption 1(b) and (c), on the limiting behavior of Z and Π as $T \to \infty$.[2] Let z_t' be the tth row of Z and let $P_Z = Z(Z'Z)^{-1}Z'$ be the usual projection with typical element p_{st} (which depends on T and K). The diagonal of P_Z is the vector d_{P_Z} and let ι_T denote the $T \times 1$ vector containing ones.

Assumption 4 (Sequence). As $T \to \infty$:

(a) $K/T = \alpha + o(T^{-1/2})$ for some $\alpha \in [0, 1)$;

(b) $\sup_{t \ge 1} \left\| z_t'\Pi \right\| < \infty$;

(c) Z and Π are such that

$$\frac{1}{T}\Pi'Z'Z\Pi \to Q_{Z\Pi} > 0, \qquad \frac{1}{T}\Pi'Z'\iota_T \to \mu_{Z\Pi}, \qquad \frac{1}{K}\Pi'Z'd_{P_Z} \to \tilde{\mu}_{Z\Pi}, \tag{2}$$

$$\frac{d_{P_Z}'d_{P_Z}}{T} \to \gamma\,\alpha, \qquad \alpha \le \gamma \le 1, \tag{3}$$

$$\sup_{T \ge K} \sup_{1 \le t \le T} \sum_{s=1}^{T} |p_{st}| = D_{P_Z} < \infty. \tag{4}$$

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

Part (a) is equivalent to convergence of $\text{tr}(C)/T$ in the previous section. The number of instruments K is allowed to grow at either the same rate ($\alpha > 0$) or more slowly ($\alpha = 0$) than the sample size T. The condition on $\text{E}(X)$ in part (b) is similar to Assumption 1(b) and implicitly imposes some structure on Z and Π; their joint behavior as $T, K \to \infty$ is such that the elements of $Z\Pi$ remain finite. Loosely speaking, (2) in part (c) assumes that for $Z\Pi$ the average, the weighted average,[3] and the average of the squares converge.[4] Therefore, the concentration parameter $\Omega_{22}^{-1/2}\Pi' Z' Z \Pi \Omega_{22}^{-1/2}$ grows at the same rate as the sample size, and the instruments are stronger than in Chao and Swanson (2006) or Stock and Yogo (2005).[5] Condition (3) is equivalent to convergence of $d_C' d_C / T$ in the previous section. The bounds on γ follow from the Cauchy–Schwarz inequality and the fact that P_Z is a projection matrix:

$$\left(\frac{K}{T}\right)^2 \leq \frac{d_P' d_P}{T} \leq \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} p_{st}^2 = \frac{K}{T}.$$

The condition in (4) corresponds to the absolute summability of C in the previous section. As mentioned before, this is a uniform integrability requirement; it implies that a Lindeberg-type condition holds (see the Appendix for more discussion). Anderson and Kunitomo (2006) directly impose a Lindeberg condition, but this comes at the cost of needing finite sixth (rather than fourth) error moments. Hansen et al. (2006) do not impose $\Pi' Z' Z \Pi = O(T)$ or condition (4). Instead, they use a restriction on $\|Z\|$ and assume that the eighth error moments are finite. Because P_Z is a projection matrix, its rows are square summable. Equation (4) strengthens this to absolute summability, which in turn allows for weaker restrictions on the distribution of (v_{t1}, v_{t2}'). The summability condition on P_Z is satisfied in a finite sample, and we assume that the asymptotic sequence is such that the same is true asymptotically. In the special case of "group asymptotics" (see Bekker and van der Ploeg, 2005) condition (4) can be verified directly; if Z is a matrix of dummy instruments, dividing each observation into one of K mutually exclusive categories, then $\sum_{s=1}^{T} |p_{st}| = 1$ for all t.

Partition Ω as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{21}' \\ \omega_{21} & \Omega_{22} \end{bmatrix},$$

where ω_{11} is a scalar, ω_{21} is $G \times 1$, and Ω_{22} is $G \times G$. The variance of u_t is $\sigma_u^2 = (1, -\beta_0')\Omega(1, -\beta_0')'$, and endogeneity of X is characterized by $\sigma_{uv_2} = \text{E}(u_t v_{t2}) = \omega_{21} - \Omega_{22}\beta_0$. Letting $M_Z = I - P_Z$, the LIML estimator is given by

$$\hat{\beta}_L = \arg\min_{\beta} \frac{(y - X\beta) P_Z (y - X\beta)}{(y - X\beta) M_Z (y - X\beta)}. \tag{5}$$

Bekker (1994) shows that when $\alpha > 0$ LIML is consistent but standard 2SLS is not.[6] Alternatively, a simple bias-corrected 2SLS estimator is given by

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

$$\hat{\beta}_B = \left(X' \left(P_Z - \frac{K}{T} I_T \right) X \right)^{-1} X' \left(P_Z - \frac{K}{T} I_T \right) y. \tag{6}$$

Under Assumption 4, $\hat{\beta}_B$ is consistent and asymptotically equivalent to the estimator analyzed in Stock and Yogo (2005). Interestingly, the bias-corrected estimator originally proposed by Nagar (1959, p. 579),

$$\hat{\beta}_N = \left(X' \left(P_Z - \frac{K-G-1}{T} M_Z \right) X \right)^{-1} X' \left(P_Z - \frac{K-G-1}{T} M_Z \right) y,$$

is not consistent under Assumption 4. The asymptotic bias is given by $\alpha^2(Q_{Z\Pi} + \alpha^2\Omega_{22})^{-1}\sigma_{uv_2}$. In the case of a single endogenous variable ($G = 1$) it is easy to see that $\hat{\beta}_N$ is less biased than standard 2SLS:

$$\frac{\operatorname{plim}(\hat{\beta}_N - \beta_0)}{\operatorname{plim}(\hat{\beta}_{2SLS} - \beta_0)} = \frac{Q_{Z\Pi} + \alpha\Omega_{22}}{\alpha^{-1}Q_{Z\Pi} + \alpha\Omega_{22}} < 1.$$

The limit distributions of $\hat{\beta}_L$ and $\hat{\beta}_B$ are given in the following two theorems.

THEOREM 2. *If Assumptions 2–4 are satisfied, the limiting distribution of* $\hat{\beta}_L$ *is given by* $T^{-1/2}(\hat{\beta}_L - \beta_0) \xrightarrow{d} N\left(0, \sigma_u^2 Q_{Z\Pi}^{-1} W_L Q_{Z\Pi}^{-1}\right)$, *where*

$$\begin{aligned} W_L = {} & Q_{Z\Pi} + \frac{\alpha(1-\gamma)}{(1-\alpha)^2}\left(\Omega_{22} - \frac{1}{\sigma_u^2}\sigma_{uv_2}\sigma_{uv_2}'\right) \\ & + \frac{\alpha(\gamma-\alpha)}{(1-\alpha)^2}\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 (v_{t2} - \frac{\sigma_{uv_2}}{\sigma_u^2}u_t)(v_{t2} - \frac{\sigma_{uv_2}}{\sigma_u^2}u_t)'\right] \\ & + \frac{\alpha}{(1-\alpha)}(\tilde{\mu}_{Z\Pi} - \mu_{Z\Pi})\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 (v_{t2} - \frac{\sigma_{uv_2}}{\sigma_u^2}u_t)'\right] \\ & + \frac{\alpha}{(1-\alpha)}\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 (v_{t2} - \frac{\sigma_{uv_2}}{\sigma_u^2}u_t)\right](\tilde{\mu}_{Z\Pi} - \mu_{Z\Pi})'. \end{aligned}$$

If (v_{t1}, v_{t2}') *is normally distributed, then* $W_L = Q_{Z\Pi} + (\alpha/(1-\alpha))(\Omega_{22} - \sigma_u^{-2}\sigma_{uv_2}\sigma_{uv_2}')$.

THEOREM 3. *If Assumptions 2–4 are satisfied, the limiting distribution of* $\hat{\beta}_B$ *is given by* $T^{-1/2}(\hat{\beta}_B - \beta_0) \xrightarrow{d} N\left(0, \sigma_u^2 Q_{Z\Pi}^{-1} W_B Q_{Z\Pi}^{-1}\right)$, *where*

$$\begin{aligned} W_B = {} & Q_{Z\Pi} + \frac{\alpha}{(1-\alpha)}(\Omega_{22} + \frac{1}{\sigma_u^2}\sigma_{uv_2}\sigma_{uv_2}') \\ & + \frac{\alpha(\gamma-\alpha)}{(1-\alpha)^2}\left\{\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 v_{t2}v_{t2}'\right] - \Omega_{22} - \frac{2}{\sigma_u^2}\sigma_{uv_2}\sigma_{uv_2}'\right\} \end{aligned}$$

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

$$+\frac{\alpha}{(1-\alpha)}\left\{(\tilde{\mu}_{Z\Pi}-\mu_{Z\Pi})\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 v_{t2}'\right]\right.$$
$$\left.+\,\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 v_{t2}\right](\tilde{\mu}_{Z\Pi}-\mu_{Z\Pi})'\right\}.$$

If (v_{t1}, v_{t2}') *is normally distributed, then* $W_B = Q_{Z\Pi} + (\alpha/(1-\alpha))(\Omega_{22} + \sigma_u^{-2}\sigma_{uv_2}\sigma_{uv_2}')$.

Remarks.

(a) From Theorems 2 and 3 it is clear that the conventional limit distribution results when $\alpha = 0$. This is no longer true when $\Pi' Z' Z \Pi$ grows at a rate slower than T. Chao and Swanson (2006) and Hansen et al. (2006) show that the asymptotic variance then also contains correction terms and the rate of convergence to normality is slower than $\sqrt{T}$.

(b) When the errors are normal and $\alpha > 0$, the distribution of LIML has been derived before by Bekker (1994). The asymptotic variance is larger than in the conventional fixed K, large T approximation, which is also true for bias-corrected 2SLS.[7] Moreover, $W_B - W_L \geq 0$, so that LIML is more efficient. This is a special case of Anderson and Kunitomo (2006, Thm. 3).[8] Chao and Swanson (2006) show that a similar efficiency result for LIML holds with many weak instruments and (u_t, v_{t2}') following an elliptical distribution.

(c) In general, i.e., without normality, it is not clear that $\hat{\beta}_L$ is more efficient than $\hat{\beta}_B$. Under some additional assumptions about the errors, however, W_L and W_B can be compared. Consider the case where $\alpha > 0$, the regression of v_{t2} on u_t is linear, and the conditional variance of v_{t2} given u_t is constant. The higher order terms in W_L vanish, and the asymptotic variance of LIML is the same as under normality. Letting $\kappa = \mathrm{E}[(u_t/\sigma_u)^4] - 3$ be the excess kurtosis of u_t/σ_u (relative to the standard normal distribution), we have (cf. Theorem 3):

$$\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 v_{t2}v_{t2}'\right] - \Omega_{22} - \frac{2}{\sigma_u^2}\sigma_{uv_2}\sigma_{uv_2}' = \frac{\kappa}{\sigma_u^2}\sigma_{uv_2}\sigma_{uv_2}',$$
$$\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^2 v_{t2}\right] = \frac{\sigma_{uv_2}}{\sigma_u}\mathrm{E}\left[\left(\frac{u_t}{\sigma_u}\right)^3\right].$$

Thus, unlike LIML, the asymptotic variance of $\hat{\beta}_B$ is sensitive to the skewness and kurtosis of u_t/σ_u. If in addition u_t is symmetrically distributed, then

$$W_B - W_L = \frac{\alpha}{(1-\alpha)^2}[2(1-\alpha) + (\gamma - \alpha)\kappa]\frac{\sigma_{uv_2}\sigma_{uv_2}'}{\sigma_u^2}.$$

Because the scalar constant in square brackets is nonnegative, LIML is in this case more efficient than bias-corrected 2SLS.[9]

(d) The difference in variance between the nonnormal and normal cases, $W^{(nn)} - W^{(n)}$, is generally an indefinite matrix. If (u_t, v_{t2}') satisfies the assumptions

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

of the previous remark, however, we find for bias-corrected 2SLS that

$$W_B^{(nn)} - W_B^{(n)} = \frac{\alpha(\gamma - \alpha)}{(1-\alpha)^2} \frac{\kappa}{\sigma_u^2} \sigma_{uv_2} \sigma_{uv_2}'.$$

Hence, when $\kappa < 0$ the variance under normality is larger.

(e) Suppose Z is a matrix of dummy variables, dividing the sample into mutually exclusive groups. If the group sizes are asymptotically equal, it can be verified from equations (2) and (3) that $\tilde{\mu}_{Z\Pi} = \mu_{Z\Pi}$ and $\gamma = \alpha$; hence, the third and fourth moments in Theorems 2 and 3 vanish.

4. CONCLUSION

We have derived the asymptotic distribution of two IV estimators when the concentration parameter and the number of instruments grow at the same rate as the sample size. The method of proof combines the martingale difference construction of Chao and Swanson (2004, 2006) and the finite-sample results of Bekker (1994) and Bekker and Van der Ploeg (2005). Unlike Morimune (1983) and Bekker (1994) we do not require normally distributed errors. The asymptotic variance generally contains correction terms that depend on third and fourth moments and on the asymptotic sequence of instruments and reduced form coefficients. For the LIML estimator Anderson and Kunitomo (2006) and Hansen et al. (2006) find the same limit distribution but under a different set of assumptions. In our paper we also present results for a bias-corrected version of 2SLS.

Under normality LIML is more efficient than bias-corrected 2SLS, a result in agreement with (Anderson and Kunitomo, 2006, Thm. 3). Without normality it is not clear which estimator is preferred: the relative efficiency depends on properties of the error distribution. Finally, when the errors are nonnormal, the many instruments asymptotic variance may increase or decrease.

NOTES

1. Van der Ploeg (1997) and Bekker and Van der Ploeg (2005) allow for nonnormal errors in the case of dummy instruments. In our paper we consider an alternative restriction on the instruments.

2. The limit results hold when T and K go to infinity simultaneously. In contrast, Stock and Yogo (2005) construct sequential limit distributions, which are large-K limits of the large-T distributions.

3. $K^{-1}\Pi' Z' d_{P_Z}$ is a weighted average because $\sum_{t=1}^{T} d_{P_Z,t} = \sum_{t=1}^{T} p_{tt} = K$.

4. The convergence of $T^{-1}\Pi' Z' Z \Pi$ and $T^{-1}\Pi' Z' \iota_T$ in (2) does not guarantee that $\|z_t' \Pi\|$ is bounded for all t. Conversely, the boundedness of $\|z_t' \Pi\|$ in Assumption 4(b) does not imply that $T^{-1}\Pi' Z' Z \Pi$ and $T^{-1}\Pi' Z' \iota_T$ actually converge. Both conditions are needed in Assumption 4.

5. Chao and Swanson (2006) also have $K/T \to \alpha$, but the growth rate of the concentration parameter lies between $\sqrt{K}$ and K. In Stock and Yogo (2005) $K^2/T \to 0$, and the concentration parameter is $O(K)$.

6. In a sequence with many weak instruments Chao and Swanson (2007) show that the bias of 2SLS can be consistently estimated. Hence, in that framework it is also possible to construct a bias-corrected estimator.

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

7. The objective function in (5) provides some intuition: under normality the numerator and denominator are χ^2 random variables with K and $T-K$ degrees of freedom, respectively. As K increases with T fixed, the variance of the numerator starts to dominate the variance of the objective function, and as a consequence the variability of $\hat{\beta}_L$ increases.

8. The first efficiency result for LIML with many strong instruments appears in Van der Ploeg (1997). Van der Ploeg establishes that with dummy instruments and normal errors, LIML is efficient within the class of consistent estimators that are differentiable functions of P_Z.

9. This follows from the definition of κ and the fact that $\mathrm{E}[(u_t/\sigma_u)^4] \geq 1$.

REFERENCES

Anderson, T. & N. Kunitomo (2006) On Asymptotic Properties of the LIML Estimator with Possibly Many Instruments. Working paper, University of Tokyo.

Bekker, P.A. (1994) Alternative approximations to the distributions of instrumental variables estimators. *Econometrica* 62, 657–681.

Bekker, P.A. & J. Van der Ploeg (2005) Instrumental variable estimation based on grouped data. *Statistica Neerlandica* 59, 239–267.

Billingsley, P. (1995) *Probability and Measure*. Wiley.

Chao, J.C. & N.R. Swanson (2004) Asymptotic Distribution of JIVE in a Heteroscedastic IV Regression with Many Weak Instruments. Working paper, Rutgers University.

Chao, J.C. & N.R. Swanson (2005) Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.

Chao, J.C. & N.R. Swanson (2006) Asymptotic normality of single-equation estimators for the case with a large number of weak instruments. In D. Corbae, S.N. Durlauf, & B.E. Hansen (eds.), *Econometric Theory and Practice: Frontiers of Analysis and Applied Research*, pp. 82–124. Cambridge University Press.

Chao, J.C. & N.R. Swanson (2007) Alternative approximations of the bias and MSE of the IV estimator under weak identification with an application to bias correction. *Journal of Econometrics* 137, 515–555.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

Hansen, C., J.A. Hausman, & W.K. Newey (2006) Estimation with Many Instrumental Variables. Working paper, MIT.

Hausman, J.A., W.K. Newey, T. Woutersen, J.C. Chao, & N.R. Swanson (2007) Instrumental Variable Estimation with Heteroscedasticity and Many Instruments. Working paper, MIT.

Morimune, K. (1983) Approximate distributions of k-class estimators when the degree of overidentifiability is large compared with the sample size. *Econometrica* 51, 821–842.

Nagar, A.L. (1959) The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* 27, 575–595.

Rohatgi, V.K. (1971) Convergence of weighted sums of independent random variables. *Proceedings of the Cambridge Philosophical Society* 69, 305–307.

Stock, J.H. & M. Yogo (2005) Asymptotic distributions of instrumental variables statistics with many instruments. In D.W.K. Andrews & J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 109–120. Cambridge University Press.

Van der Ploeg, J. (1997) Instrumental Variable Estimation and Group-Asymptotics. Ph.D. dissertation, Rijksuniversiteit Groningen, The Netherlands.

APPENDIX

Proof of Theorem 1. We focus on establishing the limit distribution of $S_{ij} \equiv u_i' C u_j$; from there the generalization to $U'CUa$ is straightforward. By symmetry of C the

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

demeaned variable $\tilde{S}_{ij} = S_{ij} - \mathrm{E}(S_{ij})$ can be written as

$$\tilde{S}_{ij} = \sum_{t=1}^{T} c_{tt}(u_{ti}u_{tj} - m_{ti}m_{tj} - \omega_{ij}) + \sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{st}(u_{si}u_{tj} - m_{si}m_{tj} + u_{ti}u_{sj} - m_{ti}m_{sj}). \quad \textbf{(A.1)}$$

A similar construction is used in Chao and Swanson (2006, Lem. A.3). Note that $\tilde{S}_{ij}$ depends on T. Defining the triangular array of σ-algebras $\left\{\mathcal{F}_{T,t} = \sigma\left((v_{1i}, v_{1j}), \ldots, (v_{ti}, v_{tj})\right)\right\}_{t=0}^{T}$, with $\mathcal{F}_{T,0}$ being the trivial σ-algebra, it is easy to see that

$$D_{ij,t} = \mathrm{E}(\tilde{S}_{ij}|\mathcal{F}_{T,t}) - \mathrm{E}(\tilde{S}_{ij}|\mathcal{F}_{T,t-1}), \qquad 1 \le t \le T, \quad \textbf{(A.2)}$$

defines a triangular martingale difference array. Given that $\mathrm{E}(\tilde{S}_{ij}|\mathcal{F}_{T,0}) = 0$ and $\tilde{S}_{ij}$ is $\mathcal{F}_{T,T}$-measurable, it follows that $\tilde{S}_{ij} = \sum_{t=1}^{T} D_{ij,t}$. Using (A.1) and (A.2) we obtain an explicit expression for $D_{ij,t}$:

$$D_{ij,t} = c_{tt}(u_{ti}u_{tj} - m_{ti}m_{tj} - \omega_{ij}) + \sum_{s=t+1}^{T} c_{st}(v_{ti}m_{sj} + m_{si}v_{tj}) + \sum_{s=1}^{t-1} c_{st}(u_{si}v_{tj} + v_{ti}u_{sj}). \quad \textbf{(A.3)}$$

We now proceed to verify the two conditions of Billingsley (1995, Thm. 35.12). First, we identify the limit of $T^{-1}\sum_{t=1}^{T}\mathrm{E}(D_{ij,t}^2|\mathcal{F}_{T,t-1})$. It follows from (A.3) that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\{\mathrm{E}\left(D_{ij,t}^2|\mathcal{F}_{T,t-1}\right) - \mathrm{E}\left(D_{ij,t}^2\right)\right\} \\
&\quad = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1} c_{st}c_{rt}(\omega_{ii}[m_{sj}v_{rj} + v_{sj}m_{rj} + v_{sj}v_{rj}] \\
&\qquad\qquad + \omega_{jj}[m_{si}v_{ri} + v_{si}m_{ri} + v_{si}v_{ri}]) \\
&\qquad + \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1} c_{st}c_{rt}\omega_{ij}\left(m_{si}v_{rj} + v_{si}m_{rj} + v_{si}v_{rj} + m_{sj}v_{ri} + m_{sj}v_{ri} + v_{sj}v_{ri}\right) \\
&\qquad + \frac{2}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{tt}c_{st}(\omega_{ii}m_{tj}v_{sj} + \omega_{jj}m_{ti}v_{si} + \omega_{ij}[v_{si}m_{tj} + v_{sj}m_{ti}]) \qquad \textbf{(A.4)} \\
&\qquad + \frac{2}{T}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\sum_{q=1}^{t-1} c_{st}c_{qt}(\omega_{ii}m_{sj}v_{qj} + \omega_{jj}m_{si}v_{qi} + \omega_{ij}[m_{sj}v_{qi} + m_{si}v_{qj}]) \\
&\qquad + \frac{2}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{tt}c_{st}(v_{si}\mathrm{E}(v_{ti}v_{tj}^2) + v_{sj}\mathrm{E}(v_{ti}^2v_{tj})) - \frac{2}{T}(\omega_{ii}\omega_{jj} + \omega_{ij}^2)\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{st}^2.
\end{aligned}$$

Terms in (A.4) of the form

$$\frac{2}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{tt}c_{st}v_{si}, \qquad \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1} c_{st}c_{rt}m_{si}v_{rj},$$

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=t+1}^{T}\sum_{q=1}^{t-1} c_{ts}c_{qt}m_{si}v_{qj}, \qquad \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{tt}c_{st}m_{ti}v_{sj}$$

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

are all $o_p(1)$. This follows from rewriting the various sums and verifying the conditions of Rohatgi (1971, Thm. 1) under Assumption 1. Regarding the coefficient of $(\omega_{ii}\omega_{jj}+\omega_{ij}^2)$ in (A.4), Assumption 1(c) and symmetry of C yield

$$\frac{2}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}c_{st}^2=\frac{\operatorname{tr}(C^2)}{T}-\frac{d_C'd_C}{T}\to\tau_{C^2}-\delta_C,\qquad T\to\infty.$$

Finally, we need to deal with terms that contain products of errors:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\sum_{r=1}^{t-1}c_{st}c_{rt}v_{si}v_{rj}. \tag{A.5}$$

Rewrite this sum as

$$\begin{aligned}\frac{1}{T-1}\sum_{s=1}^{T-1}v_{sk}\left(\sum_{t=s+1}^{T}\sum_{r=1}^{t-1}c_{st}c_{rt}v_{rl}\right)&=\frac{1}{T-1}\sum_{s=1}^{T-1}v_{sk}\left(\sum_{t=s+1}^{T}c_{st}^2v_{sl}\right)\\&\quad+\frac{1}{T-1}\sum_{s=1}^{T-1}v_{sk}\left(\sum_{t=s+1}^{T}\sum_{r=1}^{s-1}c_{st}c_{rt}v_{rl}\right)\\&\quad+\frac{1}{T-1}\sum_{s=1}^{T-1}v_{sk}\left(\sum_{t=s+1}^{T}\sum_{r=s+1}^{t-1}c_{st}c_{rt}v_{rl}\right).\end{aligned}$$

It follows from Assumption 1 and an application of Chebyshev's inequality to each of the three sums that the probability limit of (A.5) is $\left(\frac{1}{2}\right)\omega_{ij}(\tau_{C^2}-\delta_C)$. Collecting all results implies for (A.4) that

$$\frac{1}{T}\sum_{t=1}^{T}\left\{\mathrm{E}\left(D_{ij,t}^2|\mathcal{F}_{T,t-1}\right)-\mathrm{E}\left(D_{ij,t}^2\right)\right\}=o_p(1),$$

which identifies the asymptotic variance as the limit of $T^{-1}\sum_{t=1}^{T}\mathrm{E}(D_{ij,t}^2)$.

Second, to verify the Lindeberg condition in (Billingsley, 1995, Thm. 35.12), note that under Assumption 1 it is possible to construct an integrable random variable D, such that

$$\sup_{T\geq 1}\sup_{1\leq t\leq T}\frac{1}{T}D_{ij,t}^2\leq D. \tag{A.6}$$

This implies that $T^{-1}D_{ij,t}^2$ is uniformly integrable, which in turn is sufficient for the Lindeberg condition (e.g., Davidson, 1994, Thm. 23.10). Thus

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}D_{ij,t}&=\frac{1}{\sqrt{T}}(u_i'Cu_j-m_i'Cm_j-\operatorname{tr}(C)\omega_{ij})\xrightarrow{d}N\left(0,\sigma_{ij}^2\right),\\\sigma_{ij}^2&=\lim_{T\to\infty}\frac{1}{T}\operatorname{Var}(u_i'Cu_j).\end{aligned}$$

The conclusion of Theorem 1 now follows from combining Assumption 1 and Lemma 1 of Bekker and Van der Ploeg (2005). ■

Proof of Theorem 2. Assume that Assumptions 3 and 4 hold. As in Bekker (1994) we first transform the model to a canonical form. Let

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

$$F = \begin{bmatrix} 1 & 0 \\ -\beta_0 & I \end{bmatrix} \begin{bmatrix} \sigma_u^{-1} & -\sigma_u^{-2}\sigma'_{uv_2}Q^{-1/2} \\ 0 & Q^{-1/2} \end{bmatrix}, \qquad Q = \Omega_{22} - \frac{1}{\sigma_u^2}\sigma_{uv_2}\sigma'_{uv_2}.$$

Then $(y^*, X^*) = (y, X)F$ has parameter $\beta_0^* = 0$ and error variance I_G. The first-order condition for LIML yields

$$\sqrt{T}\hat{\beta}^* = -\left[\frac{\partial^2 Q_T^*(\beta_0^*)}{\partial\beta\partial\beta}\right]^{-1}\sqrt{T}\frac{\partial Q_T^*(\beta_0^*)}{\partial\beta} + o_p(1),$$

where $Q_T^*(\beta)$ is the objective function from (5). It is easy to show that the Hessian term has probability limit $2(1-\alpha)^{-1}Q^{-1/2\prime}Q_{Z\Pi}Q^{-1/2}$. Because $T^{1/2}\partial Q_T^*(\beta_0^*)/\partial\beta$ is, up to a remainder of order $o_p(1)$, a subvector of

$$-\frac{2}{(1-\alpha)^2}T^{-1/2}(y^*, X^*)'(P_Z - \alpha I_T)(y^*, X^*)(1, -\beta_0^{*\prime})',$$

we apply Theorem 1 with $U = (y^*, X^*)$ and $C = P_Z - \alpha I_T$. Then $\tau_C = 0$, $\tau_{C^2} = \alpha(1-\alpha)$, and $\delta_C = \alpha(\gamma - \alpha)$. Substituting $M = Z\Pi^*(0, I_G)$ and $\Pi^* = \Pi Q^{-1/2}$ we get

$$Q_{CM} = (1-\alpha)^2(0, I_G)'Q^{-1/2\prime}Q_{Z\Pi}Q^{-1/2}(0, I_G),$$

$$\mu_{CM} = \alpha(1-\alpha)Q^{-1/2\prime}(\tilde{\mu}_{Z\Pi} - \mu_{Z\Pi}).$$

An application of Slutsky's theorem and Lemma 1 in Bekker and Van der Ploeg (2005) yields $T^{1/2}\hat{\beta}^* \xrightarrow{d} N(0, A^{*-1}W^*A^{*-1})$, where

$$W^* = Q^{-1/2\prime}Q_{Z\Pi}Q^{-1/2} + \frac{\alpha}{(1-\alpha)}I_G + \frac{\alpha(\gamma-\alpha)}{(1-\alpha)^2}\left\{\mathrm{E}\left[v_{t1}^{*2}v_{t2}^*v_{t2}^{*\prime}\right] - I_G\right\}$$

$$+\frac{\alpha}{(1-\alpha)}\left\{Q^{-1/2\prime}(\tilde{\mu}_{Z\Pi} - \mu_{Z\Pi})\mathrm{E}\left[v_{t1}^{*2}v_{t2}^{*\prime}\right] + \mathrm{E}\left[v_{t1}^*v_{t2}^*\right](\tilde{\mu}_{Z\Pi} - \mu_{Z\Pi})'Q^{-1/2}\right\}$$

and $A^* = Q^{-1/2\prime}Q_{Z\Pi}Q^{-1/2}$. Transforming back to the original model (cf. Bekker 1994, p. 680) leads to the first conclusion of Theorem 2. In the case of normal errors, the same argument goes through with a simpler expression for W^*, and hence W_L. ■

Proof of Theorem 3. First of all,

$$\sqrt{T}(\hat{\beta}_{B2SLS} - \beta_0)$$

$$= \left(T^{-1}X'\left(P_Z - \frac{K}{T}I_T\right)X\right)^{-1}T^{-1/2}X'\left(P_Z - \frac{K}{T}I_T\right)(y, X)(1, -\beta_0')'$$

$$= ((1-\alpha)Q_{Z\Pi})^{-1}T^{-1/2}X'\left(P_Z - \frac{K}{T}I_T\right)(y, X)(1, -\beta_0')' + o_p(1). \qquad \textbf{(A.7)}$$

Under Assumptions 2–4, the conditions of Theorem 1 are satisfied with $C = P_Z - (K/T)I_T$. Then $\tau_C = 0$, $\tau_{C^2} = \alpha(1-\alpha)$, $\delta_C = \alpha(\gamma - \alpha)$. Using $M = Z\Pi(\beta_0, I_G)$ we find

$$Q_{CM} = (1-\alpha)^2(\beta_0, I_G)'Q_{Z\Pi}(\beta_0, I_G), \qquad \mu_{CM} = \alpha(1-\alpha)(\beta_0, I_G)'(\tilde{\mu}_{Z\Pi} - \mu_{Z\Pi}).$$

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press

Because $T^{-1/2}X'(P_Z-(K/T)I_T)(y,X)(1,-\beta_0')'$ has mean zero, it remains to find the variance. Applying Theorem 1, the asymptotic variance is

$$\begin{aligned}\lim_{T,K\to\infty}&\operatorname{Var}[T^{-1/2}X'(P_Z-\frac{K}{T}I_T)(y,X)(1,-\beta_0')']\\&=(1-\alpha)^2\sigma_u^2Q_{Z\Pi}+\alpha(1-\alpha)\{\sigma_u^2\Omega_{22}+\sigma_{uv_2}\sigma_{uv_2}'\}\\&\quad+\alpha(\gamma-\alpha)\{\mathrm{E}[u_t^2v_{t2}v_{t2}']-\sigma_u^2\Omega_{22}-2\sigma_{uv_2}\sigma_{uv_2}'\}\\&\quad+\alpha(1-\alpha)\{(\tilde{\mu}_{Z\Pi}-\mu_{Z\Pi})\mathrm{E}[u_t^2v_{t2}']+\mathrm{E}[u_t^2v_{t2}](\tilde{\mu}_{Z\Pi}-\mu_{Z\Pi})'\},\end{aligned}$$

which yields the first conclusion of Theorem 3. The argument under normality is similar and therefore omitted. ■

https://doi.org/10.1017/S0266466609100117 Published online by Cambridge University Press