

Many Instrument Asymptotics

Whitney K. Newey

MIT

June 2004

These notes are about the properties of instrumental variable (IV) estimators when there are many instruments. We base these notes on Bekker (1994 “Alternative Approximations to the Distributions of Instrumental Variable Estimators,” *Econometrica*), where more primitive references can be found. The theoretical innovation in these notes is to allow for nonnormality of the disturbances. We show that Bekker (1994) results for LIML still hold if joint normality is weakened to first and second moment independence.

Model:

The model we consider is given by

$$\begin{aligned} y_{T \times 1} &= X_{T \times G} \delta_0 + u_{T \times 1}, \\ X &= Z_{T \times K} \Pi_{K \times G} + V_{T \times G}. \end{aligned}$$

Here Z and Π are implicitly allowed to depend on T . In particular, we will allow K to grow with T at the same rate as T . We will consider Z as nonrandom. Alternatively, one could interpret the following results as being conditional on Z . This model differs somewhat from Bekker (1994), in that we assume that the reduced form is correctly specified.

To obtain asymptotic results for the estimators it is necessary to impose some conditions. We will make use of two assumptions, the first of which is used to show consistency and the second of which is added for asymptotic normality. Let Z_t, u_t, V_t denote the t^{th} row of Z, u , and V respectively.

Assumption 1: $(u_1, V_1), \dots, (u_T, V_T)$ are i.i.d. with mean zero and finite fourth moments, the variance of (u_t, V_t) is nonsingular, $Z'Z$ is nonsingular, and as $T \rightarrow \infty$ there is a scalar α with $0 < \alpha < 1$ and a positive definite matrix Q such that

$$K/T \rightarrow \alpha, \Pi'Z'Z\Pi/T \rightarrow Q.$$

This condition allows the number of instruments to grow at the same rate as the sample size, but requires that $\Pi'Z'Z\Pi/T$ converges. In this sense adding additional instruments does not add information. The restriction that $Z'Z$ is nonsingular is essentially a normalization. Alternatively, we could interpret K as the rank of $Z'Z$. For the second assumption, let $\sigma_{Vu} = E[V_t' u_t]$, $\sigma_u^2 = E[u_t^2]$, $\gamma = \sigma_{Vu}/\sigma_u^2$, and $\tilde{V} = V - u\gamma'$, having t^{th} row \tilde{V}_t .

Assumption 2: $E[u_t|\tilde{V}_t] = 0$, $E[u_t^2|\tilde{V}_t] = \sigma_u^2$, for some $p > 2$, $E[|u_t|^p|\tilde{V}_t]$ is bounded, and $\max_{t \leq T} \|\Pi Z'_t\|/\sqrt{T} \rightarrow 0$.

The vector \tilde{V}_t consists of residuals from the population regression of V_t on u_t and so satisfies $E[\tilde{V}_t u_t] = 0$ by construction. Under joint normality of (u_t, V_t) , u_t and \tilde{V}_t are independent, and so the first two conditions automatically hold. In general these two conditions weaken the joint normality restriction to first and second moment independence of u_t from \tilde{V}_t . The other two conditions are useful for the central limit theorem, with the last one implying asymptotic normality of $\Pi' Z' u/\sqrt{T}$. It is interesting to note that no other restrictions are imposed on Z .

Estimators:

Two important estimators for the structural coefficients δ are two stage least squares (2SLS) and limited information maximum likelihood (LIML). For $P_Z = Z(Z'Z)^{-1}Z'$ these estimators are given by

$$\begin{aligned} 2SLS & : \hat{\delta} = (X'P_Z X)^{-1} X'P_Z y, \\ LIML & : \tilde{\delta} = \arg \min_{\delta} [(y - X\delta)' P_Z (y - X\delta)] / (y - X\delta)'(y - X\delta). \end{aligned}$$

Under many instrument asymptotics 2SLS will be inconsistent but LIML will be consistent. This result can be explained by the respective first-order conditions of the estimators. For 2SLS the first-order condition is

$$\frac{1}{T} X' P_Z (y - X\hat{\delta}) = 0.$$

The reason that 2SLS is not consistent with many instruments is that $X'P_Z u/T$ does not converge to zero. By a standard calculation,

$$E[X'P_Z u/T] = E[V'P_Z u]/T = \frac{K}{T} \sigma_{V u},$$

which does not vanish asymptotically. Thus, 2SLS is setting δ so that $X'P_Z(y - X\delta)/T$ is equal to an incorrect value

In contrast, the first-order condition for LIML can be written as

$$\frac{1}{T} X' P_Z (y - X\tilde{\delta}) = \frac{\tilde{u}' P_Z \tilde{u}/T}{\tilde{u}' \tilde{u}/T} \frac{X' \tilde{u}}{T}, \tilde{u} = y - X\tilde{\delta}.$$

Note that

$$E[u' P_Z u/T] = \sigma_u^2 \frac{K}{T}, E[u'u/T] = \sigma_u^2, E[X'u/T] = \sigma_{V u},$$

so that LIML is setting $X'P_Z(y - X\delta)/T$ equal to an estimate of its expectation. In this way LIML adjusts for the fact that $E[X'P_Z u/T]$ does not vanish asymptotically.

One could also base consistent estimators on other corrections for nonzero expectation of $X'P_Z u/T$. For instance, one could use K/T in place of its estimator in the LIML first order conditions to form an estimator $\check{\delta}$ as the solution to

$$\frac{1}{T}X'P_Z(y - X\check{\delta}) = \frac{K}{T}\frac{X'\check{u}}{T}, \check{u} = y - X\check{\delta}.$$

This estimator has an explicit form as

$$\check{\delta} = (X'(P_Z - \frac{K}{T})X)^{-1}X'(P_Z - \frac{K}{T})y.$$

This estimator is similar to the Nagar (1959) bias corrected 2SLS estimator. Under certain conditions it turns out to be less efficient than LIML under many instrument asymptotics, as shown below. We conjecture that, under some set of conditions, LIML is efficient in a wide class of estimators that correct for many instrument biases.

Consistency:

Some intermediate properties lead to the consistency results. Let $\Omega = E[V_t'V_t]$.

LEMMA 1: *If Assumption 1 is satisfied and $\bar{\delta} \xrightarrow{p} \delta_0$ then for $\bar{u} = y - X\bar{\delta}$,*

$$\begin{aligned} \bar{u}'\bar{u}/T &\xrightarrow{p} \sigma_u^2, X'\bar{u}/T \xrightarrow{p} \sigma_{V_u}, X'X/T \xrightarrow{p} Q + \Omega, \\ \bar{u}'P_Z\bar{u}/T &\xrightarrow{p} \alpha\sigma_u^2, X'P_Z\bar{u}/T \xrightarrow{p} \alpha\sigma_{V_u}, X'P_ZX/T \xrightarrow{p} Q + \alpha\Omega, \end{aligned}$$

The inconsistency of 2SLS can be seen from the usual calculation (applying the continuous mapping theorem),

$$\hat{\delta} = \delta_0 + (X'P_ZX/T)^{-1}X'P_Zu/T \xrightarrow{p} \delta_0 + (Q + \alpha\Omega)^{-1}\alpha\sigma_{V_u} \neq \delta_0.$$

The consistency of the explicit bias corrected estimator $\check{\delta}$ can be shown similarly. We have

$$\begin{aligned} \check{\delta} &= \delta_0 + (X'(P_Z - \frac{K}{T}I)X/T)^{-1}X'(P_Z - \frac{K}{T})u/T \\ &= \delta_0 + (Q + \alpha\Omega - \alpha[Q + \Omega])^{-1}(\alpha\sigma_{V_u} - \alpha\sigma_{V_u}) + o_p(1) \\ &= \delta_0 + o_p(1). \end{aligned}$$

Consistency of LIML is harder to show, because it does not have an explicit formula. However, we can use the usual methods for nonlinear estimators to explain its consistency by showing that the LIML objective function converges to a function that is uniquely minimized at δ_0 . As usual, the probability limit of the minimum will be the minimum of the probability limit. Specifically,

$$\tilde{\delta} = \arg \min_{\delta} \hat{S}(\delta), \hat{S}(\delta) = \hat{A}(\delta)/\hat{B}(\delta),$$

where

$$\begin{aligned}\hat{A}(\delta) &= (y - X\delta)'P_Z(y - X\delta)/T = u'P_Zu/T - 2u'P_ZX(\delta - \delta_0)/T \\ &\quad + (\delta - \delta_0)'X'P_ZX(\delta - \delta_0)/T \\ \xrightarrow{p} A(\delta) &= \alpha\sigma_u^2 - 2\alpha\sigma'_{V_u}(\delta - \delta_0) + (\delta - \delta_0)'(Q + \alpha\Omega)(\delta - \delta_0)\end{aligned}$$

by Lemma 1, and

$$\begin{aligned}\hat{B}(\delta) &= (y - X\delta)'P_Z(y - X\delta)/T = u'u/T - 2u'X(\delta - \delta_0)/T \\ &\quad + (\delta - \delta_0)'X'X(\delta - \delta_0)/T \\ \xrightarrow{p} B(\delta) &= \sigma_u^2 - 2\sigma'_{V_u}(\delta - \delta_0) + (\delta - \delta_0)'(Q + \Omega)(\delta - \delta_0).\end{aligned}$$

Note that $A(\delta) = \alpha B(\delta) + (1 - \alpha)(\delta - \delta_0)'Q(\delta - \delta_0)$. By the continuous mapping theorem,

$$\hat{S}(\delta) \xrightarrow{p} A(\delta)/B(\delta) = \alpha + (1 - \alpha)(\delta - \delta_0)'Q(\delta - \delta_0)/B(\delta).$$

It is straightforward to show that $B(\delta)$ will be nonzero for all δ (by nonsingularity of the variance matrix of (u_t, V_t)) so that $A(\delta)/B(\delta)$ is uniquely minimized at δ_0 .

An actual proof of consistency using this approach is a little more complicated, but can be done, and is in the proof of the following result:

THEOREM 1: *If Assumption 1 is satisfied then $\tilde{\delta} \xrightarrow{p} \delta_0$.*

Asymptotic Normality

Under the large K asymptotics the LIML estimator is asymptotically normal, with an asymptotic variance that is larger than the usual one. Also, this larger asymptotic variance can be consistently estimated under the large K asymptotics. The importance of these results is that by using the consistent variance estimator we can construct standard normal confidence intervals under large K asymptotics. Since asymptotic theory is used as an approximation, this approach should provide improved inference over the usual method when there are many instruments.

The asymptotic normality of LIML can be shown using an expansion of the first-order conditions. For $u(\delta) = y - X\delta$, consider the function

$$\hat{D}(\delta) = -\frac{X'P_Zu(\delta)}{T} + \frac{u(\delta)'P_Zu(\delta)}{u(\delta)'u(\delta)} \frac{X'u(\delta)}{T}$$

The first-order conditions for $\tilde{\delta}$ can be written as

$$0 = \hat{D}(\tilde{\delta})$$

Then to show asymptotic normality and derive the asymptotic variance matrix it suffices to show that for a nonsingular matrix H , a matrix Σ , and any $\bar{\delta} \xrightarrow{p} \delta_0$,

$$\sqrt{T}\hat{D}(\delta_0) \xrightarrow{d} N(0, \Sigma), \quad \frac{\partial \hat{D}(\bar{\delta})}{\partial \delta} \xrightarrow{p} H.$$

It will then follow in the usual way that

$$\sqrt{T}(\tilde{\delta} - \delta) \xrightarrow{d} N(0, \Lambda), \Lambda = H^{-1}\Sigma H'^{-1}.$$

Also for consistent estimators \hat{H} and $\hat{\Sigma}$, a consistent estimator $\hat{\Lambda} = \hat{H}^{-1}\hat{\Sigma}\hat{H}'^{-1}$ can be constructed and inference carried out in the usual way.

The first thing we show is convergence in probability of $\partial\hat{D}(\bar{\delta})/\partial\delta$. Let $\bar{u} = u(\bar{\delta}) = y - X\bar{\delta}$. Then differentiating gives

$$\begin{aligned} \frac{\partial\hat{D}}{\partial\delta}(\bar{\delta}) &= \frac{X'P_ZX}{T} - \frac{\bar{u}'P_Z\bar{u}X'X}{\bar{u}'\bar{u}T} - \frac{X'\bar{u}\bar{u}'P_ZX}{T\bar{u}'\bar{u}} - \frac{X'P_Z\bar{u}\bar{u}'X}{\bar{u}'\bar{u}T} + 2\frac{\bar{u}P_Z\bar{u}X'\bar{u}\bar{u}'X}{(\bar{u}'\bar{u})^2T} \\ &= \frac{X'P_ZX}{T} - \frac{\bar{u}'P_Z\bar{u}X'X}{\bar{u}'\bar{u}T} + (X'\bar{u}/\bar{u}'\bar{u})\hat{D}(\bar{\delta})' + \hat{D}(\bar{\delta})\bar{u}'X/\bar{u}'\bar{u}. \end{aligned}$$

From Lemma 1 it then follows by $\hat{D}(\bar{\delta}) \xrightarrow{p} 0$ that

$$\frac{\partial\hat{D}(\bar{\delta})}{\partial\delta} \xrightarrow{p} Q + \alpha\Omega - \alpha(Q + \Omega) = (1 - \alpha)Q = H.$$

Next we consider the behavior of $\sqrt{T}\hat{D}(\delta_0)$. For $\hat{\gamma} = X'u/u'u$, $\gamma = \sigma_{V_u}/\sigma_u^2$, and $\tilde{V} = V - u\gamma'$, by $E[\tilde{V}_t u_t] = 0$ and the Lindberg-Levy central limit theorem, $\tilde{V}'u/\sqrt{T}$ is bounded in probability. Also, $\Pi'X'u/\sqrt{T}$ has bounded second moment and so is bounded in probability. Also, by Lemma 1, $u'P_Zu/u'u \xrightarrow{p} \alpha$. Therefore, by the Slutsky theorem,

$$\sqrt{T}(\hat{\gamma} - \gamma)\frac{u'P_Zu}{T} = \frac{X'u - \gamma u'u}{\sqrt{T}}\frac{u'P_Zu}{u'u} = \alpha\frac{(Z\Pi + \tilde{V})'u}{\sqrt{T}} + o_p(1).$$

We then have

$$\begin{aligned} \sqrt{T}\hat{D}(\delta_0) &= -\frac{X'P_Zu}{\sqrt{T}} + \frac{u'P_ZuX'u}{u'u\sqrt{T}} = -\frac{(X - u\hat{\gamma}')P_Zu}{\sqrt{T}} \\ &= -\frac{(X - u\gamma')P_Zu}{\sqrt{T}} + \sqrt{T}(\hat{\gamma} - \gamma)\frac{u'P_Zu}{T} \\ &= -\frac{\Pi'Z'u}{\sqrt{T}} - \frac{\tilde{V}'P_Zu}{\sqrt{T}} + \alpha\frac{(Z\Pi + \tilde{V})'u}{\sqrt{T}} + o_p(1) \\ &= \frac{W'u}{\sqrt{T}} + o_p(1), W = -(1 - \alpha)Z\Pi - (P_Z - \alpha I)\tilde{V}. \end{aligned}$$

Thus, for asymptotic normality of $\sqrt{T}\hat{D}(\delta_0)$ it suffices (by the Slutsky theorem) that a central limit theorem applies to $W'u/\sqrt{T}$. Also, as always for the central limit theorem, the asymptotic variance of $W'u/\sqrt{T}$ will be the limit of its variance. Assumption 2

implies that $E[u|\tilde{V}] = 0$ and $Var(u|\tilde{V}) = \sigma_u^2 I_T$, so that for $\tilde{\Omega} = E[\tilde{V}'\tilde{V}_t]$,

$$\begin{aligned} Var\left(\frac{W'u}{\sqrt{T}}\right) &= E\left[Var\left(\frac{W'u}{\sqrt{T}}|\tilde{V}\right)\right] = \sigma_u^2 E[W'W/T] \\ &= \sigma_u^2 \left\{ (1-\alpha)^2 \frac{\Pi'Z'Z\Pi}{T} + \frac{E[\tilde{V}'(P_Z - \alpha I)^2\tilde{V}]}{T} \right\} \\ &= \sigma_u^2 \left\{ (1-\alpha)^2 \frac{\Pi'Z'Z\Pi}{T} + \left(\frac{K}{T} - 2\alpha\frac{K}{T} + \alpha^2\right)\tilde{\Omega} \right\} \\ &\longrightarrow \sigma_u^2 \left\{ (1-\alpha)^2 Q + \alpha(1-\alpha)\tilde{\Omega} \right\} = \Sigma. \end{aligned}$$

Combining this result with the formula for H derived above, we find that the asymptotic variance of the LIML estimator, under these large K asymptotics, will be

$$\Lambda = H^{-1}\Sigma H^{-1} = \sigma_u^2 Q^{-1} + \sigma_u^2 \frac{\alpha}{1-\alpha} Q^{-1}\tilde{\Omega}Q^{-1}.$$

The leading term $\sigma_u^2 Q^{-1}$ is the asymptotic variance when $K/T \rightarrow 0$, so that the asymptotic variance of LIML under these many instrument asymptotics is larger than the usual one. It is important to note that this addition can be important even when α is small (i.e. when K/T is small). In applications the r-squared for the reduced form is often very low, corresponding to Q being very small relative to $\tilde{\Omega}$. In such cases adding the additional term can substantially raise the asymptotic variance.

The following result makes these calculations precise.

THEOREM 2: *If Assumptions 1 and 2 are satisfied then*

$$\sqrt{T}(\tilde{\delta} - \delta_0) \xrightarrow{d} N(0, \Lambda), \Lambda = \sigma_u^2 Q^{-1} + \sigma_u^2 \frac{\alpha}{1-\alpha} Q^{-1}\tilde{\Omega}Q^{-1}.$$

The asymptotic variance Λ can be consistently estimated by combining consistent estimators of its components. We can estimate σ_u^2 consistently in the usual way from the residuals $\tilde{u} = Y - X\tilde{\delta}$, as

$$\hat{\sigma}_u^2 = \tilde{u}'\tilde{u}/(T - G)$$

Since the matrix H is the limit of $\partial\hat{D}(\delta_0)/\partial\delta$, it can be estimated by

$$\hat{H} = \partial\hat{D}(\tilde{\delta})/\partial\delta = \frac{X'P_Z X}{T} - \frac{\tilde{u}'P_Z \tilde{u} X'X}{\tilde{u}'\tilde{u} T}.$$

It is straightforward to show that this matrix must be positive semi-definite. Recall that $\tilde{\delta}$ minimizes the objective function $\hat{R}(\delta) = u(\delta)'P_Z u(\delta)/u(\delta)'u(\delta)$, so that $\partial^2 \hat{R}(\tilde{\delta})/\partial\delta\partial\delta'$ is positive semi-definite by the necessary second-order conditions. Also, $\partial\hat{R}(\delta)/\partial\delta = D(\delta)/u(\delta)'u(\delta)$, so that $\partial^2 \hat{R}(\tilde{\delta})/\partial\delta\partial\delta' = [\partial D(\tilde{\delta})/\partial\tilde{\delta}]/\tilde{u}'\tilde{u}$, implying that $\partial D(\tilde{\delta})/\partial\delta$ is also positive semi-definite.

To estimate Σ , define the positive semi-definite matrix \hat{J} by

$$\hat{J} = X'P_Z X/T - X'P_Z \tilde{u}\tilde{u}'P_Z X/T\tilde{u}'P_Z \tilde{u} = X'P_Z X/T - X'\tilde{u}\tilde{u}'X[\tilde{u}'P_Z \tilde{u}/T(\tilde{u}'\tilde{u})^2],$$

where the second equality holds by $\hat{D}(\tilde{\delta}) = 0$. Taking limits gives

$$\hat{J} \xrightarrow{p} Q + \alpha\Omega - \sigma_{V_u}\sigma'_{V_u}(\alpha/\sigma_u^2) = Q + \alpha\tilde{\Omega}.$$

Let $\hat{\alpha} = \tilde{u}'P_Z \tilde{u}/\tilde{u}'\tilde{u}$. Then we can form an estimator $\hat{\Sigma}$ of Σ as

$$\hat{\Lambda} = \hat{\sigma}_u^2\{(1 - \hat{\alpha})\hat{J} - \hat{\alpha}\hat{H}\}.$$

By $\hat{\alpha} \xrightarrow{p} \alpha$ it will follow that

$$\hat{\Lambda} \xrightarrow{p} \sigma_u^2\{(1 - \alpha)(Q + \alpha\tilde{\Omega}) - \alpha(1 - \alpha)Q\} = \Sigma.$$

Therefore, a consistent estimator of the asymptotic variance of the LIML estimator under weak instrument asymptotics is $\hat{\Lambda} = \hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1}$.

The asymptotic distribution of explicit bias corrected estimator $\check{\delta}$ can be derived in a similar way. Note that

$$\sqrt{T}(\check{\delta} - \delta_0) = (X'(P_Z - \frac{K}{T})X/T)^{-1}X'(P_Z - \frac{K}{T})u/\sqrt{T}.$$

Similarly to the above derivations

$$X'(P_Z - \frac{K}{T})X/T \xrightarrow{p} Q + \alpha\Omega - \alpha(Q + \Omega) = (1 - \alpha)Q = H.$$

Assume that $\sqrt{T}(K/T - \alpha) \rightarrow 0$, so that $(K/T - \alpha)X'u/\sqrt{T} \xrightarrow{p} 0$, we have

$$\begin{aligned} X'(P_Z - \frac{K}{T})u/\sqrt{T} &= X'(P_Z - \alpha I)u/\sqrt{T} + o_p(1) \\ &= (1 - \alpha)\Pi'Z'u/\sqrt{T} + V'(P_Z - \alpha I)u/\sqrt{T} + o_p(1) \\ &= (1 - \alpha)\Pi'Z'u/\sqrt{T} + \tilde{V}'(P_Z - \alpha I)u/\sqrt{T} \\ &\quad + \gamma u'(P_Z - \alpha I)u/\sqrt{T} + o_p(1). \end{aligned}$$

Note that the sum of first two terms following the last equality are identical to those that appear in the LIML derivation, so that the variance of their sum will converge to Σ . Furthermore, the last term will be uncorrelated with each of those terms if third moments of u_t are zero conditional on \tilde{V}_t . Therefore, under these conditions the asymptotic variance of $X'(P_Z - K/T)u/\sqrt{T}$ will be larger than Σ , and hence the asymptotic variance of the bias corrected estimator larger than that of LIML.

Appendix: Proofs of Theorems.

Throughout, let C denote a generic positive constant that may be different in different uses.

Proof of Lemma 1: We first prove the results for $\bar{\delta} = \delta_0$. The first conclusion follows from Khintchine's law of large numbers. For the second conclusion, note that $X'u/T = \Pi'Z'u/T + V'u/T$, that $V'u/T \xrightarrow{p} \sigma_{V_u}$ by Khintchine's law of large number. Also, by $E[\Pi'Z'u/T] = 0$ and $Var(\Pi'Z'u/T) = \Pi'Z'Z\Pi/T^2 \rightarrow 0$, Chebyshev's law of large numbers gives $\Pi'Z'u/T \xrightarrow{p} 0$, so the second conclusion follows by the triangle inequality. The third conclusion follows similarly.

Next, suppose for the moment that V is a scalar. As above,

$$E[u'P_Z V/T] = \text{tr}(P_Z E[Vu'])/T = \sigma_{V_u} \text{tr}(P_Z)/T = \sigma_{V_u} K/T.$$

Let p_{st} denote the s, t^{th} element of P_Z . Note that for the t^{th} unit vector e_t it follows by $I - P_Z$ positive semi-definite that $p_{tt} = e_t' P_Z e_t \leq e_t' e_t = 1$. Then for $\sigma_V^2 = \text{var}(V_t)$ and $A = E[\varepsilon_t^2 V_t^2] - 2\sigma_{V_u}^2 - \sigma_u^2 \sigma_V^2$,

$$\begin{aligned} E[(\varepsilon' P_Z v/T)^2] &= \frac{1}{T^2} \sum_{t,s,r,q} E[u_t p_{ts} V_s V_r p_{rq} u_q] = \frac{1}{T^2} \sum_t E[u_t^2 V_t^2] p_{tt}^2 + \frac{1}{T^2} \sum_{t \neq s} \sigma_{V_u}^2 p_{tt} p_{ss} \\ &\quad + \frac{1}{T^2} \sum_{t \neq s} \sigma_{V_u}^2 p_{ts}^2 + \frac{1}{T^2} \sum_{t \neq s} \sigma_u^2 \sigma_V^2 p_{ts} p_{st} \\ &= \frac{1}{T^2} A \sum_t p_{tt}^2 + \frac{\sigma_{V_u}^2}{T^2} \sum_{t,s} p_{tt} p_{ss} + \frac{\sigma_{V_u}^2 + \sigma_u^2 \sigma_V^2}{T^2} \sum_{t,s} p_{ts}^2 \\ &= \frac{A}{T^2} \sum_t p_{tt}^2 + \frac{\sigma_{V_u}^2}{T^2} \text{tr}(P_Z)^2 + \frac{\sigma_{V_u}^2 + \sigma_u^2 \sigma_V^2}{T^2} \sum_{t,x} Z_t'(Z'Z)^{-1} Z_s Z_s'(Z'Z)^{-1} Z_t \\ &= A \sum_t p_{tt}^2 / T^2 + \sigma_{V_u}^2 \frac{K^2}{T^2} + (\sigma_{V_u}^2 + \sigma_u^2 \sigma_V^2) \frac{K}{T^2}. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{var}(u'P_Z V/T) &= E[(u'P_Z V/T)^2] - \{E[u'P_Z V/T]\}^2 \\ &= A \sum_t p_{tt}^2 / T^2 + (\sigma_{\varepsilon v}^2 + \sigma_{\varepsilon}^2 \sigma_v^2) K / T^2 \leq A/T + C/T \rightarrow 0. \end{aligned}$$

The since $\text{var}(u'P_Z V/T)$ converges to zero and $E[u'P_Z V/T]$ converges to $\alpha \sigma_{V_u}^2$, gives the fifth conclusion. The fourth and sixth conclusions follow similarly.

Next, for $\bar{\delta} \neq \delta_0$, note that $X'\bar{u}/T - X'u/T = (X'X/T)(\bar{\delta} - \delta_0) \xrightarrow{p} 0$, so the second conclusion follows by the triangle inequality. The rest of the conclusions follow similarly. *Q.E.D.*

Proof of Theorem 1: By Q positive definite, $\min_{\|\Delta\|=1} \{\Delta'Q\Delta/[\Delta'(Q+\Omega)\Delta]\} = C > 0$. Further, for all Δ with $\|\Delta\| = 1$,

$$\Delta'(Q + \alpha\Omega)\Delta/(\Delta'(Q + \Omega)\Delta) = \alpha + (1 - \alpha)\Delta'Q\Delta/[\Delta'(Q + \Omega)\Delta] > \alpha + C.$$

Also, by standard arguments,

$$\sup_{\|\Delta\|=1} \{[\Delta'X'P_Z X\Delta/(\Delta'X'X\Delta)] - \Delta'(Q + \alpha\Omega)\Delta/(\Delta'(Q + \Omega)\Delta)\} \xrightarrow{p} 0.$$

Therefore, it follows that with probability approaching one, for all Δ with $\|\Delta\| = 1$,

$$\Delta' X' P_Z X \Delta / (\Delta' X' X \Delta) > \alpha + C/2.$$

Next, note that by convergence in probability of $u' P_Z u / T$, $u' P_Z X / T$, it follows that for M large enough and all δ with $\|\delta - \delta_0\| > M$, for $\Delta = (\delta - \delta_0) / \|\delta - \delta_0\|$,

$$\begin{aligned} \hat{S}(\delta) &= \frac{(u' P_Z u / T) / \|\delta - \delta_0\|^2 + 2(u' P_Z X / T)(\delta - \delta_0) / \|\delta - \delta_0\|^2 + \Delta' X' P_Z X \Delta / T}{(u' u / T) / \|\delta - \delta_0\|^2 + 2(u' X / T)(\delta - \delta_0) / \|\delta - \delta_0\|^2 + \Delta' X' X \Delta / T} \\ &\geq \Delta' X' P_Z X \Delta / (\Delta' X' X \Delta) - C/4 > \alpha + C/4. \end{aligned}$$

Then, since $\hat{S}(\delta_0) \xrightarrow{p} \alpha$, it follows that with probability approaching one,

$$\inf_{\|\delta - \delta_0\| \geq M} \hat{S}(\delta) > \hat{S}(\delta_0).$$

Also, w.p.a.1 $\hat{B}(\delta) > 0$ for all δ , so that $\hat{S}(\delta)$ is continuous, so that $\bar{\delta} = \arg \min_{\|\delta - \delta_0\| \leq M} \hat{S}(\delta)$ exists, and hence $\bar{\delta} = \delta_0$. It also follows by standard arguments that that $\bar{\delta} \xrightarrow{p} \delta_0$, giving consistency. *Q.E.D.*

The next result is useful for proving Theorem 2.

LEMMA 2: If (R_t, u_t) , $(t = 1, 2, \dots)$ are i.i.d., $E[|u_t|^p | R_t]$ is bounded for $p > 0$, $E[u_t | R_t] = 0$, $\text{var}(u_t | R_t) \geq C > 0$, a_{tT} , $(t = 1, \dots, T)$ are random variables depending only on (R_1, \dots, R_T) , with

$$\max_{t \leq T} |a_{tT}| \xrightarrow{p} 0, \quad \sum_{t=1}^T a_{tT}^2 \text{var}(u_t | R_t) \xrightarrow{p} \Psi > 0,$$

then

$$\sum_{t=1}^T a_{tT} u_t \xrightarrow{d} N(0, \Psi).$$

Proof: We proceed by verifying the hypotheses of Lemma 3 of Chamberlain (1986, "Notes on Semiparametric Regression."), denoted L3 henceforth. Note that $\text{var}(u_t | R_t) \leq C$ by $E[|u_t|^p | R_t]$ bounded, so that $\sum_{t=1}^T a_{tT}^2 \text{var}(u_t | R_t) \leq C \sum_{t=1}^T a_{tT}^2$. Therefore, with probability approaching one, $\sum_{t=1}^T a_{tT}^2 \geq C > 0$. Therefore equation 1) of L3 holds. Equation 2) of L3 is also satisfied, since

$$\left(\max_{t \leq T} a_{tT}^2 \right) / \sum_{j=1}^T a_{jT}^2 \leq \left(\max_{t \leq T} |a_{tT}| \right)^2 C$$

with probability approaching one. Also, equation (3) of L3 holds, since for $\Delta > 0$,

$$\begin{aligned} E \left[1(|u_t| \geq \Delta) u_t^2 | R_t \right] &= E \left[1 \left(\left| \frac{u_t}{\Delta} \right| \geq 1 \right) \left| \frac{u_t}{\Delta} \right|^2 | R_t \right] \Delta^2 \\ &\leq E \left[1 \left(\left| \frac{u_t}{\Delta} \right| \geq 1 \right) \left| \frac{u_t}{\Delta} \right|^p | R_t \right] \Delta^2 \leq E[|u_t|^p | R_t] \Delta^{2-p} \leq C \Delta^{2-p} \end{aligned}$$

which goes to zero as $\Delta \rightarrow \infty$. Also, equation 4) of L3 is satisfied by $\text{var}(u_t|R_t)$ bounded away from zero. Let $I_T = 1$ if $\sum_{t=1}^T a_{tT}^2 > 0$, $I_T = 0$ otherwise. Note that for $\sigma_t^2 = \text{var}(u_t|R_t)$

$$\sum_{t=1}^T a_{tT} u_t = (1 - I_T) \sum_{t=1}^T a_{tT} u_t + \left[I_T \sum_{t=1}^T a_{tT} u_t / \left(\sum_{t=1}^T a_{tT}^2 \sigma_t^2 \right)^{1/2} \right] \left(\sum_{t=1}^T a_{tT}^2 \sigma_t^2 \right)^{1/2}$$

The first term converges in probability to zero by $I_T = 1$ with probability approaching one. Also $\left(\sum_{t=1}^T a_{tT}^2 \sigma_t^2 \right)^{1/2} \xrightarrow{p} \Psi^{1/2}$, so by the Slutsky theorem and the conclusion of L3,

$$\sum_{t=1}^T a_{tT} u_t \xrightarrow{d} \Psi^{1/2} N(0, 1) = N(0, \Psi). Q.E.D.$$

Proof of Theorem 2: From the discussion in the text we see that it suffices to prove that $W'u/\sqrt{T} \xrightarrow{d} N(0, \Sigma)$. Also, by the Cramer-Wold device it suffices to prove that for any vector λ , $\lambda'W'u/\sqrt{T} \xrightarrow{d} N(0, \lambda'\Sigma\lambda)$, or equivalently that the conclusion holds when $G = 1$. Without changing notation we will assume that $X, Z\Pi$, and V are vectors, representing $X\lambda, Z\Pi\lambda$ and $V\lambda$ respectively. Let

$$a_{tT} = W_t = (1 - \alpha)Z_t\Pi + \tilde{V}'Z(Z'Z)^{-1}Z_t - \alpha\tilde{V}_t$$

By hypothesis, $\max_{t \leq T} |Z_t\Pi|/\sqrt{T} \rightarrow 0$. Also, by the Markov inequality,

$$\max_{t \leq T} |\tilde{V}_t|/\sqrt{T} = \left(\max_{t \leq T} |\tilde{V}_t|^4 / T^2 \right)^{1/4} \leq \left(\sum_{t=1}^T |\tilde{V}_t|^4 / T^2 \right)^{1/4} \xrightarrow{p} 0.$$

Also, by the Marcinkiewicz-Zygmund inequality, for $w_{st} = Z_s(Z'Z)^{-1}Z'_t$,

$$E \left[|V'Z(Z'Z)^{-1}Z_t|^p \right] = E \left[\left| \sum_{s=1}^T V_s w_{st} \right|^p \right] \leq CE \left[\sum_{s=1}^T V_s^2 w_{st}^2 \right]^{p/2}.$$

As shown above, $Z_t(Z'Z)^{-1}Z'_t \leq 1$, so that

$$\sum_{s=1}^T w_{st}^2 = \sum_{s=1}^T Z_t(Z'Z)^{-1}Z'_s Z_s(Z'Z)^{-1}Z_t = Z_t(Z'Z)^{-1}Z'_t \leq 1.$$

Hence $\left(\sum_{s=1}^T w_{st}^2 \right)^{-p/2} \geq 1$, so by Jensen's inequality,

$$\begin{aligned} E \left[\left| \sum_{s=1}^T V_s^2 w_{st}^2 \right|^{p/2} \right] &\leq E \left[\left| \sum_{s=1}^T V_s^2 w_{st}^2 / \sum_{s=1}^T w_{st}^2 \right|^{p/2} \right] \leq E \left[\sum_{s=1}^T (|V_s|^2)^{p/2} w_{st}^2 / \sum_{s=1}^T w_{st}^2 \right] \\ &\leq \sum_{s=1}^T E \left[|V_s|^p \right] w_{st}^2 / \sum_{s=1}^T w_{st}^2 \leq C. \end{aligned}$$

Combining the last two equations gives $E \left[|V'Z(Z'Z)^{-1}Z_t|^p \right] \leq C$. Thus, by the Markov inequality, $\sum_{s=1}^T |V'Z(Z'Z)^{-1}Z_t|^p / T$ is bounded in probability, and

$$\begin{aligned} & \max_{t \leq T} |V'Z(Z'Z)^{-1}Z_t| / \sqrt{T} \leq \left(\sum_{t=1}^T |V'Z(Z'Z)^{-1}Z_t|^p / T^{p/2} \right)^{1/p} \\ & = \left(T^{1-p/2} \sum_{t=1}^T |V'Z(Z'Z)^{-1}Z_t|^p / T \right)^{1/p} \xrightarrow{p} 0. \end{aligned}$$

Then, by the triangle inequality,

$$\max_{t \leq T} |a_{tT}| / \sqrt{T} \xrightarrow{p} 0.$$

Next, note that by Lemma 1 and the law of large numbers,

$$\begin{aligned} \tilde{V}'(P_Z - \alpha I)^2 \tilde{V} / T &= (1 - 2\alpha) \tilde{V}' P_Z \tilde{V} / T + \alpha^2 \tilde{V}' \tilde{V} / T \xrightarrow{p} \alpha(1 - \alpha) \tilde{\Omega}, \\ \Pi' Z' (P_Z - \alpha I) \tilde{V} / T &= (1 - \alpha) \Pi' Z' u / T \xrightarrow{p} 0. \end{aligned}$$

Then by the triangle inequality,

$$\begin{aligned} W'W / T &= (1 - \alpha)^2 \Pi' Z' Z \Pi / T + 2(1 - \alpha) \Pi' Z' (P_Z - \alpha I) \tilde{V} / T \\ &\quad + \tilde{V}' (P_Z - \alpha I)^2 \tilde{V} / T \xrightarrow{p} \Sigma / \sigma_u^2. \end{aligned}$$

Since $\text{var}(u_t | \tilde{V}) = \sigma_u^2$, we have

$$\sum_{t=1}^T a_{tT}^2 \text{var}(u_t | \tilde{V}_t) / T = \sigma_u^2 W'W / T \xrightarrow{p} \Sigma.$$

It then follows by the conclusion of Lemma 2 that

$$\frac{W'u}{\sqrt{T}} = \frac{\sum_{t=1}^T a_{tT} u_t}{\sqrt{T}} \xrightarrow{d} N(0, \Sigma).$$