Estimation with Many Instrumental Variables
Author(s): Christian Hansen, Jerry Hausman and Whitney Newey
Source: *Journal of Business & Economic Statistics*, Vol. 26, No. 4 (Oct., 2008), pp. 398-422
Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association
Stable URL: https://www.jstor.org/stable/27639001
Accessed: 17-11-2018 16:59 UTC

Estimation With Many Instrumental Variables

Christian Hansen
Graduate School of Business, University of Chicago, Chicago, IL 60637 (*chansen1@chicagoGSB.edu*)

Jerry Hausman
Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139 (*jhausman@mit.edu*)

Whitney Newey
Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139 (*wnewey@mit.edu*)

Using many valid instrumental variables has the potential to improve efficiency but makes the usual inference procedures inaccurate. We give corrected standard errors, an extension of Bekker to nonnormal disturbances, that adjust for many instruments. We find that this adjustment is useful in empirical work, simulations, and in the asymptotic theory. Use of the corrected standard errors in t-ratios leads to an asymptotic approximation order that is the same when the number of instrumental variables grows as when the number of instruments is fixed. We also give a version of the Kleibergen weak instrument statistic that is robust to many instruments.

KEY WORDS: Inference; Many instruments; Standard errors; Weak instruments.

1. INTRODUCTION

Empirical applications of instrumental variables estimation often give imprecise results. Using many valid instrumental variables can improve precision. For example, as we show, using all 180 instruments in the Angrist and Krueger (1991) schooling application gives tighter correct confidence intervals than using three instruments. An important problem with using many instrumental variables is that conventional asymptotic approximations may provide poor approximations to the sampling distributions of the resulting estimators. Two-stage least squares (2SLS) is well known to have large biases when many instruments are used. The limited information maximum likelihood (LIML henceforth) or Fuller (1977, FULL henceforth) estimators correct this bias, but the usual standard errors are too small.

We give corrected standard errors (CSE) that improve upon the usual ones, leading to a better normal approximation to t-ratios under many instruments. The CSE are an extension of those of Bekker (1994) that allow for non-Gaussian disturbances. We show that the normal approximation with FULL and CSE is asymptotically correct with nonnormal disturbances under a variety of many instrument asymptotics, including the many instrument sequence of Kunitomo (1980), Morimune (1983), and Bekker (1994) and the many weak instruments sequence of Chao and Swanson (2002, 2003, 2004, 2005) and Stock and Yogo (2005b). We also find that there is no penalty for many instruments in the rate of approximation of the distribution of t-ratios when the CSE are used and an additional condition is satisfied. That is, the rate at which the distribution of the t-ratio approaches its standard normal limit is the same as with a fixed number of instruments. In addition, we give a version of the Kleibergen (2002) test statistic that is valid under many instruments, as well as under weak instruments.

We carry out a wide range of simulations to check the asymptotic approximations. We find that FULL with the CSE gives confidence intervals with actual coverage quite close to nominal. We also show that LIML with the CSE has identical asymptotic properties to FULL and performs quite well in our simulations, as in those of Hahn and Inoue (2002). Our results also demonstrate that the concentration parameter (which can be estimated) provides a better measure of accuracy for standard inference with FULL or LIML than the F-statistic, R^2, or other statistics previously considered in the literature.

In relation to previous work, the CSE, the validity of Bekker (1994) standard errors under many weak instrument asymptotics, the rate of approximation results, and our many instrument view of the Angrist and Krueger (1991) application appear to be novel. The limiting distribution results build on previous work. For many instrument asymptotics we generalize LIML results of Kunitomo (1980), Morimune (1983), Bekker (1994), and Bekker and van der Ploeg (2005) to FULL, disturbances that are not Gaussian, and general instruments. Our results also generalize recent results of Anderson, Kunitomo, and Matsushita (2005) to many weak instruments, who had generalized results from an earlier version of this article by relaxing a conditional moment restriction. We also combine and generalize results of Chao and Swanson (2002, 2003, 2005) and Stock and Yogo (2005b) by relaxing some kurtosis restrictions of Chao and Swanson (2003) and allowing a wider variety of sequences of instruments and concentration parameter than Stock and Yogo (2005b). Our theoretical results make use of some inequalities in Chao and Swanson (2004).

Hahn and Hausman (2002) gave a test for weak instruments and Hahn, Hausman, and Kuersteiner (2004) showed that FULL performs well under weak instruments. Also, the random effects estimator of Chamberlain and Imbens (2004) leads to accurate inference with many instruments. Recently Andrews and Stock (2006) derived asymptotic power envelopes for tests under several cases of many weak instrument asymptotics with Gaussian disturbances. We consider cases where the square root of the number of instruments grows more slowly than the concentration parameter. There it turns out that Wald tests using the CSE attain the power envelope. We also consider cases where the number of instruments grows as fast as the sample size, which is not covered by Andrews and Stock (2006).

Journal of Business & Economic Statistics
October 2008, Vol. 26, No. 4
DOI 10.1198/073500108000000024

The remainder of the article is organized as follows. In the next section, we briefly present the model and estimators that we will consider. We reexamine the Angrist and Krueger (1991) study of the returns to schooling in Section 3 and give a variety of simulation results in Section 4. Section 5 contains asymptotic results and Section 6 concludes.

2. MODELS AND ESTIMATORS

The model we consider is given by

$$\underset{T\times 1}{y} = \underset{T\times G}{X}\,\underset{G\times 1}{\delta_0} + \underset{T\times 1}{u},$$
$$X = \Upsilon + V,$$

where T is the number of observations, G is the number of right-side variables, Υ is a matrix of observations on the reduced form, and V is the matrix of reduced form disturbances. For the asymptotic approximations, the elements of Υ will be implicitly allowed to depend on T, although we suppress dependence of Υ on T for notational convenience. Estimation of δ_0 will be based on a $T \times K$ matrix Z of instrumental variable observations.

This model allows for Υ to be a linear combination of Z, that is, $\Upsilon = Z\pi$ for some $K \times G$ matrix π. Furthermore, columns of X may be exogenous, with the corresponding column of V being zero. The model also allows for Z to be functions meant to approximate the reduced form. For example, let Υ_t' and Z_t' denote the tth row (observation) of Υ and Z, respectively. We could have $\Upsilon_t = f_0(w_t)$ be an unknown function of a vector w_t of underlying instruments and $Z_t = (p_{1K}(w_t), \ldots, p_{KK}(w_t))'$ for approximating functions $p_{kK}(w)$, such as power series or splines. In this case linear combinations of Z_t may approximate the unknown reduced form, for example, as in Donald and Newey (2001).

It is well known that variability of Υ relative to V is important for the properties of instrumental variable (IV) estimators. For $G = 1$ this feature is well described by

$$\mu_T^2 = \sum_{t=1}^{T} \Upsilon_t^2 / E[V_t^2].$$

This concentration parameter plays a central role in the theory of IV estimators. The distribution of the estimators depends on μ_T^2, with the convergence rate being $1/\mu_T$ and the accuracy of the usual asymptotic approximation depending crucially on the size of μ_T^2.

To describe the estimators, let $P = Z(Z'Z)^- Z'$ where A^- denotes any symmetric generalized inverse of a symmetric matrix A, that is, A^- is symmetric and satisfies $AA^-A = A$. We consider estimators of the form

$$\hat{\delta} = (X'PX - \hat{\alpha}X'X)^{-1}(X'Py - \hat{\alpha}X'y)$$

for some choice of $\hat{\alpha}$. This class includes all of the familiar k-class estimators except the least squares estimator. Special cases of these estimators are two-stage least squares (2SLS), where $\hat{\alpha} = 0$, and LIML, where $\hat{\alpha} = \tilde{\alpha}$ and $\tilde{\alpha}$ is the smallest eigenvalue of the matrix $(\bar{X}'\bar{X})^{-1}\bar{X}'P\bar{X}$ for $\bar{X} = [y, X]$. FULL is also a member of this class of estimators, where $\hat{\alpha} = [\tilde{\alpha} - (1 - \tilde{\alpha})C/T]/[1 - (1 - \tilde{\alpha})C/T]$ for some constant C. FULL has moments of all orders, is approximately mean unbiased for $C = 1$, and is second-order admissible for $C \geq 4$ under standard large sample asymptotics.

For inference we consider an extension of the Bekker (1994) standard errors to nonnormality and estimators other than LIML. Let $u(\delta) = y - X\delta$, $\hat{\sigma}_u^2(\delta) = \hat{u}(\delta)'\hat{u}(\delta)/(T - G)$, $\tilde{\alpha}(\delta) = u(\delta)'Pu(\delta)/u(\delta)'u(\delta)$, $\hat{\Upsilon} = PX$, $\tilde{X}(\delta) = X - \hat{u}(\delta)(\hat{u}(\delta)'X)/\hat{u}(\delta)'\hat{u}(\delta)$, $\hat{V}(\delta) = (I - P)\tilde{X}(\delta)$, $\kappa_T = \sum_{t=1}^T p_{tt}^2/K$, $\tau_T = K/T$,

$$\hat{H}(\delta) = X'PX - \tilde{\alpha}(\delta)X'X,$$
$$\hat{\Sigma}_B(\delta) = \hat{\sigma}_u^2(\delta)\big\{(1 - \tilde{\alpha}(\delta))^2\tilde{X}(\delta)'P\tilde{X}(\delta) + \tilde{\alpha}(\delta)^2\tilde{X}(\delta)'(I - P)\tilde{X}(\delta)\big\},$$
$$\hat{\Sigma}(\delta) = \hat{\Sigma}_B(\delta) + \hat{A}(\delta) + \hat{A}(\delta)' + \hat{B}(\delta),$$
$$\hat{A}(\delta) = \sum_{t=1}^{T}(p_{tt} - \tau_T)\hat{\Upsilon}_t\left[\sum_{t=1}^{T}\hat{u}_t(\delta)^2\hat{V}_t(\delta)/T\right]',$$
$$\hat{B}(\delta) = K(\kappa_T - \tau_T)\sum_{t=1}^{T}(u_t(\delta)^2 - \hat{\sigma}_u^2(\delta))\hat{V}_t(\delta)\hat{V}_t(\delta)' \,/[T(1 - 2\tau_T + \kappa_T\tau_T)].$$

The asymptotic variance estimator is given by

$$\hat{\Lambda} = \hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1}, \qquad \hat{H} = \hat{H}(\hat{\delta}), \qquad \hat{\Sigma} = \hat{\Sigma}(\hat{\delta}).$$

When $\hat{\delta}$ is the LIML estimator, $\hat{H}^{-1}\hat{\Sigma}_B(\hat{\delta})\hat{H}^{-1}$ is identical to the Bekker (1994) variance estimator. The other terms in $\hat{\Lambda}$ account for third and fourth moment terms that are present with some forms of nonnormality. In general $\hat{\Lambda}$ is a "sandwich" formula, with $\hat{H}$ being a Hessian term.

The variance estimator $\hat{\Lambda}$ can be quite different than the usual one $\hat{\sigma}_u^2\hat{H}^{-1}$ even when K is small relative to T. This occurs because $\hat{H}$ is close to the sum of squares of predicted values for the reduced form regressions and $\hat{\Sigma}_B(\delta)$ depends on sums of squares of residuals. When the reduced form r-squared is small, the sum of squared residuals will tend to be quite large relative to $\hat{H}$, leading to $\hat{\Sigma}_B(\delta)$ being larger than $\hat{H}$. In contrast, the adjustments for nonnormality $\hat{A}(\hat{\delta})$ and $\hat{B}(\hat{\delta})$ will tend to be quite small when K is small relative to T, which is typical in applications. Thus we expect that in applied work the Bekker (1994) standard errors and CSE will often give very similar results. Also, Bekker (1994) standard errors will be consistent under many weak instrument asymptotics where K/T goes to zero.

As shown by Dufour (1997), if the parameter set is allowed to include values where $\Upsilon = 0$, then a correct confidence interval for a structural parameter must be unbounded with probability 1. Hence, confidence intervals formed using the CSE cannot be correct. Also, under the weak instrument sequence of Staiger and Stock (1997), the CSE confidence intervals will not be correct, that is, they are not robust to weak instruments. These considerations motivate a statistic that is asymptotically correct with weak or many instruments.

Such a statistic can be obtained by modifying the Lagrange multiplier statistic of Kleibergen (2002) and Moreira (2001). For any δ let

$$\widehat{LM}(\delta) = u(\delta)'P\tilde{X}(\delta)\hat{\Sigma}(\delta)^{-1}\tilde{X}(\delta)'Pu(\delta).$$

This statistic differs from previous ones in using $\hat{\Sigma}(\delta)^{-1}$ in the middle. Its validity does not depend on correctly specifying the reduced form. The statistic $\widehat{LM}(\delta)$ will be asymptotically distributed as $\chi^2(G)$ when $\delta = \delta_0$ under both many and weak instruments. Confidence intervals for δ_0 can be formed from $\widehat{LM}(\delta)$ by inverting it. Specifically, for the $1-\alpha$ quantile q of a $\chi^2(G)$ distribution, an asymptotic $1-\alpha$ confidence interval is $\{\delta : \widehat{LM}(\delta) \leq q\}$. As recently shown by Andrews and Stock (2006), the conditional likelihood ratio test of Moreira (2003) is also correct with weak and many weak instruments, though apparently not under many instruments, where K grows as fast as T. For brevity we omit a description of this statistic and the associated asymptotic theory.

We suggest that the CSE are useful despite their lack of robustness to weak instruments. Standard errors provide a simple measure of uncertainty associated with an estimate. The confidence intervals based on $\widehat{LM}(\delta)$ are more difficult to compute. Also, as we discuss below, the t-ratios for FULL based on the CSE provide a good approximation over a wide range of empirically relevant cases we considered. This observation might justify viewing the parameter space as being bounded away from $\Upsilon = 0$, thus overcoming the strict Dufour (1997) critique. Or, one might simply view that our theoretical and simulation results are relevant enough for applications to warrant using the CSE.

It does seem wise to check for weak instruments in practice. One could use the Hahn and Hausman (2002) test. One could also compare a Wald test based on the CSE with a test based on $\widehat{LM}(\delta)$. One could also develop versions of the Stock and Yogo (2005a) tests for weak instruments that are based on the CSE.

Because the concentration parameter is important for the properties of the estimators it is useful to have an estimate of it for the common case with one endogenous right-side variable. For $G = 1$ let $\hat{\sigma}_V^2 = \hat{V}'\hat{V}/(T-K)$. An estimator of μ_T^2 is

$$\hat{\mu}_T^2 = \hat{X}'\hat{X}/\hat{\sigma}_V^2 - K = K(\hat{F}-1),$$

where $\hat{F} = (\hat{X}'\hat{X}/K)/[\hat{V}'\hat{V}/(T-K)]$ is the reduced form F-statistic. This estimator is consistent in the sense that under many instrument asymptotics

$$\frac{\hat{\mu}_T^2}{\mu_T^2} \xrightarrow{p} 1.$$

In the general case with one endogenous right-side and other exogenous right-side variables we take

$$\hat{\mu}_T^2 = (K-G+1)(\tilde{F}-1),$$

where $\tilde{F}$ is the reduced form F-statistic for the variables in Z that are excluded from X.

3. QUARTER OF BIRTH AND RETURNS TO SCHOOLING

A motivating empirical example is provided by the Angrist and Krueger (1991) study of the returns to schooling using quarter of birth as an instrument. We consider data drawn from the 1980 U.S. Census for males born in 1930–1939. The model includes a constant and year and state dummies. We report results for 3 instruments and for 180 instruments. Figures 1–4 are graphs of confidence intervals at different significance levels using several different methods. The confidence intervals we consider are based on 2SLS with the usual (asymptotic) standard errors, FULL with the usual standard errors, and FULL with the CSE. We take as a standard of comparison our version of the Kleibergen (2002) confidence interval (denoted K in the graphs), which is robust to weak instruments, many instruments, and many weak instruments.

Figure 1 shows that with three excluded instruments (two overidentifying restrictions), 2SLS and K intervals are very similar. The main difference seems to be a slight horizontal shift. Because the K intervals are centered about the LIML estimator, this shift corresponds to a slight difference in the LIML and 2SLS estimators. This difference is consistent with 2SLS having slightly higher bias than LIML. Figure 2 shows that with 180 excluded instruments (179 overidentifying restrictions), the confidence intervals are quite different. In particular, there is a much more pronounced shift in the 2SLS location, as well as smaller dispersion. These results are consistent with a larger bias in 2SLS resulting from many instruments.

Figure 3 compares the confidence interval for FULL based on the usual standard error formula for 180 instruments with the K interval. Here we find that the K interval is wider than the usual one. In Figure 4, we compare FULL with CSE to K, finding that the K interval is nearly identical to the one based on the CSE.

Comparing Figures 1 and 4, we find that the CSE interval with 180 instruments is substantially narrower than the intervals with 3 instruments. Thus, in this application we find that using the larger number of instruments leads to more precise inference, as long as FULL and the CSE are used. These graphs are consistent with direct calculations of estimates and standard errors. The 2SLS estimator with 3 instruments is .1077 with standard error .0195 and the FULL estimator with 180 instruments is .1063 with CSE .0143. A precision gain is evident in the decrease in the CSE obtained with the larger number of instruments. These results are also consistent with Donald and Newey's (2001) finding that using 180 instruments gives smaller estimated asymptotic mean square error for LIML than using just 3. Furthermore, Cruz and Moreira (2005) also found that 180 instruments are informative when extra covariates are used.

We also find that the CSE and the standard errors of Bekker (1994) are nearly identical in this application. Adding significant digits, with 3 instruments the CSE is .0201002 whereas the Bekker (1994) standard error is .0200981, and with 180 instruments the CSE is .0143316 and the Bekker (1994) standard error is .0143157. They are so close in this application because even when there are 179 overidentifying restrictions, the number of instruments is very small relative to the sample size.

These results are interesting because they occur in a widely cited application. However, they provide limited evidence of the accuracy of the CSE because they are only an example. They result from one realization of the data, and so could have occurred by chance. Real evidence is provided by a Monte Carlo study.

We based the study on the application to help make it empirically relevant. The design uses the same sample size as the application, and we fixed values for the instrumental variables and other exogenous variables at the sample values, for example, as in Staiger and Stock's (1997) design for dummy variable instruments. The data were generated from a two-equation triangular simultaneous equations system with structural equation

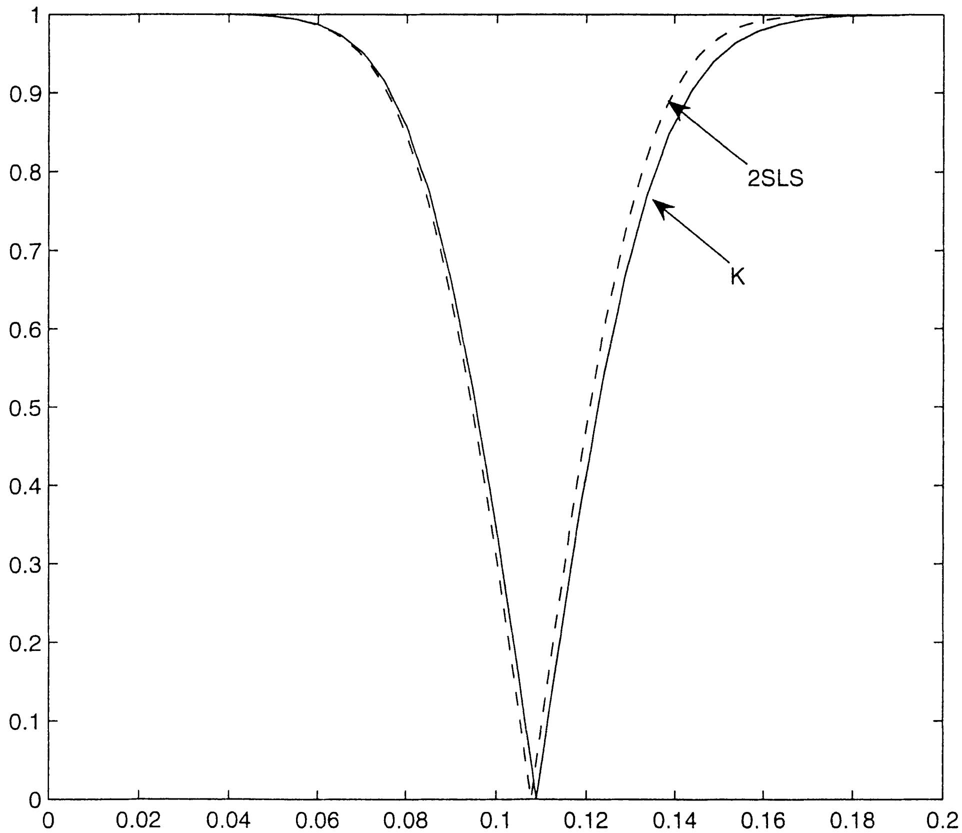

Figure 1. Three instruments: K and 2SLS.

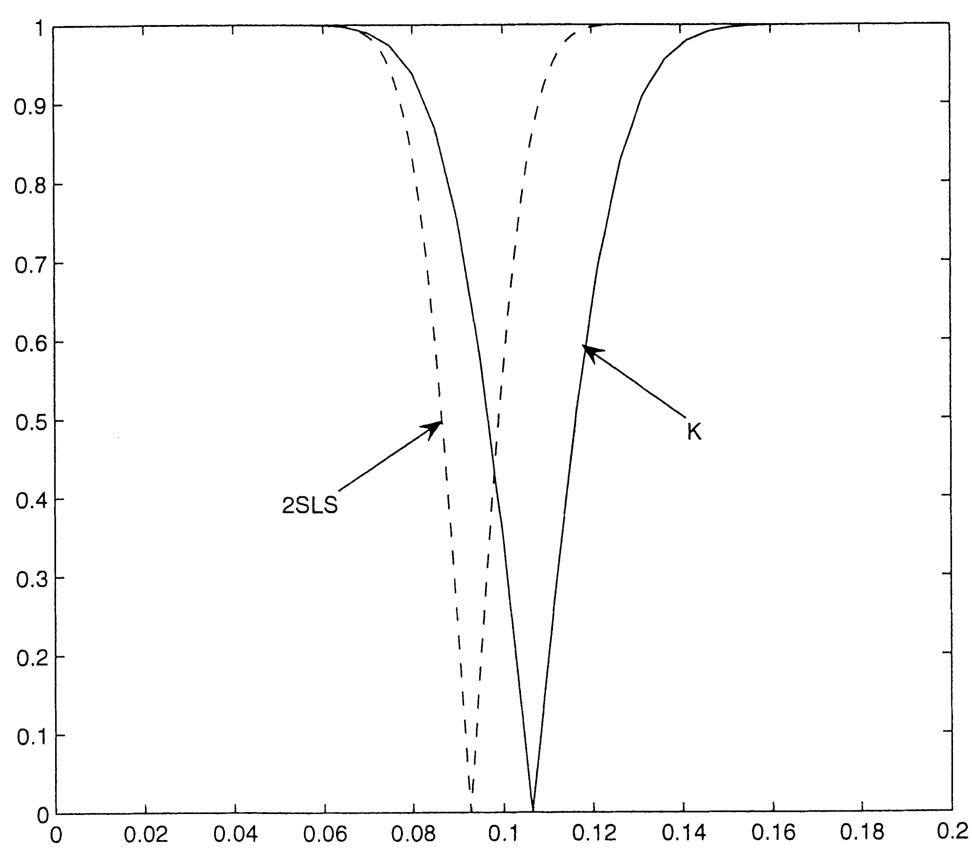

Figure 2. 180 instruments: K and 2SLS.

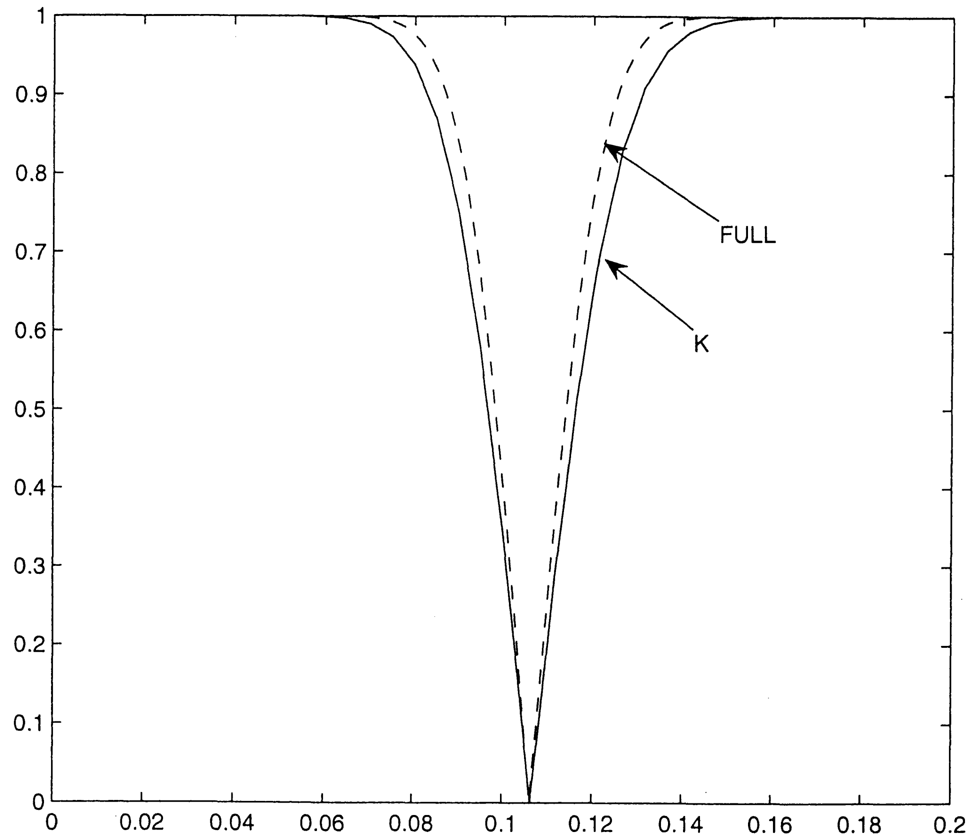

Figure 3. 180 instruments: K and Fuller.

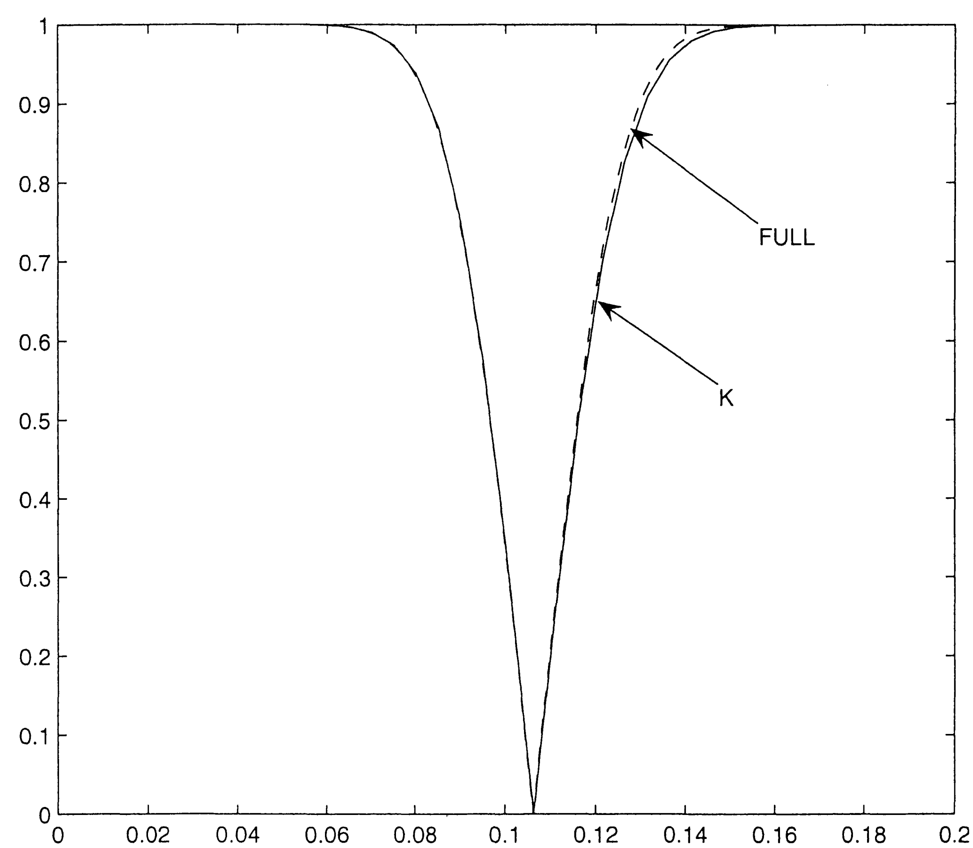

Figure 4. 180 instruments: K and Fuller with Bekker standard errors.

as in the empirical application and a reduced form consisting of a regression of schooling on all of the instruments, including the covariates from the structural equation. The structural parameters were set equal to their LIML estimated values from the 3 instruments case. The disturbances were homoscedastic Gaussian with (bivariate) variance matrix for each observation equal to the estimate from the application. Because the design has parameters equal to estimates, this Monte Carlo study could be considered a parametric bootstrap.

We carried out two experiments, one with 3 excluded instruments and one with 180 excluded instruments. In each case the reduced form coefficients were set so that the concentration parameter for the excluded instruments was equal to the unbiased estimator from the application. With 3 excluded instruments the concentration parameter value was set equal to the value of the consistent estimator $\hat{\mu}_T^2 = 95.6$ from the data and with 180 excluded instruments the value was set to $\hat{\mu}_T^2 = 257$.

Table 1 reports the results of this experiment, giving relative bias, mean square error, and rejection frequencies for nominal 5% level tests concerning the returns to schooling coefficient. Similar results hold for the median and interquartile range. We are primarily interested in accuracy of inference and not in whether confidence intervals are close to each other, as they are in the application, so we focus on rejection frequencies. We find that with 3 excluded instruments all of rejection frequencies are quite close to their nominal values, including those for 2SLS. We also find that with 180 instruments, the significance levels of the standard 2SLS, LIML, and FULL tests are quite far from their nominal values, but that with CSE the LIML and FULL confidence intervals have the right level. Thus, in this Monte Carlo study we find evidence that using CSE takes care of whatever inference problem might be present in these data.

These results provide a somewhat different view of the Angrist and Krueger (1991) application than do Bound, Jaeger, and Baker (1996) and Staiger and Stock (1997). They viewed the 180 instrument case as a *weak* instrument problem, apparently due to the low F-statistic, of about 3, for the excluded instruments. In contrast, we find that correcting for *many* instruments, by using FULL with CSE, fixes the inference problem.

Table 1. Simulation results

	Bias/β	RMSE	Size
A. 3 instruments, $\mu_T^2 = 95.6$			
2SLS	−.0021	.0217	.056
LIML	.0052	.0222	.056
CSE			.054
FULL	.0010	.0219	.057
CSE			.056
Kleibergen			.059
B. 180 instruments, $\mu_T^2 = 257$			
2SLS	−.1440	.0168	.318
LIML	−.0042	.0168	.133
CSE			.049
FULL	−.0063	.0168	.132
CSE			.049
Kleibergen			.051

NOTE: Males born 1930–1939. 1980 IPUMS $T = 329{,}509$, $\beta = .0953$.

We would not tend to find this result with weak instruments, because CSE do not correct for weak instruments as illustrated in the simulation results below. These results are reconciled by noting that a low F-statistic does not mean that FULL with CSE is a poor approximation. As we will see, a better criterion for LIML or FULL is the concentration parameter. In the Angrist and Krueger (1991) application we find estimates of the concentration parameter that are quite large. With 3 excluded instruments $\hat{\mu}_T^2 = 95.6$ and with 180 excluded instruments $\hat{\mu}_T^2 = 257$. Both of these are well within the range where we find good performance of FULL and LIML with CSE in the simulations reported below.

4. SIMULATIONS

To gain a broader view of the behavior of LIML and FULL with the CSE we consider the weak instrument limit of the FULL and LIML estimators and t-ratios with CSE under the Staiger and Stock (1997) asymptotics. This limit is obtained by letting the sample size go to infinity while holding the concentration parameter fixed. The limits of CSE and the Bekker (1994) standard errors coincide under this sequence because $K/T \longrightarrow 0$. As shown in Staiger and Stock (1997), these limits provide excellent approximations to small sample distributions. Furthermore, it seems very appropriate for microeconometric settings, where the sample size is often quite large relative to the concentration parameter.

Tables 2–5 give results for the median, interquartile range, and rejection frequencies for nominal 5% level tests based on the CSE and the usual asymptotic standard error for FULL and LIML, for a range of numbers of instruments K; concentration parameters μ_T^2; and values of the correlation coefficient ρ between u_t and V_t. These three parameters completely determine the weak instrument limiting distribution of t-ratios. Tables 2–5 give results for $\rho = 0$, $\rho = .2$, $\rho = .5$, and $\rho = .8$, respectively. Each table contains results for several different numbers of instruments and values of the concentration parameter.

Looking across the tables, there are a number of striking results. We find that LIML is nearly median unbiased for small values of the concentration parameter in all cases. This bias does increase somewhat in ρ and K, but even in the most extreme case we consider, with $\rho = .8$ and $K = 32$, the bias is virtually eliminated with a μ^2 of 16. Also, the bias is small when μ^2 is 8 in almost every case. When we look at FULL, we see that it is more biased than LIML but that it is considerably less dispersed. The difference in dispersion is especially pronounced for low values of the concentration parameter, though FULL is less dispersed than LIML in all cases.

The results for rejection frequencies are somewhat less clear-cut than the results for size and dispersion. In particular, the rejection frequencies tend to depend much more heavily on the value of K and ρ than do the results for median bias or dispersion. For LIML, the rejection frequencies when the CSE are used are quite similar to the rejection frequencies when the usual asymptotic variance is used for small values of K, but the CSE perform much better for moderate and large K, indicating that using the CSE with LIML will generally be preferable. FULL with CSE performs better in some cases and worse in others than FULL with the conventional standard errors when

Table 2. Weak instrument limit of LIML and Fuller—$\rho = 0$

			LIML				Fuller			
ρ	K	μ^2	Median	IQR	Bekker	Standard	Median	IQR	Bekker	Standard
0	1	1	.001	1.318	.001	.001	.000	.486	.001	.001
0	1	2	.001	1.000	.001	.001	.000	.502	.002	.001
0	1	4	.000	.708	.004	.004	.000	.488	.005	.003
0	1	8	−.001	.491	.010	.010	−.001	.418	.011	.009
0	1	16	.000	.342	.023	.023	.000	.319	.024	.020
0	1	32	.000	.240	.035	.035	.000	.232	.036	.033
0	1	64	.000	.169	.042	.042	.000	.167	.042	.041
0	2	1	−.002	1.418	.001	.001	−.001	.659	.001	.001
0	2	2	.000	1.099	.002	.002	.000	.629	.002	.002
0	2	4	.000	.775	.004	.005	.000	.560	.006	.005
0	2	8	.000	.525	.011	.013	.000	.450	.013	.012
0	2	16	.000	.355	.023	.026	.000	.331	.024	.023
0	2	32	.000	.244	.036	.038	.000	.236	.036	.036
0	2	64	.000	.171	.043	.045	.000	.168	.043	.043
0	4	1	.002	1.528	.002	.003	.001	.834	.003	.003
0	4	2	−.002	1.227	.003	.005	−.002	.769	.004	.005
0	4	4	.000	.879	.006	.009	.000	.656	.007	.009
0	4	8	−.001	.581	.012	.018	−.001	.500	.014	.017
0	4	16	.000	.377	.023	.033	.000	.352	.025	.030
0	4	32	.000	.252	.035	.044	.000	.244	.036	.041
0	4	64	.000	.173	.043	.048	.000	.171	.043	.046
0	8	1	−.001	1.634	.005	.012	.000	1.004	.006	.012
0	8	2	.001	1.360	.006	.015	.000	.917	.008	.015
0	8	4	.001	1.011	.009	.022	.001	.774	.010	.021
0	8	8	.001	.669	.015	.034	.001	.578	.017	.032
0	8	16	−.001	.419	.025	.049	−.001	.391	.026	.045
0	8	32	.000	.268	.036	.056	.000	.260	.036	.053
0	8	64	.000	.179	.044	.056	.000	.176	.044	.054
0	16	1	−.002	1.720	.010	.046	−.002	1.164	.011	.047
0	16	2	.000	1.496	.011	.051	.000	1.074	.013	.052
0	16	4	−.001	1.170	.014	.060	−.001	.915	.015	.059
0	16	8	−.002	.793	.019	.073	−.002	.686	.020	.070
0	16	16	−.001	.486	.026	.085	.000	.452	.028	.080
0	16	32	.000	.295	.036	.082	.000	.285	.036	.077
0	16	64	.000	.189	.043	.069	.000	.186	.043	.067
0	32	1	.002	1.795	.017	.129	.001	1.320	.019	.134
0	32	2	.001	1.622	.018	.134	.001	1.232	.020	.137
0	32	4	−.001	1.334	.020	.141	−.001	1.076	.022	.143
0	32	8	.000	.950	.023	.149	.000	.826	.025	.147
0	32	16	.000	.590	.029	.153	.000	.549	.030	.149
0	32	32	.000	.343	.036	.133	.000	.331	.037	.128
0	32	64	.000	.208	.043	.098	.000	.204	.043	.095

K is small but clearly dominates for K large. The results also show that for small values of ρ, the rejection frequencies for LIML and FULL tend to be smaller than the nominal value, whereas the frequencies tend to be larger than the nominal value for large values of ρ.

An interesting and useful result is that both LIML and FULL with the CSE perform reasonably well for all values of K and ρ in cases where the concentration parameter is 32 or higher. In these cases, the rejection frequency for LIML varies between .035 and .06, and the rejection frequency for FULL varies between .035 and .070. These results suggest that the use of LIML or FULL with the CSE and the asymptotically normal approximation should be adequate in situations where the concentration parameter is around 32 or greater, even though in many of these cases the F-statistic takes on small values.

These results are also consistent with recent Monte Carlo work of Davidson and MacKinnon (2006). From careful examination of their graphs it appears that with few instruments the bias of LIML is very small once the concentration parameter exceeds 10, and that the variance of LIML is quite small once the concentration parameter exceeds 20.

To see which cases might be empirically relevant, we summarize values of K and estimates of μ^2 and ρ from some empirical studies. We considered all microeconomic studies that contain sufficient information to allow estimation of these quan-

Table 3. Weak instrument limit of LIML and Fuller—$\rho = .2$

			LIML				Fuller			
ρ	K	μ^2	Median	IQR	Bekker	Standard	Median	IQR	Bekker	Standard
.2	1	1	.084	1.307	.002	.002	.162	.484	.048	.003
.2	1	2	.038	.989	.004	.004	.119	.498	.033	.005
.2	1	4	.010	.706	.009	.009	.065	.483	.021	.009
.2	1	8	.000	.490	.016	.016	.027	.414	.021	.016
.2	1	16	−.001	.343	.026	.026	.012	.318	.029	.025
.2	1	32	.000	.240	.036	.036	.006	.232	.038	.035
.2	1	64	.000	.169	.044	.044	.003	.166	.044	.042
.2	2	1	.096	1.402	.004	.005	.146	.650	.009	.005
.2	2	2	.054	1.088	.006	.007	.108	.619	.011	.008
.2	2	4	.019	.771	.010	.012	.062	.551	.015	.013
.2	2	8	.003	.524	.017	.020	.028	.445	.022	.020
.2	2	16	.000	.354	.027	.030	.013	.328	.030	.029
.2	2	32	.000	.244	.036	.039	.006	.236	.038	.038
.2	2	64	.000	.170	.043	.044	.003	.167	.043	.043
.2	4	1	.117	1.507	.008	.012	.146	.821	.012	.014
.2	4	2	.071	1.209	.010	.015	.108	.756	.014	.017
.2	4	4	.030	.869	.013	.021	.064	.642	.018	.022
.2	4	8	.005	.578	.019	.028	.028	.492	.023	.028
.2	4	16	.000	.376	.028	.037	.013	.349	.031	.036
.2	4	32	.000	.251	.036	.044	.006	.242	.038	.043
.2	4	64	.000	.173	.043	.048	.003	.170	.044	.047
.2	8	1	.133	1.609	.014	.033	.151	.987	.019	.037
.2	8	2	.091	1.346	.016	.037	.117	.902	.021	.040
.2	8	4	.047	1.000	.019	.042	.074	.758	.023	.044
.2	8	8	.012	.661	.023	.049	.034	.565	.027	.049
.2	8	16	.002	.415	.029	.054	.014	.386	.032	.053
.2	8	32	.000	.266	.037	.056	.006	.257	.038	.054
.2	8	64	.000	.178	.044	.055	.003	.175	.044	.054
.2	16	1	.149	1.692	.023	.082	.160	1.144	.028	.090
.2	16	2	.110	1.477	.024	.084	.127	1.057	.029	.091
.2	16	4	.064	1.154	.025	.087	.085	.898	.030	.092
.2	16	8	.022	.784	.028	.089	.041	.673	.031	.091
.2	16	16	.004	.482	.031	.088	.016	.446	.034	.087
.2	16	32	.000	.293	.037	.080	.006	.282	.039	.078
.2	16	64	.000	.188	.044	.068	.003	.185	.044	.066
.2	32	1	.161	1.769	.032	.161	.168	1.295	.036	.172
.2	32	2	.131	1.594	.033	.162	.142	1.211	.037	.171
.2	32	4	.087	1.313	.033	.163	.101	1.056	.037	.170
.2	32	8	.041	.938	.034	.161	.056	.812	.037	.164
.2	32	16	.009	.583	.034	.152	.020	.540	.037	.150
.2	32	32	.001	.341	.039	.129	.007	.329	.040	.125
.2	32	64	.000	.206	.043	.096	.003	.202	.044	.094

tities found in the March 1999 to March 2004 *American Economic Review*, the February 1999 to June 2004 *Journal of Political Economy*, and the February 1999 to February 2004 *Quarterly Journal of Economics*. We found that 50% of the articles had at least one overidentifying restriction, 25% had at least three, and 10% had seven or more. As we have seen, the CSE can provide a substantial improvement even with small numbers of overidentifying restrictions, so there appears to be wide scope for applying these results. Table 6 summarizes estimates of μ^2 and ρ from these studies.

It is interesting to note that nearly all of the studies had values of ρ that were quite low, so that the $\rho = .8$ case considered above may not be very relevant for practice. Also, the concentration parameters were mostly in the range where the many instrument asymptotics with CSE should work well.

5. MANY INSTRUMENT ASYMPTOTICS

Theoretical justification of the CSE is provided by asymptotic theory where the number of instruments grows with the sample size and using the CSE in t-ratios leads to a better asymptotic approximation (by the standard normal) than do the usual standard errors. This theory is consistent with the empirical and Monte Carlo results where the CSE improve accuracy of the Gaussian approximation.

Table 4. Weak instrument limit of LIML and Fuller—$\rho = .5$

			LIML				Fuller			
ρ	K	μ^2	Median	IQR	Bekker	Standard	Median	IQR	Bekker	Standard
.5	1	1	.200	1.221	.024	.024	.380	.470	.182	.032
.5	1	2	.091	.952	.031	.031	.268	.494	.132	.039
.5	1	4	.024	.700	.038	.038	.149	.463	.085	.046
.5	1	8	.003	.493	.042	.042	.068	.395	.061	.050
.5	1	16	−.001	.344	.043	.043	.031	.311	.053	.049
.5	1	32	.000	.240	.042	.042	.016	.229	.049	.046
.5	1	64	.000	.170	.044	.044	.008	.166	.047	.046
.5	2	1	.239	1.308	.036	.042	.360	.601	.096	.057
.5	2	2	.123	1.028	.040	.046	.260	.569	.084	.059
.5	2	4	.040	.754	.044	.051	.150	.503	.072	.062
.5	2	8	.005	.516	.045	.051	.068	.413	.061	.059
.5	2	16	.000	.351	.043	.047	.031	.318	.053	.053
.5	2	32	.000	.243	.042	.044	.015	.232	.048	.048
.5	2	64	.000	.171	.045	.046	.008	.167	.048	.048
.5	4	1	.283	1.392	.055	.081	.361	.745	.093	.105
.5	4	2	.167	1.134	.054	.080	.267	.683	.085	.100
.5	4	4	.064	.830	.052	.076	.157	.572	.075	.091
.5	4	8	.012	.564	.050	.068	.073	.453	.065	.078
.5	4	16	.001	.369	.045	.055	.032	.333	.054	.062
.5	4	32	.000	.250	.043	.049	.016	.238	.049	.053
.5	4	64	.000	.173	.044	.048	.008	.168	.047	.049
.5	8	1	.325	1.478	.078	.146	.375	.890	.109	.179
.5	8	2	.218	1.245	.073	.137	.287	.813	.099	.165
.5	8	4	.099	.937	.065	.121	.175	.673	.085	.141
.5	8	8	.024	.632	.056	.097	.081	.510	.070	.110
.5	8	16	.001	.401	.048	.071	.032	.362	.057	.079
.5	8	32	.000	.261	.043	.057	.016	.248	.049	.061
.5	8	64	.000	.176	.045	.053	.008	.172	.048	.055
.5	16	1	.367	1.553	.097	.223	.397	1.035	.120	.259
.5	16	2	.271	1.365	.091	.208	.316	.954	.111	.240
.5	16	4	.147	1.076	.080	.183	.204	.805	.097	.208
.5	16	8	.048	.733	.065	.145	.098	.599	.078	.161
.5	16	16	.005	.457	.052	.101	.035	.411	.061	.111
.5	16	32	.000	.282	.045	.074	.015	.269	.050	.078
.5	16	64	.000	.185	.045	.063	.007	.180	.048	.064
.5	32	1	.400	1.603	.113	.296	.417	1.165	.130	.331
.5	32	2	.316	1.461	.106	.281	.345	1.093	.122	.313
.5	32	4	.204	1.218	.094	.253	.244	.955	.108	.280
.5	32	8	.085	.871	.077	.209	.127	.725	.088	.228
.5	32	16	.016	.546	.060	.152	.045	.490	.067	.162
.5	32	32	.001	.323	.047	.107	.016	.307	.052	.112
.5	32	64	.000	.199	.045	.083	.008	.194	.048	.084

Some regularity conditions are important for the results. Let Z_t', u_t, V_t', and Υ_t' denote the tth row of Z, u, V, and Υ, respectively. Here we will consider the case where Z is constant, leaving the treatment of random Z to future research.

Assumption 1. Z includes among its columns a vector of ones, $\operatorname{rank}(Z) = K$, $\sum_{t=1}^T (1-p_{tt})^2/T \geq C > 0$.

The restriction that $\operatorname{rank}(Z) = K$ is a normalization that requires excluding redundant columns from Z. It can be verified in particular cases. For instance, when w_t is a continuously distributed scalar, $Z_t = p^K(w_t)$, and $p_{kK}(w) = w^{k-1}$, it can be shown that $Z'Z$ is nonsingular with probability 1 for $K < T$. [The observations $w_1, \dots, w_T$ are distinct with probability 1 and therefore, by $K < T$, cannot all be roots of a Kth-degree polynomial. It follows that for any nonzero a there must be some t with $a'Z_t = a'p^K(w_t) \neq 0$, implying $a'Z'Za > 0$.] The condition $\sum_{t=1}^T (1-p_{tt})^2/T \geq C$ implies that $K/T \leq 1 - C$, because $p_{tt} \leq 1$ implies $\sum_{t=1}^T (1-p_{tt})^2/T \leq \sum_{t=1}^T (1-p_{tt})/T = 1 - K/T$.

Assumption 2. There is a $G \times G$ matrix $S_T = \tilde{S}_T \operatorname{diag}(\mu_{1T}, \dots, \mu_{GT})$ and z_t such that $\Upsilon_t = S_T z_t/\sqrt{T}$, $\tilde{S}_T$ is bounded and the smallest eigenvalue of $\tilde{S}_T \tilde{S}_T'$ is bounded away from zero, for each j either $\mu_{jT} = \sqrt{T}$ or $\mu_{jT}/\sqrt{T} \longrightarrow 0$, $\mu_T = \min_{1 \leq j \leq G} \mu_{jT} \longrightarrow \infty$, and $\sqrt{K}/\mu_T^2 \longrightarrow 0$. Also, $\sum_{t=1}^T \|z_t\|^4/$

Table 5. Weak instrument limit of LIML and Fuller—$\rho = .8$

			LIML				Fuller			
ρ	K	μ^2	Median	IQR	Bekker	Standard	Median	IQR	Bekker	Standard
.8	1	1	.290	1.044	.113	.113	.556	.429	.495	.203
.8	1	2	.124	.891	.102	.102	.390	.400	.321	.164
.8	1	4	.027	.699	.087	.087	.220	.370	.175	.126
.8	1	8	.001	.498	.074	.074	.100	.355	.108	.100
.8	1	16	−.001	.346	.062	.062	.048	.297	.082	.079
.8	1	32	.000	.242	.053	.053	.025	.225	.066	.065
.8	1	64	.000	.170	.049	.049	.012	.164	.056	.055
.8	2	1	.347	1.088	.147	.166	.554	.480	.397	.275
.8	2	2	.160	.922	.118	.133	.394	.434	.268	.207
.8	2	4	.041	.723	.093	.102	.225	.389	.162	.146
.8	2	8	.003	.512	.076	.080	.102	.364	.108	.107
.8	2	16	.000	.350	.063	.065	.048	.301	.083	.083
.8	2	32	.000	.243	.054	.055	.025	.226	.066	.066
.8	2	64	.000	.170	.048	.049	.013	.164	.055	.055
.8	4	1	.414	1.163	.183	.238	.562	.574	.316	.351
.8	4	2	.214	.970	.141	.183	.405	.496	.230	.265
.8	4	4	.062	.763	.102	.127	.234	.420	.152	.176
.8	4	8	.007	.533	.079	.091	.106	.378	.108	.119
.8	4	16	.000	.358	.064	.069	.048	.307	.083	.087
.8	4	32	.000	.245	.054	.056	.025	.228	.066	.068
.8	4	64	.000	.171	.048	.050	.012	.165	.055	.056
.8	8	1	.489	1.213	.219	.316	.586	.679	.299	.422
.8	8	2	.286	1.043	.168	.248	.431	.594	.230	.331
.8	8	4	.101	.819	.119	.171	.253	.475	.160	.226
.8	8	8	.013	.572	.085	.111	.111	.404	.111	.142
.8	8	16	.000	.373	.065	.078	.049	.320	.084	.097
.8	8	32	.001	.250	.055	.060	.025	.232	.067	.073
.8	8	64	.000	.173	.049	.052	.012	.167	.056	.059
.8	16	1	.561	1.245	.249	.390	.620	.781	.305	.483
.8	16	2	.373	1.127	.201	.323	.473	.713	.248	.403
.8	16	4	.162	.900	.142	.232	.287	.564	.176	.291
.8	16	8	.029	.640	.094	.146	.125	.451	.118	.181
.8	16	16	.000	.405	.069	.095	.049	.346	.085	.115
.8	16	32	.000	.262	.056	.068	.024	.243	.068	.081
.8	16	64	.000	.177	.049	.056	.013	.171	.056	.063
.8	32	1	.624	1.250	.277	.457	.656	.873	.316	.534
.8	32	2	.469	1.201	.234	.401	.530	.836	.270	.473
.8	32	4	.248	.998	.172	.309	.341	.689	.201	.367
.8	32	8	.062	.731	.111	.203	.152	.523	.132	.240
.8	32	16	.004	.458	.076	.126	.053	.388	.091	.148
.8	32	32	.001	.282	.059	.084	.025	.262	.069	.098
.8	32	64	.000	.185	.049	.065	.012	.178	.056	.072

$T^2 \longrightarrow 0$, and $\sum_{t=1}^{T} z_t z_t'/T$ is bounded and uniformly nonsingular.

This condition allows for both many instruments or many weak instruments. If $\mu_T = \sqrt{T}$, then K may grow as fast as T and still satisfy this condition. This case is many instruments.

Table 6. Five years of AER, JPE, QJE

	Num. papers	Median	Q10	Q25	Q75	Q90
μ^2	28	23.6	8.95	12.7	105	588
ρ	22	.279	.022	.0735	.466	.555

Allowing for K to grow and for μ_T to grow slower than $\sqrt{T}$ is the many weak instrument case. Assumption 2 will imply that, when K grows no faster than μ_T^2, the convergence rate of $\hat{\delta}$ will be no slower than $1/\mu_T$. When K grows faster than μ_T^2, the convergence rate of $\hat{\delta}$ will be no slower than $\sqrt{K}/\mu_T^2$. This condition allows for some components of δ to be weakly identified and other components (like the constant) to be strongly identified.

Assumption 3. $(u_1, V_1), \ldots, (u_T, V_T)$ are independent with $E[u_t] = 0$, $E[V_t] = 0$, $E[u_t^8]$ and $E[\|V_t\|^8]$ are bounded in t, $\text{Var}((u_t, V_t')') = \text{diag}(\Omega^*, 0)$, and Ω^* is nonsingular.

This hypothesis includes moment existence and homoscedasticity assumptions. The consistency of the CSE depends on homoscedasticity, as does consistency of the LIML estimator itself with many instruments; see Bekker and van der Ploeg (2005), Chao and Swanson (2004), and Hausman, Newey, and Woutersen (2005).

Assumption 4. There is π_{KT} such that $\Delta_T^2 = \sum_{t=1}^T \|z_t - \pi_{KT} Z_t\|^2/T \longrightarrow 0$.

This condition allows an unknown reduced form that is approximated by a linear combination of the instrumental variables. An important example is a model with

$$X_t = \begin{pmatrix} \pi_{11} Z_{1t} + \mu_T f_0(w_t)/\sqrt{T} \\ Z_{1t} \end{pmatrix} + \begin{pmatrix} v_t \\ 0 \end{pmatrix},$$

$$Z_t = \begin{pmatrix} Z_{1t} \\ p^K(w_t) \end{pmatrix},$$

where Z_{1t} is a $G_2 \times 1$ vector of included exogenous variables, $f_0(w)$ is a $(G - G_2)$ dimensional vector function of a fixed-dimensional vector of exogenous variables w and $p^K(w) =_{\text{def}} (p_{1K}(w), \ldots, p_{K-G_2,K}(w))'$. The variables in X_t other than Z_{1t} are endogenous with reduced form $\pi_{11} Z_{1t} + \mu_T f_0(w_t)/\sqrt{T}$. The function $f_0(w)$ may be a linear combination of a subvector of $p^K(w)$, in which case $\Delta_T = 0$ in Assumption 4, or it may be an unknown function that can be approximated by a linear combination of $p^K(w)$. For $\mu_T = \sqrt{T}$ this example is like the model in Donald and Newey (2001) where Z_t includes approximating functions for the optimal (asymptotic variance minimizing) instruments Υ_t, but the number of instruments can grow as fast as the sample size. When $\mu_T^2/T \longrightarrow 0$, it is a modified version where the model is more weakly identified.

To see precise conditions under which the assumptions are satisfied, let

$$z_t = \begin{pmatrix} f_0(w_t) \\ Z_{1t} \end{pmatrix},$$

$$S_T = \tilde{S}_T \operatorname{diag}(\mu_T, \ldots, \mu_T, \sqrt{T}, \ldots, \sqrt{T}),$$

$$\tilde{S}_T = \begin{pmatrix} I & \pi_{11} \\ 0 & I \end{pmatrix}.$$

By construction we have $\Upsilon_t = S_T z_t/\sqrt{T}$. Assumption 2 imposes the requirements that

$$\sum_{t=1}^T \|z_t\|^4/T^2 \longrightarrow 0,$$

$$\sum_{t=1}^T z_t z_t'/T \text{ is uniformly nonsingular.}$$

The other requirements of Assumption 2 are satisfied by construction. Turning to Assumption 3, we require that $\operatorname{Var}(u_t, v_t')$ is nonsingular. Because the submatrix of $\tilde{S}_T$ corresponding to $V_{tj} = 0$ is the same as the submatrix corresponding to the included exogenous variables Z_{1t}, we have $\tilde{S}_{T22} = I$ is uniformly nonsingular. For Assumption 4, let $\pi_{KT} = [\tilde{\pi}_{KT}', [I_{G_2}, 0]']'$. Then Assumption 4 will be satisfied if for each T there exists $\tilde{\pi}_{KT}$ with

$$\Delta_T^2 = \sum_{t=1}^T \|z_t - \pi_{KT}' Z_t\|^2/T$$

$$= \sum_{t=1}^T \|f_0(w_t) - \tilde{\pi}_{KT}' Z_t\|^2/T \longrightarrow 0.$$

The following is a consistency result.

Theorem 1. If Assumptions 1–4 are satisfied and $\hat{\alpha} = K/T + o_p(\mu_T^2/T)$ or $\hat{\delta}$ is LIML or FULL, then $\mu_T^{-1} S_T'(\hat{\delta} - \delta_0) \longrightarrow_p 0$ and $\hat{\delta} \longrightarrow_p \delta_0$.

This result is more general than that in Chao and Swanson (2005) in allowing for strongly identified covariates but is similar to that in Chao and Swanson (2003). See Chao and Swanson (2005) for an interpretation of the condition on $\hat{\alpha}$. This result gives convergence rates for linear combinations of $\hat{\delta}$. For instance, in the linear model example setup above, it implies that $\hat{\delta}_1$ is consistent and that $\pi_{11}' \hat{\delta}_1 + \hat{\delta}_2 = o_p(\mu_T/\sqrt{T})$.

Before stating the asymptotic normality results we describe their form. Let $\sigma_u^2 = E[u_t^2]$, $\sigma_{Vu}^2 = E[V_t u_t]$, $\gamma = \sigma_{Vu}/\sigma_u^2$, $\tilde{V} = V - u\gamma'$, having tth row $\tilde{V}_t'$; and let $\tilde{\Omega} = E[\tilde{V}_t \tilde{V}_t']$. There will be two cases depending on the speed of growth of K relative to μ_T^2.

Assumption 5. Either (I) K/μ_T^2 is bounded or (II) $K/\mu_T^2 \longrightarrow \infty$.

To state a limiting distribution result it is helpful to also assume that certain objects converge. When considering the behavior of t-ratios we will drop this condition.

Assumption 6. $H = \lim_{T\longrightarrow\infty} (1 - \tau_T) z'z/T$, $\tau = \lim_{T\longrightarrow\infty} \tau_T$, $\kappa = \lim_{T\longrightarrow\infty} \kappa_T$, $A = E[u_t^2 \tilde{V}_t] \times \lim_{T\longrightarrow\infty} \sum_{t=1}^T z_t'(p_{tt} - \frac{K}{T})/\sqrt{KT}$ exist and in case (I) $\sqrt{K} S_T^{-1} \longrightarrow S_0$ or in case (II) $\mu_T S_T^{-1} \longrightarrow \bar{S}_0$.

Below we will give results for t-ratios that do not require this condition. Let $B = (\kappa - \tau) E[(u_t^2 - \sigma_u^2) \tilde{V}_t \tilde{V}_t']$. Then in case (I) we will have

$$\begin{aligned} S_T'(\hat{\delta} - \delta_0) &\xrightarrow{d} \mathrm{N}(0, \Lambda_I), \\ S_T' \hat{\Lambda} S_T &\xrightarrow{p} \Lambda_I, \\ \Lambda_I &= H^{-1} \Sigma_I H^{-1}, \\ \Sigma_I &= (1 - \tau)\sigma_u^2 \{H + S_0 \tilde{\Omega} S_0'\} \\ &\quad + (1 - \tau)(S_0 A + A' S_0') + S_0 B S_0'. \end{aligned} \tag{5.1}$$

In case (II) we will have

$$\begin{aligned} (\mu_T/\sqrt{K}) S_T'(\hat{\delta} - \delta_0) &\xrightarrow{d} \mathrm{N}(0, \Lambda_{II}), \\ (\mu_T^2/K) S_T' \hat{\Lambda} S_T &\xrightarrow{p} \Lambda_{II}, \\ \Lambda_{II} &= H^{-1} \Sigma_{II} H^{-1}, \\ \Sigma_{II} &= \bar{S}_0 [(1 - \tau)\sigma_u^2 \tilde{\Omega} + B] \bar{S}_0'. \end{aligned} \tag{5.2}$$

The asymptotic variance expressions allow for the many instrument sequence of Kunitomo (1980), Morimune (1983), and

Bekker (1994) and the many weak instrument sequence of Chao and Swanson (2003, 2005). This formula also extends that of Bekker and van der Ploeg (2005) to allow general instruments. When K and μ_T^2 grow as fast as T, the variance formula generalizes that of Anderson et al. (2005), which had previously generalized that of Hansen et al. (2004) to allow for $E[u_t|\tilde{V}_t] \neq 0$ and $E[u_t^2|\tilde{V}_t] \neq \sigma_u^2$, to include the coefficients of included exogenous variables. The formula also generalizes that of Anderson et al. (2005) to allow for μ_T^2 and K to grow slower than T. Then $\tau = \kappa = 0$, $A = 0$, and $B = 0$, giving a formula which generalizes that of Stock and Yogo (2005b) to allow for included exogenous variables and to allow for K to grow faster than μ_T^2, similarly to Chao and Swanson (2004). When K does grow faster than μ_T^2, the asymptotic variance of $\hat{\delta}$ may be singular. This occurs because the many instruments adjustment term is singular with included exogenous variables and it dominates the nonsingular matrix H when K grows that fast.

Theorem 2. If Assumptions 1–6 are satisfied, $\hat{\alpha} = \tilde{\alpha} + O_p(1/T)$ or $\hat{\delta}$ is LIML or FULL, then in case (I) (5.1) is satisfied and in case (II) (5.2) is satisfied. Also, in each case, if Σ is nonsingular, then $\widehat{LM}(\delta_0) \longrightarrow_d \chi^2(G)$.

It is straightforward to show that when the disturbances are Gaussian, the Wald test with the CSE attains the power envelope of Andrews and Stock (2006) under the conditions given here, where $\sqrt{K}/\mu_T^2 \longrightarrow 0$. Andrews and Stock (2006) showed that the LM statistic of Kleibergen (2002) attains this envelope and it is straightforward to show that the Wald statistic is asymptotically equivalent to the LM statistic under local alternatives. For brevity we omit this demonstration.

To give results for t-ratios and to understand better the performance of the CSE we now turn to approximation results. We will give order of approximation results for two t-ratios involving linear combinations of coefficients, one with the CSE and another with the usual formula, and compare results.

We first give stochastic expansions around a normalized sum with remainder rate. To describe these results we need some additional notation. Define

$$\hat{H} = X'PX - \hat{\alpha}X'X,$$
$$W = [(1-\tau_T)\Upsilon + P_Z\tilde{V} - \tau_T\tilde{V}]S_T^{-1\prime},$$
$$H_T = (1-\tau_T)z'z/T,$$
$$A_T = \left(\sum_{t=1}^{T}(p_{tt} - \tau_T)z_t/\sqrt{T}\right)E[u_t^2\tilde{V}_t']S_T^{-1\prime},$$
$$B_T = (\kappa_T - \tau_T)E[(u_t^2 - \sigma_u^2)\tilde{V}_t\tilde{V}_t'],$$
$$\Sigma_T = \sigma_u^2(1-\tau_T)(H_T + KS_T^{-1}\tilde{\Omega}S_T^{-1\prime}) + (1-\tau_T)(A_T + A_T') + KS_T^{-1}B_TS_T^{-1\prime},$$
$$\Lambda_T = H_T^{-1}\Sigma_T H_T^{-1}.$$

We will consider t-ratios for a linear combination $c'\hat{\delta}$ of the IV estimator, where c are the linear combination coefficients, satisfying the following condition:

Assumption 7. There is μ_T^c such that $\mu_T^c c'S_T^{-1\prime}$ is bounded and in case (I) $(\mu_T^c)^2 c'S_T^{-1\prime}\Lambda_T S_T^{-1}c$ and $(\mu_T^c)^2 c'S_T^{-1\prime}H_T^{-1}S_T^{-1}c$ are bounded away from zero and in case (II) $(\mu_T^c)^2 c'S_T^{-1\prime}\Lambda_T \times S_T^{-1}c\mu_T^2/K$ is bounded away from zero.

Let $\tilde{\mu}_T = \mu_T$ in case (I) and $\tilde{\mu}_T = \mu_T^2/\sqrt{K}$ in case (II).

Theorem 3. Suppose that Assumptions 1–5 and 7 are satisfied and $\hat{\alpha} = \tilde{\alpha} + O_p(1/T)$ or $\hat{\delta}$ is LIML or FULL. Then, for $\varepsilon_T = \Delta_T + 1/\tilde{\mu}_T$ in case (I) and case (II),

$$\frac{c'(\hat{\delta} - \delta_0)}{\sqrt{c'\hat{\Lambda}c}} \xrightarrow{d} \mathrm{N}(0,1),$$
$$\frac{c'(\hat{\delta} - \delta_0)}{\sqrt{c'\hat{\Lambda}c}} = \frac{c'S_T^{-1\prime}H_T^{-1}W'u}{\sqrt{c'S_T^{-1\prime}\Lambda_T S_T^{-1}c}} + O_p(\varepsilon_T).$$

Also, in case (II),

$$\Pr\left(\left|c'(\hat{\delta} - \delta_0)/\sqrt{\hat{\sigma}_u^2 c'\hat{H}^{-1}c}\right| \geq C\right) \longrightarrow 1$$

for all C whereas in case (I),

$$\frac{c'(\hat{\delta} - \delta_0)}{\sqrt{\hat{\sigma}_u^2 c'\hat{H}^{-1}c}} = \frac{c'S_T^{-1\prime}H_T^{-1}W'u}{\sqrt{\sigma_u^2 c'S_T^{-1\prime}H_T^{-1}S_T^{-1}c}} + O_p(\varepsilon_T).$$

Here we find that the t-ratio based on the linear combination $c'\hat{\delta}$ is equal to a sum of independent random variables, plus a remainder term that is of order $1/\tilde{\mu}_T + \Delta_T$. It is interesting to note that in case (I) the rate of approximation is $1/\mu_T + \Delta_T$ and $1/\mu_T$ is the rate of approximation that would hold for fixed K. For example, when $\mu_T^2 = T$ and $\Delta_T = 0$, the rate of approximation is the usual parametric rate $1/\sqrt{T}$. Thus, even when K grows as fast as T, the remainder terms in Theorem 3 can have the parametric $1/\sqrt{T}$ rate. This occurs because the specification of W accounts for the presence of many instrumental variables.

The reason that the t-ratio with the usual standard errors is unbounded when $K/\mu_T^2 \longrightarrow \infty$ is that the usual variance formula goes to zero relative to the full variance. When K grows that fast, the term that adjusts for many instruments asymptotically dominates the usual variance formula.

To obtain approximation rates for the distribution of the normalized sums in the conclusion of Theorem 3, we impose the following restriction on the joint distribution of u_t and V_t.

Assumption 8. $E[u_t|\tilde{V}_t] = 0$, $E[u_t^2|\tilde{V}_t] = \sigma_u^2$, $E[|u_t|^4|\tilde{V}_t]$ is bounded, and $\sum_{t=1}^{T}\|z_t\|^3/T^{3/2} = O(1/\mu_T)$.

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, 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 $\tilde{V}_t$. For example, if $V_t = \gamma u_t + \tilde{V}_t$ for any $\tilde{V}_t$ that is statistically independent of u_t, then Assumption 4 would be satisfied. The asymptotic variance of the estimators is simpler under these conditions. This condition implies that $E[u_t^2\tilde{V}_t] = E[E[u_t^2|\tilde{V}_t]\tilde{V}_t] = 0$ and $E[u_t^2\tilde{V}_t\tilde{V}_t'] = E[E[u_t^2|\tilde{V}_t]\tilde{V}_t\tilde{V}_t'] = \sigma_u^2 E[\tilde{V}_t\tilde{V}_t']$, so that $A_T = 0$ and $B_T = 0$.

Theorem 4. If Assumptions 1–5, 7, and 8 are satisfied, then for case (I)

$$\Pr\left(\frac{c'S_T^{-1\prime}H_T^{-1}W'u}{\sqrt{c'S_T^{-1\prime}\Lambda_T S_T^{-1}c}} \le q\right) = \Phi(q) + O(1/\mu_T),$$

$$\Pr\left(\frac{c'S_T^{-1\prime}H_T^{-1}W'u}{\sqrt{\sigma_u^2 c'S_T^{-1\prime}H_T^{-1}S_T^{-1}c}} \le q\right) = \Phi(q) + O(1/\mu_T + K/\mu_T^2).$$

When the variance Λ_T that adjusts for the presence of many instruments appears in the denominator, the approximation is the fixed K rate $1/\mu_T$. In contrast, in case (I) when the usual variance formula $\sigma_u^2 H_T^{-1}$ appears in the denominator, the rate of approximation has an additional K/μ_T^2 term. This term will go to zero more slowly than $1/\mu_T$ when K grows faster than μ_T. When K grows as fast as μ_T^2, the remainder term does not even go to zero, which corresponds to the usual standard errors being inconsistent.

We interpret this result as showing a clear advantage for the CSE with many instrumental variables. The condition for the usual standard errors to have as good an approximation rate as the CSE, that K grows slower than μ_T, may not seem very onerous when $\mu_T = \sqrt{T}$. However, when μ_T grows slower than $\sqrt{T}$ this condition would put severe limits on the number of instrumental variables. Thus, if we think of μ_T growing slowly as representing a weakly identified model, we should expect to find an improvement from using the CSE even with small numbers of instrumental variables. This interpretation is consistent with our empirical and Monte Carlo results.

It would be nice to combine Theorems 3 and 4 to obtain a result on the rate of distributional approximation for the t-ratio. It is well known that this will hold with additional tail conditions on the remainder in the stochastic expansions of Theorem 3; see Rothenberg (1984). To do this is beyond the scope of this article.

We can also show that our modified version of the Kleibergen (2002) statistic is valid under weak instruments.

Theorem 5. If Assumptions 1–3 are satisfied, for each j either $\mu_{jT} = 1$ or $\mu_{jT} = \sqrt{T}$, and $S_T^{-1} \longrightarrow S_0$, $Z'Z/T \longrightarrow M$, nonsingular, and $Z'z/T \longrightarrow R$, then $\widehat{LM}(\delta_0) \longrightarrow_d \chi^2(G)$.

6. CONCLUSION

In this article, we have given standard errors that correct for many instruments when disturbances are not Gaussian. We have also shown that the LIML and Fuller (1977) estimators with Bekker (1994) standard errors provide improved inference relative to the usual asymptotic approximation in instrumental variable settings across a wide range of applications. The Angrist and Krueger (1991) study provided an example where the CSE with 180 instruments is substantially smaller than the CSE with 3 instruments and confidence intervals closely match those of Kleibergen (2002). Through simulations, we confirm that using the CSE leads to more accurate approximations in many cases. We also provide theoretical results that show the validity of the CSE under many instruments and under many weak instruments without imposing normality. The theoretical results also show that the use of the CSE improves the approximation rate relative to when the usual standard errors are used. Overall, the results support the use of the CSE across a wide variety of applications.

ACKNOWLEDGMENTS

The NSF provided financial support for this article under grant 0136869. Amanda Kowalski provided excellent research assistance. Helpful comments were provided by A. Chesher, T. Rothenberg, J. Stock, and participants in seminars at Berkeley, Bristol, Cal Tech, Harvard–MIT, Ohio State, Stanford, the University of Texas–Austin, UCL, and the University of Virginia.

APPENDIX: PROOFS OF THEOREMS

Throughout, let C denote a generic positive constant that may be different in different uses and let M, CS, and T denote the conditional Markov inequality, the Cauchy–Schwarz inequality, and the triangle inequality, respectively.

Lemma A0. If Assumption 2 is satisfied and $\|S_T'(\hat\delta - \delta_0)/\mu_T\|^2/(1+\|\hat\delta\|^2) \longrightarrow_p 0$ then $\|S_T'(\hat\delta - \delta_0)/\mu_T\| \longrightarrow_p 0$.

Proof. When $\|\hat\delta\| \ge a \overset{\text{def}}{=} 2\|\delta_0\| + (1 + 2\|\delta_0\|^2)^{1/2}$, by subtracting $2\|\delta_0\|$ and squaring we have

$$\begin{aligned}(\|\hat\delta\| - 2\|\delta_0\|)^2 &= \|\hat\delta\|^2 - 4\|\hat\delta\|\|\delta_0\| + 4\|\delta_0\|^2 \\ &\ge 1 + 2\|\delta_0\|^2.\end{aligned}$$

Subtracting $2\|\delta_0\|^2$, adding $\|\hat\delta\|^2$, and dividing by 2 gives

$$(\|\hat\delta\| - \|\delta_0\|)^2 \ge (1 + \|\hat\delta\|^2)/2.$$

By Assumption 2, $\lambda_{\min}(S_T S_T'/\mu_T^2) \ge C$, so when $\|\hat\delta\| \ge a$,

$$\frac{\|S_T'(\hat\delta - \delta_0)/\mu_T\|^2}{1 + \|\hat\delta\|^2} \ge C\frac{\|\hat\delta - \delta_0\|^2}{1 + \|\hat\delta\|^2} \ge C/2.$$

It follows that $\|\hat\delta\| < a$ w.p.a.1, and hence $1 + \|\hat\delta\|^2 < 1 + a^2$ and

$$\|S_T'(\hat\delta - \delta_0)/\mu_T\|^2 \le (1 + a^2)\frac{\|S_T'(\hat\delta - \delta_0)/\mu_T\|^2}{1 + \|\hat\delta\|^2} \xrightarrow{p} 0.$$

Lemma A1. If conditional on the $T \times K$ matrix Z, the observations (u_t, v_t) $(t = 1, \dots, T)$ are independent with $E[u_t|Z] = E[v_t|Z] = 0$, and there is C with $E[u_t^4] \le C$, $E[v_t^4] \le C$ for all t then 1, then for $P = Z(Z'Z)^- Z'$,

$$\operatorname{Var}(u'Pv|Z) \le CK, \qquad u'Pv - E[u'Pv|Z] = O_p(\sqrt{K}).$$

Proof. For notational simplicity we suppress the conditioning on Z. Let $\sigma_{ut}^2 = E[u_t^2]$, $\sigma_{uvt} = E[u_t v_t]$, $\sigma_{vt}^2 = E[v_t^2]$. By conditional independence of observations, $E[uv'|Z] = \operatorname{diag}(\sigma_{uv1}, \dots, \sigma_{uvT}) \overset{\text{def}}{=} \Gamma$. Then $E[u'Pv|Z] = \operatorname{tr}(PE[vu']) = \operatorname{tr}(P\Gamma) = \sum_{t=1}^T p_{tt}\sigma_{uvt}$, where $p_{st} = P_{st}$. Then we have

$$\begin{aligned}&E[(u'Pv)^2|Z] \\ &\quad= \sum_{r,s,t,w=1}^{T} p_{rs}p_{tw}E[u_r v_s u_t v_w] \\ &\quad= \sum_t p_{tt}^2 E[u_t^2 v_t^2] + \sum_{s\ne t}\{(p_{ss}p_{tt} + p_{st}^2)\sigma_{uvs}\sigma_{uvt} + p_{st}^2\sigma_{us}^2\sigma_{vt}^2\} \\ &\quad= \sum_t p_{tt}^2\{E[u_t^2 v_t^2] - 2\sigma_{uvt}^2 - \sigma_{us}^2\sigma_{vt}^2\} + \sum_{s,t} p_{ss}p_{tt}\sigma_{uvs}\sigma_{uvt}\end{aligned}$$

$$+\sum_{s,t} p_{st}^2(\sigma_{uvi}\sigma_{uvj}+\sigma_{us}^2\sigma_{vt}^2)$$
$$\le C\sum_t p_{tt}^2 + C\sum_{s,t} p_{st}^2 + \operatorname{tr}(P\Gamma)^2.$$

We have $\sum_s p_{st}^2 = p_{tt}$, so that by $0 \le p_{tt} \le 1$,

$$E[(u'Pv - E[u'Pv])^2] = E[(u'Pv)^2] - \operatorname{tr}(P\Gamma)^2 \le C\sum_t (p_{tt}^2 + p_{tt}) \le CK.$$

The second conclusion follows by M.

Lemma A2. If (i) P is a constant idempotent matrix with $\operatorname{rank}(P) = K$; (ii) $(W_{1T}, V_1, u_1), \dots, (W_{1T}, V_T, u_T)$ are independent and $D_T = \sum_{t=1}^T E[W_{tT}W_{tT}']$ is bounded; (iii) (V_t', u_t) has bounded fourth moments, $E[V_t] = 0$, $E[u_t] = 0$, and $E[(V_t', u_t)'(V_t', u_t)]$ is constant; (iv) $\sum_{t=1}^T E[\|W_{tT}\|^4] \longrightarrow 0$; (v) $K \to \infty$; then for $\bar\Sigma =_{\text{def}} E[V_tV_t']E[u_t^2] + E[V_tu_t]E[u_tV_t']$, $\kappa_T = \sum_{t=1}^T p_{tt}^2/K$, and any sequence of bounded vectors c_{1T}, c_{2T} such that $V_T = c_{1T}'D_Tc_{1T} + (1-\kappa_T)c_{2T}'\bar\Sigma c_{2T}$ is bounded away from zero it follows that

$$Y_T = V_T^{-1/2}\left(\sum_{t=1}^T c_{1T}'W_{tT} + c_{2T}'\sum_{s\ne t} V_s p_{st} u_t/\sqrt{K}\right) \xrightarrow{d} \mathrm{N}(0,1).$$

Proof. Without changing notation let $c_{1T} = c_{1T}/V_T^{-1/2}$ and $c_{2T} = c_{2T}/V_T^{-1/2}$, and note that these are bounded in T by V_T bounded away from zero. Let $w_{tT} = c_{1T}'W_{tT}$ and $v_t = c_{2T}'V_t$, where we suppress the T subscript on v_t for convenience. Then we have

$$Y_T = w_{1T} + \sum_{t=2}^T y_{tT},$$
$$y_{tT} = w_{tT} + \sum_{s<t}(v_s p_{st} u_t + v_t p_{st} u_t)/\sqrt{K}.$$

Also, by $E[\|W_{1T}\|^4] \le \sum_{t=1}^T E[\|W_{tT}\|^4] \longrightarrow 0$, so that $E[w_{1T}^2] \longrightarrow 0$ and hence

$$Y_T = \sum_{t=2}^T y_{tT} + o_p(1).$$

Note that y_{tT} is martingale difference, so that we can apply a martingale central limit theorem. It follows by P idempotent that $\sum_{s=1}^T p_{st}^2 = p_{tt}$ and $\sum_{t=1}^T p_{tt} = K$. Then, for $D_T = \sum_{t=1}^T E[W_{tT}W_{tT}']$,

$$s_T^2 = E\left[\left(\sum_{t=2}^T y_{tT}\right)^2\right]$$
$$= \sum_{t=2}^T E[w_{tT}^2] + E\left[\left(\sum_{s\ne t} v_s p_{st} u_t\right)^2\right]/K$$
$$= c_{1T}'D_Tc_{1T} - E[w_{1T}^2] + \sum_{s\ne t}\sum_{q\ne r} p_{st}p_{qr}E[v_su_tv_qu_r]/K$$
$$= c_{1T}'D_Tc_{1T} + \left\{E[v_t^2]E[u_t^2] + (E[v_tu_t])^2\right\}(1-\kappa_T) + o(1)$$
$$= c_{1T}'D_Tc_{1T} + c_{2T}'(1-\kappa_T)\bar\Sigma c_{2T} + o(1) \longrightarrow 1.$$

Note that s_T^2 is bounded and bounded away from zero. Also

$$\sum_{t=2}^T E[y_{tT}^4] \le C\sum_{t=2}^T E[\|W_{tT}\|^4] + C\sum_{t=2}^T E\left[\left(\sum_{j<t}\{v_tp_{tj}u_j + v_jp_{tj}u_t\}\right)^4\right]/K^2.$$

By condition (iv), $\sum_{t=2}^T E[\|W_{tT}\|^4] \longrightarrow 0$. Also, by $|p_{st}| \le 1$ and $\sum_{j=1}^T p_{tj}^2 = p_{tt}$,

$$\sum_{t=2}^T E\left[\left(\sum_{j<t} v_tp_{tj}u_j\right)^4\right]/K^2$$
$$= \frac{1}{K^2}\sum_{t=2}^T \sum_{j,k,\ell,m<t} p_{tj}p_{tk}p_{t\ell}p_{tm}E[v_t^4u_ju_ku_\ell u_m]$$
$$= \frac{1}{K^2}\sum_{t=2}^T \sum_{j,k,\ell,m<t} E[v_t^4]p_{tj}p_{tk}p_{t\ell}p_{tm}E[u_ju_ku_\ell u_m]$$
$$\le \frac{C}{K^2}\sum_{t=2}^T\left(\sum_{j<t} p_{tj}^4 + \sum_{j,k<t} p_{tj}^2p_{tk}^2\right)$$
$$\le \frac{C}{K^2}\left(\sum_{t=1}^T\sum_{j=1}^T p_{tj}^2 + \sum_{t=1}^T\left(\sum_{j=1}^T p_{tj}^2\right)\left(\sum_{k=1}^T p_{tk}^2\right)\right)$$
$$= \frac{C}{K^2}\left(\sum_{t=1}^T p_{tt} + \sum_{t=1}^T p_{tt}^2\right) \le \frac{C}{K} \longrightarrow 0.$$

Therefore, $\sum_{t=2}^T E[y_{tT}^4] \longrightarrow 0$, so the Lindbergh condition is satisfied. To apply the martingale central limit theorem it now suffices to show that for $Z_t = (W_{tT}, V_t, u_t)$,

$$\sum_{t=2}^T E[y_{tT}^2|Z_1,\dots,Z_{t-1}] - s_T^2 \xrightarrow{p} 0. \tag{A.1}$$

Note first that by independence of $W_{1T}, \dots, W_{TT}$,

$$\sum_{t=2}^T (E[w_{tT}^2|Z_1,\dots,Z_{t-1}] - E[w_{tT}^2]) = 0.$$

Also

$$E\left[w_{tT}\sum_{j<t}(v_tp_{tj}u_j + v_jp_{tj}u_t)\right] = 0$$

and

$$E\left[w_{tT}\sum_{j<t}(v_tp_{tj}u_j + v_jp_{tj}u_t)/\sqrt{K}\,\middle|\,Z_1,\dots,Z_{t-1}\right]$$
$$= E[w_{tT}v_t]\sum_{j<t} p_{tj}u_j/\sqrt{K} + E[w_{tT}u_t]\sum_{j<t} p_{tj}v_j/\sqrt{K}.$$

Let $\delta_t = E[w_{tT}v_t]$ and consider the first term $\delta_t\sum_{j<t} p_{tj}u_j/\sqrt{K}$. Let $\bar P$ be the upper triangular matrix with $\bar P_{tj} = P_{tj}$ for $j > t$ and $\bar P_{tj} = 0$, $j \le t$, and let $\delta = (\delta_1, \dots, \delta_T)$. Then $\sum_{t=2}^T\sum_{j<t}\delta_tp_{tj} \times$

$u_j/\sqrt{K} = \delta'\bar{P}'u/\sqrt{K}$. By CS, $\delta'\delta = \sum_{t=1}^{T}(E[w_{tT}v_t])^2 \leq \sum_{t=1}^{T} E[w_{tT}^2]E[v_t^2] \leq C$. by lemma A3 of Chao and Swanson (2004), $\|\bar{P}'\bar{P}\| \leq \sqrt{K}$. It then follows that

$$E[(\delta'\bar{P}'u/\sqrt{K})^2] \leq C\delta'\bar{P}'\bar{P}\delta/K \leq \|\delta\|^2\|\bar{P}'\bar{P}\|/K \leq C\sqrt{K}/K \longrightarrow 0,$$

so that $\delta'\bar{P}'u/\sqrt{K} \longrightarrow_p 0$ by M. Similarly, we have $\sum_{t=2}^{T} E[w_{tT}u_t]\sum_{j<t} p_{tj}v_j/\sqrt{K} \longrightarrow 0$. Therefore it follows by T that

$$\sum_{t=2}^{T} E\left[w_{tT}\sum_{j<t}(v_t p_{tj}u_j + v_j p_{tj}u_t)/\sqrt{K}\Big|Z_t,\ldots,Z_{t-1}\right] \xrightarrow{p} 0.$$

To finish showing that (A.1) is satisfied it only remains to show that for $\bar{y}_{tT} = \sum_{j<t}(v_t p_{tj}u_j + v_j p_{tj}u_t)/\sqrt{K}$,

$$\sum_{t=2}^{T} E[\bar{y}_{tT}^2|Z_1,\ldots,Z_{t-1}] - E[\bar{y}_{tT}^2] \xrightarrow{p} 0. \qquad \text{(A.2)}$$

Note that for $\sigma_u^2 = E[u_t^2]$, $\sigma_v^2 = E[v_t^2]$, $\sigma_{uv} = E[u_t v_t]$,

$$\begin{aligned}
&E[\bar{y}_{tT}^2|Z_1,\ldots,Z_{t-1}] - E[\bar{y}_{tT}^2] \\
&\quad = \sigma_v^2\sum_{j<t} p_{tj}^2(u_j^2 - \sigma_u^2)/K + 2\sigma_v^2\sum_{j<k<t} p_{tj}p_{tk}u_j u_k/K \\
&\qquad + \sigma_u^2\sum_{j<t} p_{tj}^2(v_j^2 - \sigma_v^2)/K + 2\sigma_u^2\sum_{j<k<t} p_{tj}p_{tk}v_j v_k/K \\
&\qquad + 2\sigma_{uv}\sum_{j<t} p_{tj}^2(u_j v_j - \sigma_{uv})/K + 4\sigma_{uv}\sum_{j<k<t} p_{tj}p_{tk}u_j v_k/K.
\end{aligned}$$

Consider the last two terms. Note that

$$\begin{aligned}
&E\left[\left(\sum_{t=2}^{T}\sum_{j<t} p_{tj}(u_j v_j - \sigma_{uv})\right)^2\right]\Big/K^2 \\
&\quad = \sum_{j<t}\sum_{k<s} p_{tj}^2 p_{sk}^2 E[(u_j v_j - \sigma_{uv})(u_k v_k - \sigma_{uv})]/K^2 \\
&\quad = \sum_{j<t,s} p_{tj}^2 p_{j\ell}^2 E[(u_j v_j - \sigma_{uv})^2]/K^2 \\
&\quad \leq C\sum_{j<t,s} p_{tj}^2 p_{sj}^2/K \leq \frac{C}{K^2}\sum_{t,s,j} p_{tj}^2 p_{sj}^2 \\
&\quad = C\sum_j\left(\sum_t p_{jt}^2\right)\left(\sum_s p_{js}^2\right)\Big/K^2 \\
&\quad = C\sum_j p_{jj}^2/K^2 \leq CK/K^2 \longrightarrow 0.
\end{aligned}$$

Also, we have

$$\begin{aligned}
&E\left[\left(\sum_{t=2}^{T}\sum_{j<k<t} p_t p_{tk}u_j v_k\right)^2\right]\Big/K^2 \\
&\quad = \sum_{t,\ell}\sum_{j<k<t}\sum_{m<q<\ell} p_{tj}p_{tk}p_{\ell m}p_{\ell q}E[u_j v_k u_m v_q]/K^2 \\
&\quad = \sum_{t,\ell}\sum_{j<k<t}\sum_{j<k<\ell} p_{tj}p_{tk}p_{\ell j}p_{\ell k}\sigma_u^2\sigma_v^2/K^2 \\
&\quad = C\sum_{j<k<t,\ell} p_{tj}p_{tk}p_{\ell j}p_{\ell k}/K^2 \\
&\quad = C\sum_{j<k<t} p_{tj}^2 p_{tk}^2/K^2 + C\sum_{j<k<t<\ell} p_{tj}p_{tk}p_{\ell j}p_{\ell k}/K^2.
\end{aligned}$$

Note that

$$\sum_{j<k<t} p_{tj}^2 p_{tk}^2/K^2 \leq \sum_t\left(\sum_j p_{tj}^2\right)\left(\sum_k p_{tk}^2\right)\Big/K^2 \leq \sum_t p_{tt}^2/K^2 \longrightarrow 0.$$

Also by lemma A2 of Chao and Swanson (2004),

$$\begin{aligned}
\sum_{j<k<t<\ell} p_{tj}p_{tk}p_{\ell j}p_{\ell k}/K^2 &= \sum_{t<j<k<\ell} p_{kt}p_{kj}p_{\ell t}p_{\ell j}/K^2 \\
&= \sum_{i<j<k<\ell} p_{ik}p_{i\ell}p_{jk}p_{j\ell}/K^2 \\
&= O(K)/K^2 \longrightarrow 0.
\end{aligned}$$

It follows similarly that $E[(\sum_t\sum_{j<k<t} p_{tj}p_{tk}u_k v_j)^2]/K^2 \longrightarrow 0$. Similar arguments can also be applied to show that each of the other four terms following the equality in (A.3) converges in probability to zero. It then follows by T and M that (A.3) is satisfied. By T it then follows that (A.1) is satisfied. Thus all the conditions of the Martingale central limit theorem are satisfied, so that $\sum_{t=2}^{T} y_{tT} \longrightarrow_d \mathrm{N}(0,1)$. Then by the Slutzky theorem the conclusion holds.

For the next result let $\bar{S}_T = \mathrm{diag}(\mu_T, S_T)$, $\tilde{X} = [u, X]\bar{S}_T^{-1\prime}$, and $H_T = (1-\tau_T)\sum_{t=1}^{T} z_t z_t'/T$.

Lemma A2a. If Assumptions 1–4 are satisfied and $\sqrt{K}/\mu_T^2 \longrightarrow 0$, then

$$\tilde{X}'P\tilde{X} - (K/T)(\tilde{X}'\tilde{X}) = \mathrm{diag}(0, H_T) + o_p(1).$$

Proof. Note that $\bar{S}_T^{-1}\bar{\Upsilon} = (0, z_t')'/\sqrt{T}$ and $\bar{S}_T^{-1\prime}\bar{S}_T^{-1} \leq CI/\mu_T^2$, so that

$$\begin{aligned}
E[\|\bar{S}_T^{-1}\bar{\Upsilon}'\bar{V}\bar{S}_T^{-1\prime}\|^2] &= \mathrm{tr}(E[z'\bar{V}S_T^{-1\prime}S_T^{-1}\bar{V}z])/T \\
&\leq C\,\mathrm{tr}(z'z)/T\mu_T^2 \longrightarrow 0.
\end{aligned}$$

Then $\bar{S}_T^{-1}\bar{\Upsilon}'\bar{V}\bar{S}_T^{-1\prime} \longrightarrow_p 0$ by M. Similarly, we have $\bar{S}_T^{-1}\bar{\Upsilon}'P \times \bar{V}\bar{S}_T^{-1\prime} \longrightarrow_p 0$. Also,

$$\bar{S}_T^{-1}\bar{\Upsilon}'(I-P)\bar{\Upsilon}S_T^{-1\prime} = \mathrm{diag}(0, z'(I-P)z/T) \longrightarrow 0.$$

We also have, by $\bar{S}_T^{-1} = O(1/\mu_T)$ and Lemma A1,

$$\begin{aligned}
&\bar{S}_T^{-1}\left(\bar{V}'P\bar{V} - \frac{K}{T}\bar{V}'\bar{V}\right)\bar{S}_T^{-1\prime} \\
&\quad = \bar{S}_T^{-1}\big(K\bar{\Omega} + O_p(\sqrt{K}) - K\bar{\Omega} + O_p(K/\sqrt{T})\big)\bar{S}_T^{-1\prime} \\
&\quad = O_p(\sqrt{K}/\mu_T^2) + O_p(K/\mu_T^2\sqrt{T}) \xrightarrow{p} 0.
\end{aligned}$$

Let $\hat{A} = \tilde{X}'P\tilde{X} - (K/T)\tilde{X}'\tilde{X}$. Then by T, w.p.a.1,

$$\begin{aligned}\hat{A} &= (1-\tau_T)\operatorname{diag}(0, z'z/T)\\ &\quad - \bar{S}_T^{-1}\Big\{\bar{\Upsilon}'(I-P)\bar{\Upsilon} + \bar{\Upsilon}'P\bar{V} + \bar{V}'P\bar{\Upsilon}\\ &\quad -\frac{K}{T}\bar{V}'\bar{\Upsilon} - \frac{K}{T}\bar{\Upsilon}'\bar{V} + \bar{V}'P\bar{V} - \frac{K}{T}\bar{V}'\bar{V}\Big\}\bar{S}_T^{-1\prime}\\ &= \operatorname{diag}(0, H_T) + o_p(1).\end{aligned}$$

Let $z = [z_1, \dots, z_T]'$, so that $\Upsilon = zS_T'/\sqrt{T}$.

Lemma A3. If Assumptions 1–4 are satisfied, then $S_T' \times (\hat{\delta}_{LIML} - \delta_0)/\mu_T \longrightarrow_p 0$.

Proof. Let $\bar{\Upsilon} = [0, \Upsilon]$, $\bar{V} = [u, V]$, $\bar{X} = [y, X]$, so that $\bar{X} = (\bar{\Upsilon} + \bar{V})D$ for

$$D = \begin{bmatrix} 1 & 0 \\ \delta_0 & I \end{bmatrix}.$$

Let $\hat{B} = \bar{X}'\bar{X}/T$. Note that $\|S_T/\sqrt{T}\| \le C$. Then by $\operatorname{tr}(\bar{\Upsilon}'\bar{\Upsilon}) = \operatorname{tr}(S_T z'zS_T')/T$ and $E[\bar{V}\bar{V}'] \le CI_T$,

$$\begin{aligned}E[\|\bar{\Upsilon}'\bar{V}\|^2/T^2] &= \operatorname{tr}(\bar{\Upsilon}'E[\bar{V}'\bar{V}]\bar{\Upsilon})/T^2\\ &\le C\operatorname{tr}(S_T z'zS_T')/T^3 \longrightarrow 0,\end{aligned}$$

so that $\bar{\Upsilon}'\bar{V}/T \longrightarrow_p 0$ by M. Let $\bar{\Omega} = E[\bar{V}_i\bar{V}_i'] = \operatorname{diag}(\Omega^*, 0) \ge C\operatorname{diag}(I_{G-G_2+1}, 0)$ by Assumption 3. By M we have $\bar{V}'\bar{V}/T - \bar{\Omega} \longrightarrow_p 0$, so it follows that w.p.a.1,

$$\begin{aligned}\hat{B} &= (\bar{V}'\bar{V} + \bar{\Upsilon}'\bar{V} + \bar{V}'\bar{\Upsilon} + \bar{\Upsilon}'\bar{\Upsilon})/T\\ &= \bar{\Omega} + \bar{\Upsilon}'\bar{\Upsilon}/T + o_p(1) \ge C\operatorname{diag}(I_{G-G_2+1}, 0).\end{aligned}$$

Note that $\bar{\Upsilon}'\bar{\Upsilon}/T$ is bounded, so that $\hat{B}$ minus a constant, bounded matrix converges to zero; w.p.a.1 it follows that

$$\begin{aligned}C &\le (1, -\delta')\hat{B}(1, -\delta')' = (y - X\delta)'(y - X\delta)/T\\ &\le C\|(1, -\delta')\|^2 = C(1 + \|\delta\|^2).\end{aligned}$$

Next, let $\bar{S}_T = \operatorname{diag}(\mu_T, S_T)$, so that $\bar{S}_T^{-1}\bar{\Upsilon} = (0, z_t')'/\sqrt{T}$. Note that $\bar{S}_T^{-1\prime}\bar{S}_T^{-1} \le CI/\mu_T^2$, so that

$$\begin{aligned}E[\|\bar{S}_T^{-1}\bar{\Upsilon}'\bar{V}\bar{S}_T^{-1\prime}\|^2] &\le C\operatorname{tr}(\bar{S}_T^{-1}\bar{\Upsilon}'\bar{\Upsilon}\bar{S}_T^{-1\prime})/\mu_T^2\\ &= C\operatorname{tr}(z'z)/T\mu_T^2 \longrightarrow 0.\end{aligned}$$

Then $\bar{S}_T^{-1}\bar{\Upsilon}'\bar{V}\bar{S}_T^{-1\prime} \longrightarrow_p 0$. Similarly, we have $\bar{S}_T^{-1}\bar{\Upsilon}'P\bar{V} \times \bar{S}_T^{-1\prime} \longrightarrow_p 0$. Also,

$$\bar{S}_T^{-1}\bar{\Upsilon}'(I-P)\bar{\Upsilon}\bar{S}_T^{-1\prime} = \operatorname{diag}(0, z'(I-P)z/T) \longrightarrow 0.$$

We also have, by $\bar{S}_T^{-1} = O(1/\mu_T)$,

$$\begin{aligned}&\bar{S}_T^{-1}\left(\bar{V}'P\bar{V} - \frac{K}{T}\bar{V}'\bar{V}\right)\bar{S}_T^{-1\prime}\\ &= \bar{S}_T^{-1}\big(K\bar{\Omega} + O_p(\sqrt{K}) - K\bar{\Omega} + O_p(K/\sqrt{T})\big)\bar{S}_T^{-1\prime}\\ &= O_p\big(\sqrt{K}/\mu_T^2\big) + O_p(K/\mu_T^2\sqrt{T}) \xrightarrow{p} 0. \qquad \text{(A.3)}\end{aligned}$$

Let $\tilde{X}_t = \bar{S}_T^{-1}(\bar{\Upsilon}_t + \bar{V}_t)$, $\tilde{X} = [\tilde{X}_1, \dots, \tilde{X}_T]'$, and $\hat{A} = \tilde{X}'P\tilde{X} - (K/T)\tilde{X}'\tilde{X}$. By Assumption 2, $\bar{S}_T^{-1}\bar{\Upsilon}'\bar{\Upsilon}\bar{S}_T^{-1\prime} = \operatorname{diag}(0, z'z/T) \ge \operatorname{diag}(0, I_G)$. Then by T, w.p.a.1,

$$\begin{aligned}\hat{A} &= \left(1 - \frac{K}{T}\right)\operatorname{diag}(0, z'z/T)\\ &\quad - \bar{S}_T^{-1}\Big\{\bar{\Upsilon}'(I-P)\bar{\Upsilon} + \bar{\Upsilon}'P\bar{V} + \bar{V}'P\bar{\Upsilon}\\ &\quad - \frac{K}{T}\bar{V}'\bar{\Upsilon} - \frac{K}{T}\bar{\Upsilon}'\bar{V} + \bar{V}'P\bar{V} - \frac{K}{T}\bar{V}'\bar{V}\Big\}\bar{S}_T^{-1\prime}\\ &\ge C\operatorname{diag}(0, I_G).\end{aligned}$$

Note that $\bar{S}_T'D(1, -\delta')' = (\mu_T, (\delta_0 - \delta)'S_T)'$. It follows that w.p.a.1, by $\bar{X}_i = D'\bar{S}_T\tilde{X}_i$, for all δ

$$\begin{aligned}&\mu_T^{-2}(y - X\delta)'\left(P - \frac{K}{T}I\right)(y - X'\delta)\\ &= \mu_T^{-2}(1, -\delta')[\bar{X}'P\bar{X} - (K/T)\bar{X}'\bar{X}](1, -\delta')'\\ &= \mu_T^{-2}(1, -\delta')D'\bar{S}_T\hat{A}\bar{S}_T'D(1, -\delta')'\\ &\ge C\|S_T'(\delta - \delta_0)/\mu_T\|^2.\end{aligned}$$

Let $\hat{Q}(\delta) = \mu_T^{-2}(y - X\delta)'(P - \frac{K}{T}I)(y - X'\delta)/T^{-1}(y - X\delta)'(y - X'\delta)$. Note that $\hat{\delta} = \arg\min_\delta \hat{Q}(\delta)$. Also, it follows by (A.3) that $\mu_T^{-2}u'(P - \frac{K}{T}I)u \longrightarrow_p 0$, so that by $u'u/T \ge C$ w.p.a.1, $\hat{Q}(\delta_0) \longrightarrow_p 0$. Therefore,

$$0 \le \hat{Q}(\hat{\delta}) \le \hat{Q}(\delta_0) \xrightarrow{p} 0,$$

and hence $\hat{Q}(\hat{\delta}) \longrightarrow_p 0$. By $(y - X\delta)'(y - X\delta)/T \le C(1 + \|\delta\|^2)$, it follows that

$$\frac{\|S_T'(\hat{\delta} - \delta_0)/\mu_T\|^2}{1 + \|\hat{\delta}\|^2} \le C\hat{Q}(\hat{\delta}) \xrightarrow{p} 0.$$

Then by Lemma A0 we have $S_T'(\hat{\delta} - \delta_0)/\mu_T \longrightarrow_p 0$.

Let $\breve{\alpha} = u'Pu/u'u$.

Lemma A4. If Assumptions 1–4 are satisfied, then $\breve{\alpha} = K/T + O_p(\sqrt{K}/T)$.

Proof. By Lemma A1, $u'Pu/K = \sigma_u^2 + O_p(1/\sqrt{K})$. Also $\tilde{\sigma}_u^2 = u'u/T = \sigma_u^2 + O_p(1/\sqrt{T})$ by M. Then

$$\begin{aligned}u'Pu/u'u - K/T &= \frac{K}{T}\left(\frac{u'Pu/K}{\tilde{\sigma}_u^2} - 1\right)\\ &= \frac{K}{T\tilde{\sigma}_u^2}\left(\frac{u'Pu}{K} - \sigma_u^2 - (\tilde{\sigma}_u^2 - \sigma_u^2)\right)\\ &= O_p\left(\frac{K}{T}\right)\left[O_p\left(\frac{1}{\sqrt{K}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right)\right]\\ &= O_p\left(\frac{\sqrt{K}}{T}\right).\end{aligned}$$

Lemma A5. If Assumptions 1–4 are satisfied, $\hat{\alpha} = \breve{\alpha} + O_p(\varepsilon_T^\alpha)$, and $S_T'(\hat{\delta} - \delta_0)/\mu_T = O_p(\varepsilon_T^\delta)$ for $\varepsilon_T^\alpha T/\mu_T^2 \longrightarrow 0$, $\varepsilon_T^\delta \longrightarrow 0$, then

$$\begin{aligned}S_T^{-1}(X'PX - \hat{\alpha}X'X)S_T^{-1\prime} &= H_T + O_p(\Delta_T^2 + \tilde{\mu}_T^{-1} + \varepsilon_T^\alpha T/\mu_T^2),\\ S_T^{-1}(X'P\hat{u} - \hat{\alpha}X'\hat{u})/\mu_T &= O_p(\tilde{\mu}_T^{-1} + \varepsilon_T^\delta + \varepsilon_T^\alpha T/\mu_T^2).\end{aligned}$$

Proof. Note that in case (I), $\sqrt{K}/\mu_T^2 \le C/\tilde{\mu}_T$ and in case (II), $\sqrt{K}/\mu_T^2 = 1/\tilde{\mu}_T$, so that $\sqrt{K}/\mu_T^2 = O(1/\tilde{\mu}_T)$. Also by M, $X'X = O_p(T)$, $X'\hat{u} = O_p(T)$. Therefore,

$$(\hat{\alpha} - \breve{\alpha})S_T^{-1}X'XS_T^{-1\prime} = O_p(\varepsilon_T^{\alpha}T/\mu_T^2),$$
$$(\hat{\alpha} - \breve{\alpha})S_T^{-1}X'\hat{u}/\mu_T = O_p(\varepsilon_T^{\alpha}T/\mu_T^2).$$

Also, by Lemma A4,

$$(\breve{\alpha} - K/T)S_T^{-1}X'XS_T^{-1\prime} = O_p(\sqrt{K}/\mu_T^2) = O_p(\tilde{\mu}^{-1}),$$
$$(\breve{\alpha} - K/T)S_T^{-1}X'\hat{u}/\mu_T = O_p(\tilde{\mu}^{-1}).$$

Also, for $A_T = \Upsilon'(P - I)\Upsilon$, $B_T = \Upsilon'PV - (K/T)\Upsilon'V$, and $D_T = V'PV - (K/T)V'V$ we have

$$S_T^{-1}[X'PX - (K/T)X'X]S_T^{-1\prime} = H_T + S_T^{-1}(A_T + B_T + B_T' + D_T)S_T^{-1\prime}.$$

Note that $-A_T$ is p.s.d. and by Assumption 4

$$-S_T^{-1}A_TS_T^{-1\prime} = z'(I - P)z/T \le (z - Z\pi_{KT}')'(z - Z\pi_{KT}')/T = O(\Delta_T^2).$$

Also, $S_T^{-1\prime}S_T^{-1} \le I/\mu_T^2$ and $E[VV'] \le CI$, so that

$$E[\|S_T^{-1}\Upsilon'PVS_T^{-1\prime}\|^2] \le C\,\mathrm{tr}(z'PPz/T)/\mu_T^2 \le \mathrm{tr}(z'z/T)/\mu_T^2 = O(1/\mu_T^2),$$

and $S_T^{-1}\Upsilon'PVS_T^{-1\prime} = O_p(1/\mu_T)$ by CM. Similarly, $S_T^{-1}\Upsilon' \times VS_T^{-1\prime} = O_p(1/\mu_T)$, so that $S_T^{-1}B_TS_T^{-1\prime} = O_p(1/\mu_T)$ by T. Also, $V'V = T\Omega + O_p(\sqrt{T})$ by M and $V'PV = K\Omega + O_p(\sqrt{K})$ by Lemma A1, so that

$$\begin{aligned} S_T^{-1}D_TS_T^{-1\prime} &= S_T^{-1}(K\Omega - (K/T)T\Omega)S_T^{-1\prime} + O_p(\sqrt{K}/\mu_T^2 + K/\mu_T^2\sqrt{T}) \\ &= O_p(1/\tilde{\mu}_T). \end{aligned}$$

The first conclusion then follows by T.

To show the second conclusion, it follows similarly to above that $S_T^{-1}\Upsilon'Pu/\mu_T = O_p(1/\tilde{\mu}_T)$ and $S_T^{-1}\Upsilon'u/\mu_T = O_p(1/\tilde{\mu}_T)$. Also by Lemma A1 and M,

$$\begin{aligned} S_T^{-1}\left(V'Pu - \frac{K}{T}V'u\right)/\mu_T &= S_T^{-1}(K\sigma_{Vu} - (K/T)T\sigma_{Vu})/\mu_T + O_p(\sqrt{K}/\mu_T^2) \\ &= O_p(1/\tilde{\mu}_T). \end{aligned}$$

Then by $X = \Upsilon + V$ and T we have $S_T^{-1}(X'Pu - \hat{\alpha}X'u)/\mu_T = O_p(1/\tilde{\mu}_T)$. Also, by H_T bounded and the first conclusion, $\hat{H}_T = S_T^{-1}(X'PX - \hat{\alpha}X'X)S_T^{-1\prime} = O_p(1)$. Then the last conclusion follows by T and

$$S_T^{-1}(X'P\hat{u} - \hat{\alpha}X'\hat{u})/\mu_T = S_T^{-1}(X'Pu - \hat{\alpha}X'u)/\mu_T - \hat{H}_TS_T'(\hat{\delta} - \delta_0)/\mu_T.$$

Lemma A6. If Assumptions 1–4 are satisfied and $S_T'(\hat{\delta} - \delta_0)/\mu_T = O_p(\varepsilon_T)$ for $\varepsilon_T \longrightarrow 0$ and $\varepsilon_T \ge 1/\mu_T$, then $\hat{u}'P \times \hat{u}/\hat{u}'\hat{u} = \breve{\alpha} + O_p(\varepsilon_T^2\mu_T^2/T)$.

Proof. Let $\hat{\beta} = S_T'(\hat{\delta} - \delta_0)/\mu_T$. Also, $\hat{\sigma}_u^2 = \hat{u}'\hat{u}/T$ satisfies $1/\hat{\sigma}_u^2 = O_p(1)$ by M. Therefore, $\tilde{H}_T = S_T^{-1}(X'PX - \breve{\alpha}X'X)S_T^{-1\prime} = O_p(1)$ and $S_T^{-1}(X'Pu - \breve{\alpha}X'u)/\mu_T = O_p(1/\mu_T)$ by Lemma A5 with $\hat{\alpha} = \breve{\alpha}$ and $\varepsilon_T^{\alpha} = \varepsilon_T^{\delta} = 0$ there, so that

$$\begin{aligned} \frac{\hat{u}'P\hat{u}}{\hat{u}'\hat{u}} - \breve{\alpha} &= \frac{1}{\hat{u}'\hat{u}}(\hat{u}'P\hat{u} - u'Pu - \breve{\alpha}(\hat{u}'\hat{u} - u'u)) \\ &= \frac{\mu_T^2}{T}\frac{1}{\hat{\sigma}_u^2}\big(\hat{\beta}'S_T^{-1}(X'PX - \breve{\alpha}X'X)S_T^{-1\prime}\hat{\beta} \\ &\qquad - 2\hat{\beta}'S_T^{-1}(X'Pu - \breve{\alpha}X'u)/\mu_T\big) \\ &= O_p\left(\frac{\mu_T^2}{T}\varepsilon_T^2\right). \end{aligned}$$

Proof of Theorem 1

By $\hat{\alpha} = K/T + o_p(\mu_T^2/T)$ there exists $\zeta_T \longrightarrow 0$, such that $\hat{\alpha} = K/T + O_p(\zeta_T\mu_T^2/T)$. Then by Lemma A4 and T, $\hat{\alpha} = \breve{\alpha} + O_p(\sqrt{K}/T + \zeta_T\mu_T^2/T)$. Then by Lemma A5 with $\varepsilon_T^{\alpha} = \sqrt{K}/T + \zeta_T\mu_T^2/T$ we have

$$\begin{aligned} S_T^{-1}(X'PX - \hat{\alpha}X'X)S_T^{-1\prime} &= H_T + O_p(\Delta_T^2 + \tilde{\mu}_T^{-1} + \zeta_T + \sqrt{K}/\mu_T^2) \\ &= H_T + o_p(1). \end{aligned}$$

Also $S_T^{-1}(X'Pu - \hat{\alpha}X'u)/\mu_T \longrightarrow_p 0$ by Lemma A5 with $\varepsilon_T^{\delta} = 0$. By uniform nonsingularity of H_T we have $(H_T + o_p(1))^{-1} = O_p(1)$. Then we have

$$\begin{aligned} S_T'(\hat{\delta} - \delta_0)/\mu_T &= S_T'(X'PX - \hat{\alpha}X'X)^{-1}(X'Pu - \hat{\alpha}X'u)/\mu_T \\ &= [S_T^{-1}(X'PX - \hat{\alpha}X'X)S_T^{-1\prime}]^{-1}S_T^{-1}(X'Pu - \hat{\alpha}X'u)/\mu_T \\ &= (H_T + o_p(1))^{-1}o_p(1) \xrightarrow{p} 0. \end{aligned}$$

For LIML, the conclusion follows by Lemma A3. For FULL, note $S_T'(\hat{\delta}_{LIML} - \delta_0)/\mu_T \longrightarrow_p 0$ implies that there is $\varepsilon_T \longrightarrow 0$ with $S_T'(\hat{\delta}_{LIML} - \delta_0)/\mu_T = O_p(\varepsilon_T)$, so by Lemma A6 we have $\hat{\alpha}_{LIML} = \hat{u}'P\hat{u}/\hat{u}'\hat{u} = \breve{\alpha} + O_p(\varepsilon_T\mu_T^2/T) = o_p(\mu_T^2/T)$. Also, $(T/\mu_T^2)(\sqrt{K}/T) = \sqrt{K}/\mu_T^2 \longrightarrow 0$, so that $O_p(\sqrt{K}/T) = o_p(\mu_T^2/T)$. Then $\breve{\alpha} = K/T + o_p(\mu_T^2/T)$ by Lemma A4 so that $\hat{\alpha}_{LIML} = K/T + o_p(\mu_T^2/T)$ by T. Also, $(T/\mu_T^2)(1/T) = 1/\mu_T^2 \longrightarrow 0$, so by T,

$$\begin{aligned} \hat{\alpha}_{FULL} &= \hat{\alpha}_{LIML} + O_p(1/T) = \hat{\alpha}_{LIML} + o_p(\mu_T^2/T) \\ &= K/T + o_p(\mu_T^2/T). \end{aligned}$$

Let $\hat{D}(\delta) = \partial[u(\delta)'Pu(\delta)/2u(\delta)'u(\delta)]/\partial\delta = X'Pu(\delta) - \tilde{\alpha}(\delta)X'u(\delta)$.

Lemma A7. If Assumptions 1–4 are satisfied and $S_T'(\bar{\delta} - \delta_0)/\mu_T = O_p(\varepsilon_T)$ for $\varepsilon_T \longrightarrow 0$, then

$$-S_T^{-1}[\partial\hat{D}(\bar{\delta})/\partial\delta]S_T^{-1\prime} = H_T + O_p(\Delta_T^2 + \tilde{\mu}_T^{-1} + \varepsilon_T).$$

Proof. Let $\bar{u} = u(\bar{\delta}) = y - X\bar{\delta}$ and $\bar{\gamma} = X'\bar{u}/\bar{u}'\bar{u}$. Then differentiating gives

$$-\frac{\partial \hat{D}}{\partial \delta}(\bar{\delta}) = X'PX - \frac{\bar{u}'P\bar{u}}{\bar{u}'\bar{u}}X'X - X'\bar{u}\frac{\bar{u}'PX}{\bar{u}'\bar{u}} - \frac{X'P\bar{u}}{\bar{u}'\bar{u}}\bar{u}'X + 2\frac{\bar{u}P\bar{u}}{(\bar{u}'\bar{u})^2}X'\bar{u}\bar{u}'X$$
$$= X'PX - \bar{\alpha}X'X + \bar{\gamma}\hat{D}(\bar{\delta})' + \hat{D}(\bar{\delta})\bar{\gamma}',$$
$$\bar{\alpha} = \bar{u}'P\bar{u}/\bar{u}'\bar{u} = \tilde{\alpha}(\bar{\delta}).$$

By Lemma A6 we have $\bar{\alpha} = \breve{\alpha} + O_p(\varepsilon_T^2\mu_T^2/T)$. Then by Lemma A5 with $\varepsilon_T^{\alpha} = \varepsilon_T^2\mu_T^2/T$ and $\varepsilon_T^{\delta} = \varepsilon_T$ we have

$$S_T^{-1}(X'PX - \bar{\alpha}X'X)S_T^{-1\prime} = H_T + O_p(\Delta_T^2 + \tilde{\mu}_T^{-1} + \varepsilon_T^2),$$
$$\mu_T^{-1}S_T^{-1}\hat{D}(\bar{\delta}) = S_T^{-1}(X'P\bar{u} - \bar{\alpha}X'\bar{u})/\mu_T = O_p(\tilde{\mu}_T^{-1} + \varepsilon_T).$$

Note that by standard arguments $\bar{\gamma} = O_p(1)$, so that $\mu_T S_T^{-1}\bar{\gamma} = O_p(1)$, and hence

$$S_T^{-1}\hat{D}(\bar{\delta})\bar{\gamma}'S_T^{-1\prime} = \mu_T^{-1}S_T^{-1}\hat{D}(\bar{\delta})O_p(1) = O_p(\tilde{\mu}_T^{-1} + \varepsilon_T).$$

The conclusion then follows by T.

Next, we give an expansion that is useful for the asymptotic normality results. Let $W = [(1-\tau_T)\Upsilon + P\tilde{V} - \tau_T\tilde{V}]S_T^{-1\prime}$ as in the text.

Lemma A8. If Assumptions 1–4 are satisfied, then

$$S_T^{-1}\hat{D}(\delta_0) = W'u + O_p\left(\frac{\sqrt{K}}{\mu_T\sqrt{T}} + \Delta_T\right).$$

Proof. Let $\breve{\alpha} = u'Pu/u'u$. By Lemma A4, $\breve{\alpha} = K/T + O_p(\sqrt{K}/T)$. Also, $S_T^{-1}\Upsilon'u = z'u/\sqrt{T} = O_p(1)$ and $S_T^{-1}\tilde{V}'u = O_p(\sqrt{T}/\mu_T)$ by M, so that $S_T^{-1}(\Upsilon + \tilde{V})'uO_p(\sqrt{K}/T) = O_p(\sqrt{K}/\mu_T\sqrt{T})$. Note also that similarly to the proof of Lemma A5, we have

$$E\left[\|S_T^{-1}\Upsilon'(I-P)u\|^2\right] = \sigma_u^2\,\mathrm{tr}(z'(I-P)z/T) = O_p(\Delta_T^2),$$

so by M, $S_T^{-1}\Upsilon'(I-P)u = O_p(\Delta_T)$. It then follows by T and $\breve{\alpha} = K/T + O_p(\sqrt{K}/T)$ that

$$S_T^{-1}\hat{D}(\delta_0) = S_T^{-1}[(X - u\gamma')'Pu - \breve{\alpha}(X - u\gamma')'u]$$
$$= S_T^{-1}\left\{\Upsilon'u + \tilde{V}'Pu - (\Upsilon + \tilde{V})'u\left[\frac{K}{T} + O_p\left(\frac{\sqrt{K}}{T}\right)\right] - \Upsilon'(I-P)u\right\}$$
$$= W'u + O_p\left(\frac{\sqrt{K}}{\mu_T\sqrt{T}} + \Delta_T\right).$$

Let $\tilde{\mu}_T = \mu_T$ in case (I) and $\tilde{\mu}_T = \mu_T^2/\sqrt{K}$ in case (II) and let $\bar{V} = (I-P)\tilde{V}$.

Lemma A9. If Assumptions 1–4 are satisfied and $S_T'(\hat{\delta} - \delta_0)/\mu_T = O_p(1/\tilde{\mu}_T)$, then

$$\|\hat{V} - \bar{V}\|^2/T = O_p(\Delta_T^2 + \tilde{\mu}_T^{-2}) \xrightarrow{p} 0,$$
$$\bar{V}'\bar{V}/T = (1-\tau_T)\tilde{\Omega} + O_p(\Delta_T + 1/\tilde{\mu}_T),$$
$$\hat{V}'\hat{V}/T = (1-\tau_T)\tilde{\Omega} + O_p(\Delta_T + 1/\tilde{\mu}_T).$$

Proof. By Lemma A1 we have $\tilde{V}'P\tilde{V}/T = \tau_T\tilde{\Omega} + O_p(\sqrt{K}/T) = \tau_T\tilde{\Omega} + O_p(1/\sqrt{T})$. Also, by CLT, $\tilde{V}'\tilde{V}/T = \tilde{\Omega} + O_p(1/\sqrt{T})$, so that by the CLT,

$$\bar{V}'\bar{V}/T = \tilde{V}'\tilde{V}/T - \tilde{V}'P\tilde{V}/T = (1-\tau_T)\tilde{\Omega} + O_p(1/\sqrt{T}).$$

Note that by construction $\mu_T^2 S_T^{-1}S_T^{-1\prime} \le CI$ so that $\|\mu_T S_T^{-1\prime}a\| \le C\|a\|$. Therefore, $\|\hat{\delta} - \delta_0\| \le \|\mu_T S_T^{-1\prime}S_T'(\hat{\delta} - \delta_0)/\mu_T\| \le \|S_T'(\hat{\delta} - \delta_0)/\mu_T\| = O_p(1/\tilde{\mu}_T)$. Then by $X'X = O_p(T)$ we have

$$\|u - \hat{u}\|^2/T \le \|X\|^2\|\hat{\delta} - \delta_0\|^2/T \le (\|X\|^2/T)O_p(\tilde{\mu}_T^{-2}) = O_p(\tilde{\mu}_T^{-2}).$$

It then follows by standard calculations that for $\hat{\gamma} = X'\hat{u}/\hat{u}'\hat{u}$, $\|\hat{\gamma} - \gamma\|^2 = O_p(\tilde{\mu}_T^{-2})$. Note that $\hat{V} - \bar{V} = (I-P)(\Upsilon + u\gamma' - \hat{u}\hat{\gamma}')$. Also by $S_T S_T'/T \le I$ we have

$$\mathrm{tr}[\Upsilon'(I-P)\Upsilon/T] = \mathrm{tr}[S_T z'(I-P)zS_T'/T^2] = O_p(\Delta_T^2).$$

Then it follows that

$$\|\hat{V} - \bar{V}\|^2/T \le C\|u\gamma' - \hat{u}\hat{\gamma}'\|^2/T + C\,\mathrm{tr}[\Upsilon'(I-P)\Upsilon/T],$$

giving the first conclusion. It then follows by standard arguments that

$$\hat{V}'\hat{V}/T - \bar{V}'\bar{V}/T = O_p(\Delta_T + \tilde{\mu}_T^{-1}).$$

The final conclusion then follows by T.

Let $\hat{a} = (\hat{u}_1^2 - \sigma_u^2, \ldots, \hat{u}_T^2 - \sigma_u^2)'$ and $a = (u_1^2 - \sigma_u^2, \ldots, u_T^2 - \sigma_u^2)'$.

Lemma A10. If Assumptions 1–4 are satisfied and $S_T'(\hat{\delta} - \delta_0)/\mu_T = O_p(1/\tilde{\mu}_T)$, then

$$S_T^{-1}\hat{A}(\hat{\delta})S_T^{-1\prime} = (1-\tau_T)A_T + O_p\big((\sqrt{K}/\mu_T)(1/\tilde{\mu}_T + \Delta_T)\big).$$

Proof. By Z including a constant we have $\sum_t \hat{u}_t^2\hat{V}_t/T = \hat{V}'\hat{a}/T$. Also, $\|\hat{a} - a\|^2/T = O_p(\tilde{\mu}_T^{-2})$ follows by standard arguments and $\|\hat{V} - \bar{V}\|^2/T = O_p(\Delta_T^2 + \tilde{\mu}_T^{-2})$ by Lemma A9. By Lemma A9 $\bar{V}'\bar{V}/T = O_p(1)$ and $a'a/T = O_p(1)$ by M, so by CS,

$$\frac{\sum_t \hat{u}_t^2\hat{V}_t'}{T} - a'\bar{V}/T = (\hat{a} - a)'(\hat{V} - \bar{V})/T + (\hat{a} - a)'\bar{V}/T + a'(\hat{V} - \bar{V})/T = O_p(1/\tilde{\mu}_T + \Delta_T).$$

It also follows by Lemma A1, similarly to the proof of Lemma A9, that $a'\bar{V}/T = (1-\tau_T)E[u_t^2\tilde{V}_t] + O_p(1/\sqrt{T})$, so it follows by T that

$$\sum_t \hat{u}_t^2\hat{V}_t/T = (1-\tau_T)E[u_t^2\tilde{V}_t] + O_p(1/\tilde{\mu}_T + \Delta_T).$$

Let $d_t = (p_{tt} - \tau_T)/\sqrt{K}$ and $d = (d_1, \ldots, d_T)'$. Note that $\|d\|^2 \le 1$ and $E[\|V'Pd\|^2] \le Cd'd \le C$, so that $V'Pd = O_p(1)$ and $\|S_T^{-1}\Upsilon d\| \le \|z/\sqrt{T}\|\|d\| \le C$. Also, $S_T^{-1}\Upsilon'(I-P)d = O_p(\Delta_T)$. Then

$$\begin{aligned}\sum_{t=1}^{T} S_T^{-1}\hat{\Upsilon}_t(p_{tt} - \tau_T)/\sqrt{K}&\\ &= S_T^{-1}X'Pd = S_T^{-1}(\Upsilon + V)P'd\\ &= z'd/\sqrt{T} + O_p(\Delta_T + 1/\mu_T).\end{aligned}$$

Then we have, for $\varepsilon_T = \Delta_T + 1/\mu_T$,

$$\begin{aligned}S_T^{-1}\hat{A}(\hat{\delta})S_T^{-1\prime} &= [z'd/\sqrt{T} + O_p(\varepsilon_T)]\\ &\quad \times \{(1-\tau_T)E[u_t^2\tilde{V}_t'] + O_p(\varepsilon_T)\}\sqrt{K}S_T^{-1\prime}\\ &= A_T + O_p(\varepsilon_T)O_p(\sqrt{K}/\mu_T)\\ &= A_T + O_p\big((\sqrt{K}/\mu_T)(1/\tilde{\mu}_T + \Delta_T)\big).\end{aligned}$$

Lemma A11. If Assumptions 1–4 are satisfied, then

$$\begin{aligned}\sum_{t=1}^{T}(\hat{u}_t^2 - \hat{\sigma}_u^2)\hat{V}_t\hat{V}_t'/T = (1 - 2\tau_T + \tau_T\kappa_T)E[(u_t^2 - \sigma_u^2)\tilde{V}_t\tilde{V}_t']&\\ + O_p(1/\tilde{\mu}_T + \Delta_T).&\end{aligned}$$

Proof. Let $A = \operatorname{diag}(a_1, \ldots, a_T)$. Let ε and v be columns of $\tilde{V}$ and $\bar{\varepsilon} = (I-P)\varepsilon$, $\bar{v} = (I-P)v$, so that $\sum_t a_t\bar{\varepsilon}_t\bar{v}_t/T$ is an element of $\sum_t a_t\bar{V}_t\bar{V}_t'/T$. We also have

$$\sum_t a_t\bar{\varepsilon}_t\bar{v}_t/T = \varepsilon'(I-P)A(I-P)v/T.$$

By CLT, $\varepsilon'Av/T = \sum_t a_t\varepsilon_t v_t/T = E[a_t\varepsilon_t v_t] + O_p(1/\sqrt{T})$. Let $e = (1, \ldots, 1)'$ and $\overline{av} = E[a_t v_t]e$. Then

$$\begin{aligned}E[(\varepsilon'P\overline{av})^2/T^2] &= \overline{av}'PE[\varepsilon\varepsilon']P\overline{av}/T^2\\ &\le C\overline{av}'\overline{av}/T^2 = O(1/T),\end{aligned}$$

so that $(\varepsilon'P\overline{av})/T = O_p(1/\sqrt{T})$ by M. Also, by Lemma A1, $\varepsilon'P(Av - \overline{av})/T = \tau_T E[a_t\varepsilon_t v_t] + O_p(\sqrt{K}/T)$. Then by T it follows that $\varepsilon'PAv = \tau_T E[a_t\varepsilon_t v_t] + O_p(1/\sqrt{T})$. It then follows similarly that $\varepsilon'APv = \tau_T E[a_t\varepsilon_t v_t] + O_p(1/\sqrt{T})$.

Next, let $D = \operatorname{diag}(p_{11}, \ldots, p_{TT})$ and $H = P - D$. Then for any α with $\|\alpha\| = 1$,

$$\begin{aligned}1 \ge \alpha'P\alpha &= \alpha'P^2\alpha = \alpha'H^2\alpha + 2\alpha'HD\alpha + \alpha'D^2\alpha\\ &\ge \alpha'H^2\alpha + \alpha'D^2\alpha - 2(\alpha'H^2\alpha)^{1/2}(\alpha'D^2\alpha)^{1/2}\\ &= \big|(\alpha'H^2\alpha)^{1/2} - (\alpha'D^2\alpha)^{1/2}\big|^2.\end{aligned}$$

Note that $\alpha'D^2\alpha \le 1$ by $p_{tt}^2 \le 1$ so that $(\alpha'H^2\alpha)^{1/2} \le 2$. Then for $\alpha = De/(\sum_{t=1}^T p_{tt}^2)^{1/2}$,

$$\begin{aligned}E[(\varepsilon'HD\overline{av})^2]/T^2 &\le C\frac{e'DH^2De}{T^2} = \alpha'H^2\alpha\frac{\sum_{t=1}^T p_{tt}^2}{T^2}\\ &\le C/T,\end{aligned}$$

so that $\varepsilon'HD\overline{av}/T = O_p(1/\sqrt{T})$ by M. Also, for $w_t = (a_t v_t - E[a_t v_t])p_{tt}$ we have

$$\begin{aligned}E[(\varepsilon'HADv &- \varepsilon'HD\overline{av})^2]/T^2\\ &= E\bigg[\bigg(\sum_{s\ne t}\varepsilon_s p_{st}w_t\bigg)^2\bigg]/T^2\\ &= \sum_{t\ne s}\sum_{i\ne j}p_{st}p_{ij}E[\varepsilon_s w_t\varepsilon_i w_j]/T^2\\ &= \sum_{t\ne s}p_{st}^2(E[\varepsilon_s^2]E[w_t^2] + E[\varepsilon_s w_s]E[\varepsilon_t w_t])/T^2\\ &\le C\sum_{s,t}p_{st}^2/T^2 = C\sum_t p_{tt}/T^2 \le C/T,\end{aligned}$$

so that

$$\varepsilon'HADv - \varepsilon'HD\widehat{av}/T = O_p\bigg(\frac{1}{\sqrt{T}}\bigg).$$

Then by T, $\varepsilon'HADv/T = O_p(1/\sqrt{T})$. We also have

$$\begin{aligned}E\bigg[\bigg(\varepsilon'DADv &- \sum_t p_{tt}^2E[a_t\varepsilon_t v_t]\bigg)^2/T^2\bigg]\\ &= \sum_{t=1}^T p_{tt}^4(E[a_t^2\varepsilon_t^2v_t^2] - E[a_t\varepsilon_t v_t]^2)/T^2\\ &= O(1/T),\end{aligned}$$

so that $\varepsilon'DADv/T = (\sum_t p_{tt}^2/T)E[a_t\varepsilon_t v_t] + O_p(1/\sqrt{T}) = \tau_T \times \kappa_T E[a_t\varepsilon_t v_t] + O_p(1/\sqrt{T})$.

Next, let L be an upper triangular matrix with zero diagonal such that $L + L' = H$. Consider $\varepsilon'HAHv/T = \varepsilon'(L+L')A(L+L')v/T$. Note that

$$\varepsilon'LAL'v/T = \sum_t a_t\bigg(\sum_{j<t}p_{jt}\varepsilon_j\bigg)\bigg(\sum_{k<t}p_{kt}v_k\bigg)/T$$

is an average of a martingale difference. Therefore,

$$\begin{aligned}E[(\varepsilon'LAL'v)^2/T^2]&\\ &= \sum_t E[a_t^2]E\bigg[\bigg(\sum_{j<t}p_{jt}\varepsilon_j\bigg)^2\bigg(\sum_{k<t}p_{kt}v_k\bigg)^2\bigg]/T^2\\ &\le C\sum_t\sum_{j,k,\ell,m<t}p_{jt}p_{kt}p_{\ell t}p_{mt}E[\varepsilon_j\varepsilon_k v_\ell v_m]/T^2\\ &\le C\sum_t\sum_{j,k<t}p_{jt}^2p_{kt}^2(E[\varepsilon_j^2]E[v_k^2] + 2E[\varepsilon_j v_j]E[\varepsilon_k v_t])/T^2\\ &\le C\sum_t\bigg(\sum_j p_{jt}^2\bigg)\bigg(\sum_k p_{kt}^2\bigg)/T^2 = \sum_t p_{tt}^2/T^2 = O\bigg(\frac{1}{T}\bigg).\end{aligned}$$

Thus $\varepsilon'LAL'v/T = O_p(1/\sqrt{T})$ by M. It follows similarly that $\varepsilon'L'ALv/T = O_p(1/\sqrt{T})$. We also have

$$\begin{aligned}\varepsilon'LALv/T &= \sum_t a_t\bigg(\sum_{j<t}p_{jt}\varepsilon_j\bigg)\bigg(\sum_{k>t}p_{kt}v_k\bigg)\\ &= \sum_{j<t<k}p_{jt}p_{kt}a_t\varepsilon_j v_k.\end{aligned}$$

Therefore, since for $j<t<k$, $\ell<s<m$, $E[a_t a_s \varepsilon_j \varepsilon_\ell v_k v_m]$ is nonzero only when $t=s$, $j=\ell$, $k=m$,

$$\begin{aligned}
&E[(\varepsilon' LALv/T)^2]\\
&\quad=\sum_{j<t<k}\sum_{\ell<s<m} p_{jt}p_{kt}p_{\ell s}p_{ms}E[a_t a_s \varepsilon_j \varepsilon_\ell v_k v_m]/T^2\\
&\quad=\sum_{j<t<k} p_{jt}^2 p_{kt}^2 E[a_t^2]E[\varepsilon_j^2]E[v_k^2]/T^2\\
&\quad\le C\sum_t\Big(\sum_j p_{jt}^2\Big)\Big(\sum_k p_{kt}^2\Big)/T^2 = C\sum_t p_{tt}^2/T^2\\
&\quad=O\Big(\frac{1}{T}\Big),
\end{aligned}$$

so that $\varepsilon' TATv/T = O_p(1/\sqrt{T})$. It follows similarly that $\varepsilon' T'AT'v/T = O_p(1/\sqrt{T})$. Then by T we have

$$\varepsilon' PAPv/T = \tau_T\kappa_T E[a_t\varepsilon_t v_t] + O_p\Big(\frac{1}{\sqrt{T}}\Big).$$

Also by CLT $\varepsilon' Av/T = E[a_t\varepsilon_t v_t] + O_p(1/\sqrt{T})$. Then by T, $\varepsilon'(I-P)A(I-P)v/T = (1-2\tau_T+\kappa_T\tau_T)E[a_t\varepsilon_t v_t]+O_p(1/\sqrt{T})$. Applying this result to each component, we have

$$\sum_t(u_t^2-\sigma_u^2)\bar V_t\bar V_t'/T = (1-2\tau_T+\kappa_T\tau_T)E[(u_t^2-\sigma_u^2)\bar V_t\bar V_t'] + O_p(1/\sqrt{T}).$$

Now, there is C big enough such that for $d_t = C(1+y_t^2+X_t'X_t)$, $(y_t-X_t'\delta)^2\le d_t$ and $|(y_t-X_t'\tilde\delta)^2-(y_t-X_t'\delta)^2|\le d_t\|\tilde\delta-\delta\|$ for all $\delta,\tilde\delta$ in some neighborhood of δ_0. It also follows similarly to previous arguments that by the fourth moment of d_t bounded in t, $\sum_t d_t\|\bar V_t\|^2/T = O_p(1)$. In particular, for $\tilde D = \operatorname{diag}(d_1,\ldots,d_T)$,

$$\begin{aligned}
E[\varepsilon' P\tilde D P\varepsilon]/T &= \sum_{j,k,t} p_{jt}p_{kt}E[d_t\varepsilon_j\varepsilon_k]\\
&=\sum_t p_{tt}^2(E[d_t\varepsilon_t^2]-E[d_t]E[\varepsilon_t^2])/T\\
&\quad+\sum_{j,t}p_{jt}^2E[d_t]E[\varepsilon_j^2]/T\\
&\le C
\end{aligned}$$

and $\varepsilon'\tilde D\varepsilon/T = \sum_t d_t\varepsilon_t^2/T = O_p(1)$, so that by CS,

$$\begin{aligned}
|\varepsilon' P\tilde D v/T| &\le (\varepsilon' P\tilde D P\varepsilon/T)^{1/2}(v'\tilde D v/T)^{1/2} = O_p(1),\\
|\varepsilon P\tilde D P v/T| &\le (\varepsilon' P\tilde D P\varepsilon/T)^{1/2}(vP\tilde D P v/T)^{1/2} = O_p(1).
\end{aligned}$$

It then follows that

$$\begin{aligned}
&\Big\|\sum_t[\hat u_t^2-\hat\sigma_u^2-(u_t^2-\sigma_u^2)]\bar V_t\bar V_t'/T\Big\|\\
&\quad\le O_p(1)(\|\hat\delta-\delta_0\|+\|\hat\sigma_u^2-\sigma_u^2\|)\\
&\quad=O_p(1/\tilde\mu_T).
\end{aligned}$$

We also have by CS and T,

$$\begin{aligned}
&\Big\|\sum_t(\hat u_t^2-\hat\sigma_u^2)(\hat V_t\hat V_t'-\bar V_t\bar V_t)/T\Big\|\\
&\quad\le\sum_t d_t(\|\hat V_t-\bar V_t\|^2+2\|\bar V_t\|\|\hat V_t-\bar V_t\|)/T\\
&\quad\le\sum_t d_t\|\hat V_t-\bar V_t\|^2/T\\
&\qquad+2\Big(\sum_t d_t\|\bar V_t\|^2/T\Big)^{1/2}\Big(\sum_t d_t\|\hat V_t-\bar V_t\|^2/T\Big)^{1/2}.
\end{aligned}$$

It follows similarly to previous arguments that

$$\sum_t d_t\|\hat V_t-\bar V_t\|^2/T = O_p(\Delta_T^2+\tilde\mu_T^{-2}).$$

The conclusion then follows by T.

Lemma A12. If Assumptions 1–4 are satisfied and $S_T'(\hat\delta-\delta_0)/\mu_T = O_p(\tilde\mu_T^{-1})$, then

$$S_T^{-1}\hat H(\hat\delta)S_T^{-1\prime} = H_T + O_p(\Delta_T^2+1/\tilde\mu_T).$$

Proof. By Lemma A6 with $\varepsilon_T = \tilde\mu_T^{-1}$ we have $\tilde\alpha(\hat\delta) = \breve\alpha + O_p(\mu_T^2/T\tilde\mu_T^2)$. The conclusion then follows by Lemma A5 with $\varepsilon_T^\alpha = \mu_T^2/T\tilde\mu_T^2$.

Lemma A13. If Assumptions 1–4 are satisfied and $S_T'(\hat\delta-\delta_0)/\mu_T = O_p(\tilde\mu_T^{-1})$, then

$$\begin{aligned}
&S_T^{-1}\tilde\alpha(\hat\delta)\tilde X(\hat\delta)'\tilde X(\hat\delta)S_T^{-1\prime}\\
&\quad=\tau_T(1-\tau_T)^{-1}H_T + KS_T^{-1}\tilde\Omega S_T^{-1\prime}+O_p(\tilde\mu_T^{-1}).
\end{aligned}$$

Proof. By Lemma A6 with $\varepsilon_T = \tilde\mu_T^{-1}$ we have $\hat\alpha = \tilde\alpha(\hat\delta) = \breve\alpha + O_p(\mu_T^2/T\tilde\mu_T^2)$. Also, note that $(T/\sqrt{K})\mu_T^2/T\tilde\mu_T^2 = \mu_T^2/\sqrt{K}\tilde\mu_T^2 = 1/\sqrt{K}$ in case (I) and is equal to $\sqrt{K}/\mu_T^2\longrightarrow 0$ in case (II), so that $O_p(\mu_T^2/T\tilde\mu_T^2) = o_p(\sqrt{K}/T)$. Then by T we have $\hat\alpha = \tau_T + O_p(\sqrt{K}/T) = O_p(K/T)$. Let $\hat X = \tilde X(\hat\delta)$ and $\tilde X = X - u\gamma' = \Upsilon + \tilde V$. It follows by standard arguments that $\|\hat X-\tilde X\| = O_p(\sqrt{T}/\tilde\mu_T)$ and $\|\tilde X\| = O_p(\sqrt{T})$, so that $\|\hat X'\hat X - \tilde X'\tilde X\| = O_p(T/\tilde\mu_T)$. Therefore we have

$$\begin{aligned}
&\|S_T^{-1}\hat\alpha\hat X'\hat X S_T^{-1\prime} - S_T^{-1}\hat\alpha\tilde X'\tilde X S_T^{-1\prime}\|\\
&\quad=O_p(K/T)O_p(1/\mu_T^2)O_p(T/\tilde\mu_T)\\
&\quad=O_p(\sqrt{K}/\mu_T^2\tilde\mu_T) = o_p(1/\tilde\mu_T).
\end{aligned}$$

We also have

$$\|(\hat\alpha-\tau_T)S_T^{-1}\tilde X'\tilde X S_T^{-1\prime}\| = O_p(T\sqrt{K}/T\mu_T^2) = O_p(1/\tilde\mu_T).$$

Furthermore, by M, $\tau_T S_T^{-1}\Upsilon'\tilde V S_T^{-1\prime} = O_p(K/T\mu_T) = O_p(1/\tilde\mu_T)$. Also, $K\sqrt{T}/T\mu_T^2 = (\sqrt{K}/\mu_T^2)\sqrt{K/T}\le C/\tilde\mu_T$ so that by M

$$\begin{aligned}
\tau_T S_T^{-1}\tilde V'\tilde V S_T^{-1\prime} &= \tau_T S_T^{-1}(T\tilde\Omega)S_T^{-1\prime} + O_p(K\sqrt{T}/T\mu_T^2)\\
&=KS_T^{-1}\tilde\Omega S_T^{-1\prime} + O_p(1/\tilde\mu_T).
\end{aligned}$$

It then follows by T that

$$
\begin{aligned}
S_T^{-1}\hat{\alpha}\hat{X}'\hat{X}S_T^{-1\prime} &= \tau_T S_T^{-1}\tilde{X}'\tilde{X}S_T^{-1\prime} + O_p(\tilde{\mu}_T^{-1}) \\
&= \tau_T S_T^{-1}(\Upsilon+\tilde{V})'(\Upsilon+\tilde{V})S_T^{-1\prime} + O_p(\tilde{\mu}_T^{-1}) \\
&= \tau_T S_T^{-1}\Upsilon'\Upsilon S_T^{-1\prime} + KS_T^{-1}\tilde{\Omega}S_T^{-1\prime} + O_p(\tilde{\mu}_T^{-1}).
\end{aligned}
$$

Lemma A14. If Assumptions 1–4 are satisfied and $S_T'(\hat{\delta}-\delta_0)/\mu_T = O_p(\tilde{\mu}_T^{-1})$, then

$$
S_T^{-1}\hat{\Sigma}(\hat{\delta})S_T^{-1\prime} = \Sigma_T + O_p\big((1+\sqrt{K}/\mu_T)(1/\tilde{\mu}_T+\Delta_T)\big).
$$

Proof. By standard arguments we have $\hat{\sigma}_u^2(\hat{\delta}) = \sigma_u^2 + O_p(1/\tilde{\mu}_T)$ and it follows as in the proof of Lemma A13 that $\tilde{\alpha}(\hat{\delta}) = \tau_T + O_p(\sqrt{K}/T)$. It also follows similarly to the proofs of Lemmas A5 and A9 that

$$
\begin{aligned}
&S_T^{-1}(\hat{X}'P\hat{X} - \hat{\alpha}\hat{X}'\hat{X})S_T^{-1\prime} \\
&\quad = S_T^{-1}(\tilde{X}'P\tilde{X} - \tilde{\alpha}\tilde{X}'\tilde{X})S_T^{-1\prime} + O_p(1/\tilde{\mu}_T) \\
&\quad = H_T + O_p(\Delta_T^2 + 1/\tilde{\mu}_T).
\end{aligned}
$$

Also, we have $O_p(\sqrt{K}/T)KS_T^{-1}\tilde{\Omega}S_T^{-1\prime} = O_p((K/T)(\sqrt{K}/\mu_T^2)) = O_p(1/\tilde{\mu}_T)$. Note that

$$
\hat{\Sigma}_B(\hat{\delta}) = \hat{\sigma}_u^2[(1-2\hat{\alpha})(\hat{X}'P\hat{X} - \hat{\alpha}\hat{X}'\hat{X}) + \hat{\alpha}(1-\hat{\alpha})\hat{X}'\hat{X}].
$$

Then by Lemma A13 and T it follows that

$$
\begin{aligned}
&S_T^{-1}\hat{\Sigma}_B(\hat{\delta})S_T^{-1\prime} \\
&\quad = (\sigma_u^2 + O_p(1/\tilde{\mu}_T)) \\
&\qquad \times \big\{(1-2\tau_T + O_p(\sqrt{K}/T))(H_T + O_p(\Delta_T^2 + 1/\tilde{\mu}_T)) \\
&\qquad + (1-\tau_T + O_p(\sqrt{K}/T)) \\
&\qquad \times \big(\tau_T(1-\tau_T)^{-1}H_T + KS_T^{-1}\tilde{\Omega}S_T^{-1\prime} + O_p(1/\tilde{\mu}_T)\big)\big\} \\
&\quad = \sigma_u^2\{(1-2\tau_T)H_T + \tau_T H_T + (1-\tau_T)KS_T^{-1}\tilde{\Omega}S_T^{-1\prime}\} \\
&\qquad + O_p(\Delta_T^2 + 1/\tilde{\mu}_T) \\
&\quad = \sigma_u^2(1-\tau_T)(H_T + KS_T^{-1}\tilde{\Omega}S_T^{-1\prime}) + O_p(\Delta_T^2 + 1/\tilde{\mu}_T).
\end{aligned}
$$

The conclusion now follows by Lemmas A10 and A11 and T.

Lemma A15. If Assumptions 1–5 are satisfied and $S_T'(\hat{\delta}-\delta_0)/\mu_T = O_p(\tilde{\mu}_T^{-1})$, then in case (I), $S_T'\hat{\Lambda}S_T - \Lambda_T = O_p(1/\tilde{\mu}_T + \Delta_T)$, and in case II, $(\mu_T^2/K)(S_T'\hat{\Lambda}S_T - \Lambda_T) = O_p(1/\tilde{\mu}_T + \Delta_T)$.

Proof. Let $\hat{H} = S_T^{-1}\hat{H}(\hat{\delta})S_T^{-1\prime}$. Note that H_T is uniformly nonsingular by τ_T bounded away from 1 and uniform nonsingularity of $z'z/T$. Then by Lemma A12 we have, in both cases,

$$
\hat{H}^{-1} = H_T^{-1} + O_p(\tilde{\mu}_T^{-1} + \Delta_T),
$$

$$
\hat{H}^{-1} = O_p(1), \qquad H_T^{-1} = O(1).
$$

In case (I) note that $\sqrt{K}/\mu_T$ is bounded, so that by Lemma A14, $S_T^{-1}\hat{\Sigma}(\hat{\delta})S_T^{-1\prime} = \Sigma_T + O_p(1/\tilde{\mu}_T + \Delta_T)$ and $\Sigma_T = O(1)$. The conclusion then follows by

$$
\begin{aligned}
S_T'\hat{\Lambda}S_T &= \hat{H}^{-1}S_T^{-1}\hat{\Sigma}(\hat{\delta})S_T^{-1\prime}\hat{H}^{-1} \\
&= [H_T^{-1} + O_p(\tilde{\mu}_T^{-1} + \Delta_T)][\Sigma_T + O_p(1/\tilde{\mu}_T + \Delta_T)] \\
&\quad \times [H_T^{-1} + O_p(\tilde{\mu}_T^{-1} + \Delta_T)] \\
&= \Lambda_T + O_p(1/\tilde{\mu}_T + \Delta_T).
\end{aligned}
$$

In case (II) note that by Lemma A14,

$$
(\mu_T^2/K)S_T^{-1}\hat{\Sigma}(\hat{\delta})S_T^{-1\prime} = (\mu_T^2/K)\Sigma_T + O_p(1/\tilde{\mu}_T + \Delta_T),
$$

and that $(\mu_T^2/K)\Sigma_T = O(1)$. The conclusion then follows from

$$
\begin{aligned}
&(\mu_T^2/K)S_T'\hat{\Lambda}S_T \\
&\quad = (\mu_T^2/K)\hat{H}^{-1}S_T^{-1}\hat{\Sigma}(\hat{\delta})S_T^{-1\prime}\hat{H}^{-1} \\
&\quad = [H_T^{-1} + O_p(\tilde{\mu}_T^{-1} + \Delta_T)][(\mu_T^2/K)\Sigma_T + O_p(1/\tilde{\mu}_T + \Delta_T)] \\
&\qquad \times [H_T^{-1} + O_p(\tilde{\mu}_T^{-1} + \Delta_T)] \\
&\quad = (\mu_T^2/K)\Lambda_T + O_p(1/\tilde{\mu}_T + \Delta_T).
\end{aligned}
$$

Proof of Theorem 2

Consider first the case where $\hat{\delta}$ is LIML. Then $\mu_T^{-1}S_T'(\hat{\delta}-\delta_0) \longrightarrow_p 0$ by Theorem 1, implying $\hat{\delta} \longrightarrow_p \delta_0$. The first-order conditions for LIML are $\hat{D}(\hat{\delta}) = 0$. Expanding gives

$$
0 = \hat{D}(\delta_0) + \frac{\partial\hat{D}}{\partial\delta}(\bar{\delta})(\hat{\delta}-\delta_0),
$$

where $\bar{\delta}$ lies on the line joining $\hat{\delta}$ and δ_0 and hence $\bar{\beta} = \mu_T^{-1}S_T'(\bar{\delta}-\delta_0) \longrightarrow_p 0$. Then there is $\varepsilon_T \longrightarrow 0$ such that $\bar{\beta} = O_p(\varepsilon_T)$, so by Lemma A5, $\bar{H}_T = S_T^{-1}[\partial\hat{D}(\bar{\delta})/\partial\delta]S_T^{-1\prime} = H_T + o_p(1)$. Then $\partial\hat{D}(\bar{\delta})/\partial\delta$ is nonsingular w.p.a.1 and solving gives

$$
\begin{aligned}
S_T'(\hat{\delta}-\delta) &= -S_T'[\partial\hat{D}(\bar{\delta})/\partial\delta]^{-1}\hat{D}(\delta_0) \\
&= -\bar{H}_T^{-1}S_T^{-1}\hat{D}(\delta_0).
\end{aligned}
$$

Next, apply Lemma A2 with $V_t = \tilde{V}_t$ and

$$
W_{tT} = \begin{pmatrix} S_T^{-1}(1-\tau_T)\Upsilon_t u_t \\ K^{-1/2}(p_{tt}-\tau_T)\tilde{V}_t u_t \end{pmatrix}.
$$

By u_t having bounded fourth moment,

$$
\sum_{t=1}^T E[\|S_T^{-1}\Upsilon_t u_t\|^4] \le C\sum_{t=1}^T \|z_t\|^4/T^2 \longrightarrow 0.
$$

Also, by u_t and V_t having bounded eighth moment and $p_{tt}^4 \le K$,

$$
\begin{aligned}
\sum_{t=1}^T E\big[\|K^{-1/2}(p_{tt}-\tau_T)\tilde{V}_t u_t\|^4\big] &\le C\left[\sum_{t=1}^T p_{tt}^4 + T\tau_T^4\right]\Big/K^2 \\
&\le \frac{C}{K} + \tau_T^2/T \longrightarrow 0.
\end{aligned}
$$

By Assumption 3, we have

$$
\sum_{t=1}^T E[W_{tT}W_{tT}'] \longrightarrow \begin{bmatrix} \sigma_u^2(1-\tau)H & (1-\tau)A' \\ (1-\tau)A & (\kappa-\tau)(\tilde{\Omega}+B) \end{bmatrix} = \bar{\Psi}.
$$

Let $\Gamma = \operatorname{diag}(\bar{\Psi}, \sigma_u^2\tilde{\Omega}(1-\kappa))$ and

$$U_T = \begin{pmatrix} \sum_{t=1}^T W_{tT} \\ \sum_{t\neq s} \tilde{V}_t p_{ts} u_s/\sqrt{K} \end{pmatrix}.$$

Consider c such that $c'\Gamma c > 0$. Then by the conclusion of Lemma A2 we have $c'U_T \longrightarrow_d \mathrm{N}(0, c'\Gamma c)$. Also, if $c'\Gamma c = 0$, then it is straightforward to show that $c'U_T \longrightarrow_p 0$. Then it follows that

$$U_T = \begin{pmatrix} \sum_{t=1}^T W_{tT} \\ \sum_{t\neq s} \tilde{V}_t p_{ts} u_s/\sqrt{K} \end{pmatrix} \xrightarrow{d} \mathrm{N}(0,\Gamma),$$
$$\Gamma = \operatorname{diag}(\bar{\Psi}, \sigma_u^2\tilde{\Omega}(1-\kappa)).$$

Next, we consider the two cases. Case (I) has K/μ_T^2 bounded. In this case $\sqrt{K}S_T^{-1} \longrightarrow S_0$, so that

$$F_T \stackrel{\text{def}}{=} [I, \sqrt{K}S_T^{-1}, \sqrt{K}S_T^{-1}] \longrightarrow F_0 = [I, S_0, S_0],$$
$$F_0\Gamma F_0' = \Lambda_I.$$

Then by Lemma A8 and S and $W'u = F_T U_T$,

$$\begin{aligned} S_T^{-1}\hat{D}(\delta_0) &= W'u + o_p(1) \\ &= F_T U_T + o_p(1) \xrightarrow{d} \mathrm{N}(0, \Lambda_I), \\ S_T'(\hat{\delta} - \delta_0) &= -\bar{H}_T^{-1} S_T^{-1}\hat{D}(\delta_0) \xrightarrow{d} \mathrm{N}(0, H^{-1}\Lambda_I H^{-1}). \end{aligned}$$

In case (II) we have $K/\mu_T^2 \longrightarrow \infty$. Here

$$(\mu_T/\sqrt{K})F_T \longrightarrow \bar{F}_0 = [0, \bar{S}_0, \bar{S}_0], \qquad \bar{F}_0\Gamma\bar{F}_0' = \Lambda_{II}$$

and $(\mu_T/\sqrt{K})o_p(1) = o_p(1)$. Then by Lemma A8 and S and $W'u = F_T U_T$,

$$\begin{aligned} &(\mu_T/\sqrt{K})S_T\hat{D}(\delta_0) \\ &\quad = (\mu_T/\sqrt{K})W'u + o_p(1) \\ &\quad = (\mu_T/\sqrt{K})F_T U_T + o_p(1) \xrightarrow{d} \mathrm{N}(0, \Lambda_{II}), \\ &(\mu_T/\sqrt{K})S_T'(\hat{\delta} - \delta_0) \\ &\quad = -\bar{H}_T^{-1}(\mu_T/\sqrt{K})S_T^{-1}\hat{D}(\delta_0) \xrightarrow{d} \mathrm{N}(0, H^{-1}\Lambda_{II}H^{-1}). \end{aligned}$$

Also, Lemma A15 gives the convergence of the covariance matrix estimators. Finally, if Σ_I is nonsingular, then by Lemma A14 we have $(S_T^{-1}\hat{\Sigma}(\hat{\delta})S_T^{-1\prime})^{-1} = \Sigma_T^{-1} + o_p(1)$, so that

$$\begin{aligned} \hat{K}(\delta_0) &= \hat{D}(\delta_0)' S_T^{-1\prime}(S_T^{-1}\hat{\Sigma}(\hat{\delta})S_T^{-1\prime})^{-1} S_T^{-1}\hat{D}(\delta_0) \\ &= \hat{D}(\delta_0)' S_T^{-1\prime}\Sigma_T^{-1} S_T^{-1}\hat{D}(\delta_0) + o_p(1) \\ &\xrightarrow{d} \chi^2(G). \end{aligned}$$

The result for case (II) follows similarly by replacing S_T by $(\mu_T/\sqrt{K})S_T$.

Let $\hat{t} = c'(\tilde{\delta} - \delta_0)/(c'\hat{\Lambda}c)^{1/2}$.

Proof of Theorem 3

First, consider LIML. Let $\bar{\delta}$ be the mean value as in the proof of Theorem 2. It follows similarly to the proof of Theorem 2 that $S_T'(\hat{\delta} - \delta_0)/\mu_T = O_p(\tilde{\mu}_T^{-1})$, so that $S_T'(\bar{\delta} - \delta_0)/\mu_T = O_p(\tilde{\mu}_T^{-1})$ also holds for the mean value. Then by Lemma A7 we have $S_T^{-1}[\partial\hat{D}(\bar{\delta})/\partial\delta]S_T^{-1\prime} = H_T + O_p(\Delta_T^2 + \tilde{\mu}_T^{-1})$. Also, by Lemma A8 we have $S_T^{-1}\hat{D}(\delta_0) = W'u + O_p(\Delta_T + \tilde{\mu}_T^{-1}) = O_p(1)$, so that in case (I), by $F_T = \mu_T^c c' S_T^{-1\prime}$ bounded,

$$\begin{aligned} &\mu_T^c c'(\tilde{\delta} - \delta_0) \\ &\quad = F_T\big[S_T^{-1}[\partial\hat{D}(\bar{\delta})/\partial\delta]S_T^{-1\prime}\big]^{-1} S_T^{-1}\hat{D}(\delta_0) \\ &\quad = F_T[H_T + O_p(\Delta_T^2 + \tilde{\mu}_T^{-1})]^{-1}[W'u + O_p(\Delta_T + \tilde{\mu}_T^{-1})] \\ &\quad = F_T H_T^{-1} W'u + O_p(\Delta_T + \tilde{\mu}_T^{-1}). \end{aligned}$$

Note also that Lemma A15 by F_T bounded,

$$(\mu_T^c)^2 c'\hat{\Lambda}c = F_T S_T'\hat{\Lambda}S_T F_T' = F_T\Lambda_T F_T' + O_p(\Delta_T + \tilde{\mu}_T^{-1}).$$

Then by $F_T\Lambda_T F_T'$ bounded and bounded away from zero we also have

$$((\mu_T^c)^2 c'\hat{\Lambda}c)^{-1/2} = (F_T\Lambda_T F_T')^{-1/2} + O_p(\Delta_T + \tilde{\mu}_T^{-1}).$$

The second conclusion now follows by the delta method and $F_T H_T^{-1} W'u = O_p(1)$, which gives

$$\begin{aligned} \hat{t} &= \frac{\mu_T^c c'(\tilde{\delta} - \delta_0)}{(\mu_T^{c2} c'\hat{\Lambda}c)^{1/2}} = \frac{F_T S_T'(\tilde{\delta} - \delta_0)}{(F_T S_T'\hat{\Lambda}S_T F_T')^{1/2}} \\ &= \frac{F_T H_T^{-1} W'u}{(F_T\Lambda_T F_T')^{1/2}} + O_p(\Delta_T + \tilde{\mu}_T^{-1}). \end{aligned}$$

The last conclusion, for case (I), follows similarly. In case (II) we have, by Lemma A15 and $F_T H_T^{-1} W'u\mu_T/\sqrt{K} = O_p(1)$,

$$\begin{aligned} \hat{t} &= \frac{\mu_T^c c'(\tilde{\delta} - \delta_0)}{(\mu_T^{c2} c'\hat{\Lambda}c)^{1/2}} \\ &= \frac{F_T S_T'(\tilde{\delta} - \delta_0)\mu_T/\sqrt{K}}{(F_T S_T'\hat{\Lambda}S_T F_T'\mu_T^2/K)^{1/2}} \\ &= \frac{F_T H_T^{-1} W'u\mu_T/\sqrt{K}}{(F_T\Lambda_T F_T'\mu_T^2/K)^{1/2}} \\ &= \frac{F_T H_T^{-1} W'u}{(F_T\Lambda_T F_T')^{1/2}} + O_p(\Delta_T + \tilde{\mu}_T^{-1}), \end{aligned}$$

giving the second conclusion in case (II). The first conclusion now follows from the second conclusion and Lemma A2.

For the third conclusion, let $\tilde{t} = c'(\tilde{\delta} - \delta_0)/(\hat{\sigma}_u^2 c'\hat{H}c)^{1/2}$ and $\hat{\rho} = (\hat{\sigma}_u^2 c'\hat{H}c)^{1/2}(c'\hat{\Lambda}c)^{-1/2}$, so that $\tilde{t} = \hat{t}/\hat{\rho}$. In case (II), by $(S_T^{-1}\hat{H}S_T^{-1\prime})^{-1}$ and $\hat{\sigma}_u^2$ bounded in probability and $F_T\Lambda_T F_T'\mu_T^2/K$ bounded away from zero, we have

$$\hat{\rho} = \frac{\{\hat{\sigma}_u^2 F_T(S_T^{-1}\hat{H}S_T^{-1\prime})^{-1}F_T'\mu_T^2/K\}^{1/2}}{\{F_T S_T'\hat{\Lambda}S_T F_T'\mu_T^2/K\}^{1/2}} \xrightarrow{p} 0.$$

Then by the Slutzky theorem, $(\hat{t}, \hat{\rho}) \longrightarrow_d (\mathrm{N}(0,1), 0)$ jointly. Therefore, for any $C, \varepsilon > 0$,

$$\begin{aligned} \Pr(|\tilde{t}| \geq C) &\geq \Pr(|\hat{t}| \geq C\varepsilon, |\hat{\rho}| < \varepsilon) \\ &\longrightarrow 1 - \{\Phi(C\varepsilon) - \Phi(-C\varepsilon)\}. \end{aligned}$$

For any C the expression on the right side can be made arbitrarily close to 1 by choosing ε small enough. Thus, $\Pr(|\tilde{t}| \geq C) \longrightarrow 1$.

To show the same result for estimators with $\hat{\alpha} = \tilde{\alpha} + O_p(1/T)$, note that

$$(\hat{\alpha} - \tilde{\alpha})S_T^{-1}X'XS_T^{-1\prime} = O_p(1/T)O_p(T/\mu_T^2) = O_p(1/\mu_T^2),$$
$$(\hat{\alpha} - \tilde{\alpha})S_T^{-1}X'u = O_p(1/T)O_p(T/\mu_T) = O_p(1/\mu_T).$$

Then it follows from the formula $(\hat{\delta} - \delta_0) = (X'PX - \hat{\alpha}X'X)^{-1} \times (X'Pu - \hat{\alpha}X'u)$ that

$$\begin{aligned}
&\mu_T^c c'(\hat{\delta} - \tilde{\delta}) \\
&= F_T S_T'(\hat{\delta} - \tilde{\delta}) \\
&= F_T[S_T^{-1}(X'PX - \hat{\alpha}X'X)S_T^{-1\prime}]^{-1}S_T(X'Pu - \hat{\alpha}X'u) \\
&\quad - F_T[S_T^{-1}(X'PX - \tilde{\alpha}X'X)S_T^{-1\prime}]^{-1}S_T(X'Pu - \tilde{\alpha}X'u) \\
&= O_p(1/\mu_T).
\end{aligned}$$

The results then follow as before, with this additional remainder present.

Lemma A16. If Assumptions 1–3 are satisfied, then $\sum_{t=1}^T E[\|\tilde{V}'Z(Z'Z)^{-}Z_t\|^3] \leq CK$.

Proof. Consider first the case where $\tilde{V}_t$ is a scalar. By the Marcinkiewicz–Zygmund inequality,

$$E[|\tilde{V}'Z(Z'Z)^{-}Z_t|^3] = E\left[\left|\sum_{s=1}^T \tilde{V}_s p_{st}\right|^3\right] \leq CE\left[\left|\sum_{s=1}^T \tilde{V}_s^2 p_{st}^2\right|^{3/2}\right].$$

By $p_{tt} \leq 1$ it follows that $p_{tt}^{3/2} \leq p_{tt}$. Also, $f(r) = r^{3/2}$ is a convex function of r. Then by Jensen's inequality and $\sum_t p_{st}^2 = p_{tt}$ we have

$$\begin{aligned}
E\left[\left|\sum_{s=1}^T \tilde{V}_s^2 p_{st}^2\right|^{3/2}\right] &\leq p_{tt}^{3/2} E\left[\left|\sum_{s=1}^T \tilde{V}_s^2 p_{st}^2/p_{tt}\right|^{3/2}\right] \\
&\leq p_{tt}\sum_{s=1}^T E[|\tilde{V}_s|^3]p_{st}^2/p_{tt} \leq Cp_{tt}.
\end{aligned}$$

Combining the last two equations gives $E[|\tilde{V}'Z(Z'Z)^{-1}Z_t|^3] \leq Cp_{tt}$. The conclusion then follows by $\sum_{t=1}^T p_{tt} = K$ and summing up. The conclusion for the vector $\tilde{V}_t$ case follows by T.

Lemma A17. If Assumptions 1–5 and 7 are satisfied, then $\sum_{t=1}^T E[|\mu_T^c c'S_T^{-1\prime}H_T^{-1}W_{tT}|^3] \leq C/\mu_T$ in case (I).

Proof. By T, CS, and $F_T = \mu_T^c c'S_T^{-1\prime}$ and H_T bounded,

$$\begin{aligned}
&\sum_{t=1}^T E[|F_T H_T^{-1}W_{tT}|^3] \\
&\quad \leq C\sum_{t=1}^T E\big(\|z_t\|^3/T^{3/2} + \|\tilde{V}'Z(Z'Z)^{-}Z_t\|^3/\mu_T^3 \\
&\qquad + \tau_T^3\|\tilde{V}_t\|^3/\mu_T^3\big) \\
&\quad \leq C\sum_{t=1}^T \|z_t\|^3/T^{3/2} + K/\mu_T^3 + \tau_T^3 T/\mu_T^3 \\
&\quad \leq C(1/\mu_T + K/\mu_T^3).
\end{aligned}$$

In case (I) we have K/μ_T^2 bounded, giving the conclusion.

Lemma A18. If Assumptions 1–5, 7, and 8 are satisfied and $b_T > 0$ are constants such that b_T is bounded and bounded away from zero, then in case (I),

$$\left|\Pr\left(\frac{\mu_T^c c'S_T^{-1\prime}H_T^{-1}W'u}{\sqrt{b_T}} \leq q\right) - \Phi(q)\right| \leq C/\mu_T + C|b_T - \mu_T^{c2}c'S_T^{-1\prime}\Lambda_T S_T^{-1}c|.$$

Proof. Let $F_T = \mu_T^c c'S_T^{-1\prime}$ as previously. Assumption 7 implies

$$E[u_t^2\tilde{V}_t] = E\big[E[u_t^2|\tilde{V}_t]\tilde{V}_t\big] = E[\sigma_u^2\tilde{V}_t] = 0,$$
$$E[(u_t^2 - \sigma_u^2)\tilde{V}_t\tilde{V}_t'] = E\big[E[u_t^2 - \sigma_u^2|\tilde{V}_t]\tilde{V}_t\tilde{V}_t'\big] = 0.$$

Then $\Sigma_T = \sigma_u^2(1 - \tau_T)(H_T + KS_T^{-1}\tilde{\Omega}S_T^{-1\prime})$. Without changing notation let $W = WH_T^{-1}F_T'$, $\Lambda_T = F_T\Lambda_T F_T'$, and

$$\begin{aligned}
\bar{\Lambda}_T &= \sigma_u^2 W'W \\
&= \sigma_u^2 F_T H_T^{-1}\{(1 - \tau_T)H_T + S_T^{-1}\tilde{V}'(P - \tau_T I)^2\tilde{V}S_T^{-1\prime} \\
&\qquad + \hat{J} + \hat{J}'\}H_T^{-1}F_T', \\
\hat{J} &= (1 - \tau_T)S_T^{-1}\Upsilon'(P - \tau_T I)\tilde{V}S_T^{-1\prime}.
\end{aligned}$$

Note that

$$\begin{aligned}
E[\|\hat{J}\|^2] &\leq CE[\|z'P\tilde{V}\|^2/T]/\mu_T^2 + CE[\|z'\tilde{V}\|^2/T]/\mu_T^2 \\
&= C\,\mathrm{tr}(z'Pz/T)/\mu_T^2 + C\,\mathrm{tr}(z'z/T)/\mu_T^2 \\
&= O(1/\mu_T^2).
\end{aligned}$$

Also by Lemma A1 and M we have

$$E[\|\tilde{V}'P\tilde{V} - K\tilde{\Omega}\|^2] = O(K), \qquad E[\|\tilde{V}'\tilde{V} - T\tilde{\Omega}\|^2] = O(T).$$

Then by T and by $(1 - \tau_T)K\tilde{\Omega} = (1 - 2\tau_T)K\tilde{\Omega} + \tau_T^2 T\tilde{\Omega}$,

$$\begin{aligned}
&E\big[\|S_T^{-1}(\tilde{V}'(P - \tau_T I)^2\tilde{V} - (1 - \tau_T)K\tilde{\Omega})S_T^{-1\prime}\|^2\big] \\
&\quad \leq C(1 - 2\tau_T)E[\|\tilde{V}'P\tilde{V} - K\tilde{\Omega}\|^2]/\mu_T^4 \\
&\qquad + C\tau_T^2 E[\|\tilde{V}'\tilde{V} - T\Omega\|^2]/\mu_T^4 \\
&\quad \leq CK/\mu_T^4 \leq C/\mu_T^2.
\end{aligned}$$

Then by T we have $E[|\bar{\Lambda}_T - \Lambda_T|^2] \leq C/\mu_T^2$ whereas by Assumption 7 there is $\varepsilon > 0$ such that $\Lambda_T \geq \varepsilon$ for all T large enough. Then for $\mathcal{A}_T = \{\bar{\Lambda}_T > \varepsilon/2\}$, by Chebyshev's inequality,

$$\begin{aligned}
\Pr(\mathcal{A}_T^c) &\leq \Pr(|\bar{\Lambda}_T - \Lambda_T| > \varepsilon/2) \\
&\leq CE[|\bar{\Lambda}_T - \Lambda_T|^2] \leq C/\mu_T^2.
\end{aligned}$$

Note that $\mathrm{Var}(W'u|\tilde{V}) = \bar{\Lambda}_T$ and $\Pr(W'u/\sqrt{b_T} \leq q|\tilde{V}) = \Pr(W'u/\sqrt{\bar{\Lambda}_T} \leq q\sqrt{b_T/\bar{\Lambda}_T}|\tilde{V})$. Also, by independent observations $u_1, \ldots, u_T$ are independent conditional on $\tilde{V}$ and have conditional mean zero and bounded conditional third moment. Then by a standard approximation result,

$$1(\mathcal{A}_T)\big|\Pr(W'u/\sqrt{b_T}\le q|\tilde{V})-\Phi(q\sqrt{b_T/\bar{\Lambda}_T})\big|$$
$$\le 1(\mathcal{A}_T)C\sum_{t=1}^{T}|W_{tT}|^3/\bar{\Lambda}_T^{3/2}\le C\sum_{t=1}^{T}|W_{tT}|^3.$$

By an expansion of the Gaussian distribution,

$$1(\mathcal{A}_T)\big|\Phi(q\sqrt{b_T/\bar{\Lambda}_T})-\Phi(q)\big|\le C|\bar{\Lambda}_T-b_T|.$$

It then follows by Lemma A17, T, and CS that

$$\begin{aligned}
&|\Pr(W'u/\sqrt{b_T}\le q)-\Phi(q)|\\
&=\big|E[\Pr(W'u/\sqrt{b_T}\le q|\tilde{V})-\Phi(q)]\big|\\
&\le E\big[|\Pr(W'u/\sqrt{b_T}\le q|\tilde{V})-\Phi(q)|\big]\\
&\le E\big[\{1(\mathcal{A}_T^c)+1(\mathcal{A}_T)\}|\Pr(W'u/\sqrt{b_T}\le q|\tilde{V})-\Phi(q)|\big]\\
&\le C/\mu_T^2+CE\left[\sum_{t=1}^{T}|W_{tT}|^3\right]+CE[|\bar{\Lambda}_T-\Lambda_T|]\\
&\le C/\mu_T+C\big\{E[|\bar{\Lambda}_T-\Lambda_T|^2]\big\}^{1/2}+C|\Lambda_T-b_T|\\
&\le C/\mu_T+C|\Lambda_T-b_T|.
\end{aligned}$$

Proof of Theorem 4

For the first conclusion apply Lemma A18 with $V_T=\Lambda_T$, using the notation from Lemma A18. For the second conclusion, do the same with $b_T=\sigma_u^2F_TH_T^{-1}F_T'$, so that by the conclusion of Lemma A18 and by $\tau_T\le K/\mu_T^2$,

$$\begin{aligned}
\left|\Pr\left(\frac{W'u}{\sqrt{b_T}}\le q\right)-\Phi(q)\right|&\le C/\mu_T+|\sigma_u^2F_TH_T^{-1}F_T'-\Lambda_T|\\
&\le C/\mu_T+C\|H_T-\Sigma_T\|\\
&\le C(\mu_T^{-1}+K/\mu_T^2).
\end{aligned}$$

Proof of Theorem 5

$S_T=\operatorname{diag}(I_{G_1},\sqrt{T}I_{G_2})$. By Lemma A11, when $\hat{\delta}=\delta_0$ we have $\sum_{t=1}^{T}(\hat{u}_t^2-\hat{\sigma}_T^2)\hat{V}_t\hat{V}_t'/T=O_p(1)$. Also, as in McFadden (1982), $Z'Z/T$ converging implies that $\max_{t\le T}p_{tt}\longrightarrow 0$, so that

$$\sum_t p_{tt}^2\le\max_{t\le T}p_{tt}\sum_t p_{tt}=K\max_{t\le T}p_{tt}\longrightarrow 0.$$

Therefore, it follows that

$$|S_T^{-1}\hat{B}(\delta_0)S_T^{-1\prime}|\le\left(\sum_t p_{tt}^2+K^2/T\right)O_p(1)\xrightarrow{p}0.$$

Note that by standard calculations, $E[V'PV]\le CK$, so that $V'PV=O_p(1)$ by M. Then by T we have

$$\operatorname{tr}(S_T^{-1}\Upsilon'\Upsilon S_T^{-1\prime})\le C\operatorname{tr}(z'Pz/T)+C\operatorname{tr}(S_T^{-1}V'PVS_T^{-1})=O_p(1).$$

We also have, by $\sum_t p_{tt}=K$,

$$\sum_t(p_{tt}-K/T)^2\le\sum_t p_{tt}^2\longrightarrow 0.$$

Also, it follows by the proof of Lemma A10 that $\sum_t u_t^2\hat{V}_t\hat{V}_t'/T=O_p(1)$, so that

$$\begin{aligned}
S_T^{-1}\hat{A}(\delta_0)S_T^{-1\prime}&\le\left[\sum_t(p_{tt}-K/T)^2\sum_t\|S_T^{-1}\hat{\Upsilon}_t\|^2\right]^{1/2}O_p(1)\\
&\xrightarrow{p}0.
\end{aligned}$$

Let $\hat{X}=\tilde{X}(\delta_0)=X-u\hat{\gamma}'$, $\hat{\gamma}=X'u/u'u$, so that $\hat{X}-\tilde{X}=-u(\hat{\gamma}-\gamma)'$. Then we have, by $u'Pu=O_p(1)$,

$$\begin{aligned}
&S_T^{-1}(\hat{X}-\tilde{X})'P(\hat{X}-\tilde{X})S_T^{-1\prime}\\
&\qquad=(u'Pu)S_T^{-1}(\hat{\gamma}-\gamma)(\hat{\gamma}-\gamma)'S_T^{-1\prime}\xrightarrow{p}0.
\end{aligned}$$

We also have by the Lindberg–Feller central limit theorem,

$$\begin{aligned}
&[Z'\tilde{X}S_T^{-1\prime}/\sqrt{T},Z'u/\sqrt{T}]\\
&\qquad=[Z'z/T+Z'\tilde{V}S_T^{-1\prime}/\sqrt{T},Z'u/\sqrt{T}]\xrightarrow{d}[\tilde{G},\tilde{Y}],
\end{aligned}$$

where $\operatorname{vec}(\tilde{G})$ and $\tilde{Y}$ are Gaussian, independent by $\tilde{V}_t$ and u_t uncorrelated, and $\operatorname{Var}(\tilde{Y})=\sigma_u^2M$. Then by CMT and Slutzky,

$$\begin{aligned}
&S_T^{-1}\tilde{X}'P\tilde{X}S_T^{-1\prime}\\
&\qquad=(S_T^{-1}\tilde{X}'Z/\sqrt{T})(Z'Z/T)^{-1}Z'\tilde{X}S_T^{-1\prime}/\sqrt{T}\xrightarrow{d}\tilde{G}'M^{-1}\tilde{G}.
\end{aligned}$$

It follows that $S_T^{-1}\tilde{X}'P\tilde{X}S_T^{-1}=O_p(1)$, so that $S_T^{-1}\hat{X}'P\hat{X}S_T^{-1}=S_T^{-1}\tilde{X}'P\tilde{X}S_T^{-1}+o_p(1)$. It follows similarly that

$$S_T^{-1}\hat{X}'Pu=S_T^{-1}\tilde{X}'Pu+o_p(1)\xrightarrow{d}\tilde{G}'M^{-1}\tilde{Y},$$

where this convergence is joint with that of $S_T^{-1}\hat{X}'P\hat{X}S_T^{-1}$. Note that by independence of $\tilde{G}$ and $\tilde{Y}$, the conditional variance of $\tilde{G}'M^{-1}\tilde{Y}$ given $\tilde{G}$ is $\sigma_u^2\tilde{G}'M^{-1}\tilde{G}$. Also, $\tilde{G}'M^{-1}\tilde{G}$ is nonsingular with probability 1. Hence, the conditional distribution of $\tilde{Y}'M^{-1}\tilde{G}(\tilde{G}'M^{-1}\tilde{G})^{-1}\tilde{G}'M^{-1}\tilde{Y}/\sigma_u^2$ is $\chi^2(G)$. Because this distribution does not depend on $\tilde{G}$ it follows that this is also the unconditional distribution. Note also that $u'Pu=O_p(1)$ by K fixed, so $\tilde{\alpha}(\delta_0)=O_p(1/T)$. Also, $\tilde{X}(\delta_0)'\tilde{X}(\delta_0)=O_p(T)$ by standard arguments, so that by $\hat{\sigma}_u^2\longrightarrow_p\sigma_u^2$,

$$\begin{aligned}
\hat{\Sigma}_B(\delta_0)&=\{\sigma_u^2+o_p(1)\}\{1+O_p(1/T)\}\tilde{X}(\delta_0)'P\tilde{X}(\delta_0)\\
&\quad+O_p(1/T^2)\tilde{X}(\delta_0)'\tilde{X}(\delta_0)\\
&=\sigma_u^2\tilde{X}(\delta_0)'P\tilde{X}(\delta_0)+o_p(1)\xrightarrow{d}\tilde{G}'M^{-1}\tilde{G}.
\end{aligned}$$

Then by the CMT,

$$\widehat{LM}(\delta_0)\xrightarrow{d}\tilde{Y}'M^{-1}\tilde{G}(\tilde{G}'M^{-1}\tilde{G})^{-1}\tilde{G}'M^{-1}\tilde{Y}/\sigma_u^2\overset{d}{\sim}\chi^2(G).$$

[Received August 2006. Revised March 2007.]

REFERENCES

Anderson, T. W., Kunitomo, N., and Matsushita, Y. (2005), "A New Light From Old Wisdoms: Alternative Estimation Methods of Simultaneous Equations and Microeconometric Models," working paper, University of Tokyo.

Andrews, D. W. K., and Stock, J. H. (2006), "Inference With Weak Instruments," in *Advances in Economics and Econometrics*, eds. R. Blundell, W. K. Newey, and T. Persson, Cambridge, U.K.: Cambridge University Press.

Angrist, J., and Krueger, A. (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*, 63, 657–681.

Bekker, P. A., and van der Ploeg, J. (2005), "Instrumental Variable Estimation Based on Grouped Data," *Statistica Neerlandica*, 59, 239–267.

Bound, J., Jaeger, D., and Baker, R. (1996), "Problems With Instrumental Variables Estimation When the Correlation Between Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90, 443–450.

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

Chao, J., and Swanson, N. (2002), "Consistent Estimation With a Large Number of Weak Instruments," working paper, University of Maryland.

——— (2003), "Asymptotic Normality of Single-Equation Estimators for the Case With a Large Number of Weak Instruments," working paper, University of Maryland.

——— (2004), "Estimation and Testing Using Jackknife IV in Heteroskedastic Regressions With Many Weak Instruments," working paper, University of Maryland.

——— (2005), "Consistent Estimation With a Large Number of Weak Instruments," *Econometrica*, 73, 1673–1692.

Cruz, L. M., and Moreira, M. J. (2005), "On the Validity of Econometric Techniques With Weak Instruments: Inference on Returns to Education Using Compulsory Schooling Laws," *Journal of Human Resources*, 40, 393–410.

Davidson, R., and MacKinnon, J. G. (2006), "The Case Against Jive" (with discussion and reply), *Journal of Applied Econometrics*, 21, 827–833.

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

Dufour, J. M. (1997), "Some Impossibility Theorems in Econometrics, With Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1388.

Fuller, W. A. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939–954.

Hahn, J., and Hausman, J. (2002), "A New Specification Test for the Validity of Instrumental Variables," *Econometrica*, 70, 163–189.

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

Hahn, J., Hausman, J. A., and Kuersteiner, G. M. (2004), "Estimation With Weak Instruments: Accuracy of Higher-Order Bias and MSE Approximations," *Econometrics Journal*, 7, 272–306.

Hansen, C., Hausman, J., and Newey, W. (2004), "Weak Instruments, Many Instruments, and Microeconometric Practice," working paper, MIT.

Hausman, J., Newey, W., and Woutersen, T. (2005), "Instrumental Variable Estimation With Heteroskedasticity and Many Instruments," working paper, MIT.

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

Kunitomo, N. (1980), "Asymptotic Expansions of Distributions of Estimators in a Linear Functional Relationship and Simultaneous Equations," *Journal of the American Statistical Association*, 75, 693–700.

McFadden, D. (1982), "Asymptotic Properties of Least Squares," lecture notes, UC Berkeley.

Moreira, M. J. (2001), "Tests With Correct Size When Instruments Can Be Arbitrarily Weak," Center for Labor Economics Working Paper 37, UC Berkeley.

——— (2003), "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71, 1027–1048.

Morimune, K. (1983), "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability Is Large Compared With the Sample Size," *Econometrica*, 51, 821–841.

Newey, W. K. (1997), "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

Newey, W. K., and Windjmeier, F. (2005), "GMM With Many Weak Moment Conditions," working paper, MIT.

Rothenberg, T. J. (1984), "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, Vol. 2, eds. Z. Griliches and M. D. Intriligator, New York: Elsevier, pp. 881–935.

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

Stock, J., and Yogo, M. (2005a), "Testing for Weak Instruments," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. W. K. Andrews and J. H. Stock, Cambridge, U.K.: Cambridge University Press, Chap. 5.

——— (2005b), "Asymptotic Distributions of Instrumental Variables Statistics With Many Instruments," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, eds. D. W. K. Andrews and J. H. Stock, Cambridge, U.K.: Cambridge University Press, Chap. 6.