Estimation with Many Instrumental Variables*

Christian Hansen
Graduate School of Business
University of Chicago

Jerry Hausman
Department of Economics
M.I.T.

Whitney Newey
Department of Economics
M.I.T.

December 2004
Revised, August 2006

Abstract

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 (1994) 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 (2002) weak instrument statistic that is robust to many instruments.

JEL Classification: C13, C21, C31
Keywords: Weak Instruments, Many Instruments, Standard Errors, Inference

Corresponding Author: Whitney Newey, E52-262D, MIT, Cambridge, MA 02139, wnewey@mit.edu

*The NSF provided financial support for this paper under Grant No. 0136869. 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.

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 3 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, 3004, 2005) and Stock and Yogo (2004). We also find that there is no penalty for many instruments in the rate of approximation for t-ratios when the CSE are used and an additional condition is satisfied. That is, the rate of approximation 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 give 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 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 (2006) to many weak instruments, who had generalized results from an earlier version of this paper by relaxing a conditional moment restriction. We also combine and generalize results of Chao and Swanson (2002, 2003, 2005) and Stock and Yogo (2004) 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 (2004). Our theoretical results make use of some inequalities in Chao and Swanson (2004).

Hahn and Hausman (2002) give a test for weak instruments and Hahn, Hausman, and Kuersteiner (2004) show that FULL performs well under weak instruments. Recently Andrews and Stock (2006) derive 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 slower than the concentration parameter. There it turns out that Wald tests using the CSE attain the power envelope. We also consider a case where the number of instruments grows as fast as the sample size, which is not covered by Andrews and Stock (2006).

The remainder of the paper 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 the number of right-hand 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, i.e. $\Upsilon = Z\pi$ 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 t^{th} row (observation) for 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 $Z_t = (p_{1K}(w_t), ..., p_{KK}(w_t))'$ 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, e.g. 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 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, i.e. 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 then LIML. Let $\hat{u}(\delta) = y - X\delta$, $\hat{\sigma}_u^2(\delta) = \hat{u}(\delta)'\hat{u}(\delta)/(T-G)$, $\tilde{\alpha}(\delta) = \hat{u}(\delta)'P\hat{u}(\delta)/\hat{u}(\delta)'\hat{u}(\delta)$, $\hat{\Upsilon} = PX$, $\tilde{X}(\delta) = X - \hat{u}(\delta)(\hat{u}(\delta)'X)/\hat{u}(\delta)'\hat{u}(\delta)$, $\hat{V}(\delta) = (I-P)\tilde{X}(\delta)$, $\kappa_T = \sum_{t=1}^T P_{tt}^2/K, \tau_T = K/T$,

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

The asymptotic variance estimator is given by

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

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^2 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.

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 one. 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, i.e. they are not robust to weak instruments. These considerations motivate consideration of 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

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

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 $L\hat{M}(\delta)$ 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 $L\hat{M}(\delta)$ by inverting it. Specifically, for the $1-\alpha$ quantile q of a $\chi^2(G)$ distribution, an asymptotic $1-\alpha$ confidence interval is $\{\delta : L\hat{M}(\delta) \leq q\}$. 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 $L\hat{M}(\delta)$ 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. 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 (2004) test. One could also compare a Wald test based on the CSE with a test based on $L\hat{M}(\delta)$. One could also develop versions of the Stock and Yogo (2004) 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-hand 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-hand side and other exogenous right-hand 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. Since 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 find 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 while the Bekker (1994) standard error is .0200981, and with 180 instruments the CSE .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 a study on the application to help make it empirically relevant. The design had the same sample size as the application and instrument observations fixed at the sample values, e.g. as in Staiger and Stock's (1997) design for dummy variable instruments. The data was generated from a two equation triangular simultaneous equations system with structural equation 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 homoskedastic 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 three excluded instruments and one with 179 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 overidentifying restrictions 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 179 overidentifying restrictions the value was set to $\hat{\mu}_T^2 = 257$.

Table 1. Simulation Results
Males born 1930-1939. 1980 IPUMS
n =329,509, $\beta = .0953$

	Bias/β	RMSE	Size
A. 3 instruments, $\mu_T^2 = 95.6$			
2SLS	-0.0021	0.0217	0.056
LIML	0.0052	0.0222	0.056
CSE			0.054
FULL	0.0010	0.0219	0.057
CSE			0.056
Kleibergen			0.059
B. 180 instruments, $\mu_T^2 = 257$			
2SLS	-0.1440	0.0168	0.318
LIML	-0.0042	0.0168	0.133
CSE			0.049
FULL	-0.0063	0.0168	0.132
CSE			0.049
Kleibergen			0.051

Table 1 reports the results of this experiment, giving relative bias, mean-square error, and rejection frequencies for nominal five percent 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 this 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. 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 179 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 provides 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 percent 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 $\rho = 0$, $\rho = 0.2$, $\rho = 0.5$, and $\rho = 0.8$ 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 differences in dispersion are 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 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, while 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 (2004). 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 quantities 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 percent of the papers had at least one overidentifying restriction, 25 percent had at least three, and 10 percent had 7 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 7 summarizes estimates of μ^2 and ρ from these studies.

TABLE 7. 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

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 is not 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.

Some regularity conditions are important for the results. Let Z_t', u_t, V_t', and Υ_t' denote the t^{th} 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, $rank(Z) = K$, $\sum_{t=1}^{T}(1-p_{tt})^2/T \geq C > 0$.

The restriction that $rank(Z) = K$ 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 one for $K < T$.[1] 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 $S_T = \tilde{S}_T \operatorname{diag}(\mu_{1T}, ..., \mu_{GT})$ 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, $\sum_{t=1}^{T} \|z_t\|^4/T^2 \longrightarrow 0$, and $\sum_{t=1}^{T} z_t z_t'/T$ is uniformly nonsingular.

Allowing for K to grow and for μ_T to grow slower than $\sqrt{T}$ models having many instruments without strong identification. Assumption 2 will imply that, when K grows no faster that 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: $(u_1, V_1), ..., (u_T, V_T)$ are independent with $E[u_t] = 0$, $E[V_t] = 0$, $E[u_t^8]$ and $E[\|V_t\|^8]$ are bounded in t, $Var((u_t, V_t')') = diag(\Omega^*, 0)$, Ω^* is nonsingular, and for all $j \in \{1, ..., G\}$ such that $V_{tj} = 0$ and the corresponding submatrix $\tilde{S}_{T22}$ of $\tilde{S}_T$ it is the case that $\mu_{jT} = \sqrt{T}$ and $\tilde{S}_{T22}$ is uniformly nonsingular.

This hypothesis includes moment existence and homoskedasticity assumptions. The consistency of the CSE depends on homoskedasticity, 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 (2006).

[1]The observations $w_1, ..., w_T$ are distinct with probability one and therefore, by $K < T$, cannot all be roots of a K^{th} 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 $a'Z'Za > 0$.

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

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

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 $p^K(w) \stackrel{def}{=} (p_{1K}(w), ..., p_{K-G_2,K}(w))'$. The other 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

$$z_t = \begin{pmatrix} f_0(w_t) \\ Z_{1t} \end{pmatrix}, S_T = \tilde{S}_T diag\left(\mu_T, ..., \mu_T, \sqrt{T}, ..., \sqrt{T}\right), \tilde{S}_T = \begin{pmatrix} I & \pi_{11} \\ 0 & I \end{pmatrix}.$$

By construction we have $\Upsilon_t = S_T z_t / T$. 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 satsified by construction. Turning to Assumption 3, we require that $Var(u_t, V_{1t}')$ is nonsingular. Since 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* $\mu_T^{-1} S_T'(\hat{\delta} - \delta_0) \xrightarrow{p} 0$ *and* $\hat{\delta} \xrightarrow{p} \delta_0$.

This result is more general than Chao and Swanson (2005) in allowing for strongly identified covariates but is similar to 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 set up 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 t^{th} 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: $H = \lim_{T\longrightarrow\infty} (1-\tau_T) z'z/T$, $\tau = \lim_{T\longrightarrow\infty} \tau_T$, $\kappa = \lim_{T\longrightarrow\infty} \kappa_T$, $A = E[u_t^2 \tilde{V}_t] \lim_{T\longrightarrow\infty} \sum_{t=1}^{T} z_t'(p_{tt} - \frac{K}{T})/\sqrt{KT}$ *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

$$S_T'(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \Lambda_I), S_T'\hat{\Lambda} S_T \xrightarrow{p} \Lambda_I, \Lambda_I = H^{-1}\Sigma_I H^{-1}, \tag{5.1}$$

$$\Sigma_I = (1-\tau)\sigma_u^2\{H + S_0\tilde{\Omega}S_0'\} + (1-\tau)(S_0 A + A' S_0') + S_0 B S_0'.$$

In case II we will have

$$(\mu_T/\sqrt{K}) S_T'(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, \Lambda_{II}), (\mu_T^2/K) S_T'\hat{\Lambda} S_T \xrightarrow{p} \Lambda_{II}, \Lambda_{II} = H^{-1}\Sigma_{II} H^{-1}, \tag{5.2}$$

$$\Sigma_{II} = \bar{S}_0[(1-\tau)\sigma_u^2\tilde{\Omega} + B]\bar{S}_0'.$$

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). When K and μ_T^2 grow as fast as T the variance formula generalizes that of Anderson et. al. (2006) to include the coefficients of included exogenous variables, which had previously generalized Hansen et. al. (2004) to allow for $E[u_t|\tilde{V}_t] \neq 0$ and $E[u_t^2|\tilde{V}_t] \neq \sigma_u^2$. This formula also extends that of Bekker and van der Ploeg (1995) to general instruments. The formula also generalizes Anderson et. al. (2006) 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 Stock and Yogo (1994) 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) equation (5.1) is satisfied and in case II) equation (5.2) is satisfied. Also, in each case if Σ is nonsingular then $L\hat{M}(\delta_0) \xrightarrow{d} \chi^2(G)$.*

Recently Andrews and Stock (2006) have derived the power envelope for a test of $H_0 : \delta_0 = \bar{\delta}$ under many weak instruments with Gaussian disturbances and scalar δ. Under the restriction $\sqrt{K}/\mu_T^2 \longrightarrow 0$ that we impose the Wald test with the CSE is optimal, attaining this power envelope. This result follows from optimality of the LM statistic of Kleibergen (2002), as shown by Andrews and Stock (2006), and asymptotic equivalence of the Wald and LM statistic under local alternatives. For brevity we omit the demonstration of asymptotic equivalence of the Wald and LM statistics.

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

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

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) $(\mu_T^c)^2 c'S_T^{-1\prime}\Lambda_T S_T^{-1}c\mu_T^2/K$ 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),*

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

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 while 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 are 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*

$$\begin{aligned} \Pr(\frac{c' S_T^{-1\prime} H_T^{-1} W' u}{\sqrt{c' S_T^{-1\prime} \Lambda_T S_T^{-1} c}} &\leq q) = \Phi(q) + O(1/\mu_T), \\ \Pr(\frac{c' S_T^{-1\prime} H_T^{-1} W' u}{\sqrt{\sigma_u^2 c' S_T^{-1\prime} H_T^{-1} S_T^{-1} c}} &\leq q) = \Phi(q) + O(1/\mu_T + K/\mu_T^2). \end{aligned}$$

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 slower 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 seem not 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 paper.

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* $L\hat{M}(\delta_0) \xrightarrow{d} \chi^2(G)$.

6 Conclusion

In this paper, 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 provides 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.

7 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-Schwartz inequality, and the Triangle inequality respectively. Also, for notational convenience, we drop the T subscript on μ_T throughout.

LEMMA A1: If (u_i, v_i, z_i) are independent with $E[u_i|z_i] = E[v_i|z_i] = 0$, $E[u_i^4|z_i] \leq C$, $E[v_i^4|z_i] \leq C$, z_i is $K \times 1$, then for $Z = [z_1, ..., z_T]$ and $P = Z(Z'Z)^- Z'$,

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

Proof: Let $\sigma_{uvi} = E[u_i v_i|z_i]$, $\mu_{ui}^j = E[(u_i)^j|z_i]$, $\mu_{vi}^j = E[(v_i)^j|z_i]$. By independent observations, $E[uv'|Z] = diag(\sigma_{uv1}, ..., \sigma_{uvT}) = \Gamma$. Then $E[u'Pv|Z] = tr(PE[vu'|Z]) = tr(P_Z\Gamma)$. Also, for $p_{ij} = P_{ij}$,

$$\begin{aligned}
& E[(u'Pv)^2|Z] \qquad (7.3)\\
= & \sum_{i,j,k,\ell=1}^{T} p_{ij}p_{k\ell}E[u_i v_j u_k v_\ell|Z] = \sum_{i=1}^{T} p_{ii}^2 E[u_i^2 v_i^2|z_i] + \sum_{i\neq j=1}^{T} \{(p_{ii}p_{jj} + p_{ij}^2)\sigma_{uvi}\sigma_{uvj} + p_{ij}^2 \mu_{ui}^2 \mu_{vj}^2\} \\
= & \sum_{i=1}^{T} p_{ii}^2 \{E[u_i^2 v_i^2|z_i] - 2\sigma_{uvi}^2 - \mu_{ui}^2\mu_{vi}^2\} + tr(P_Z\Gamma)^2 + \sum_{i,j=1}^{T} p_{ij}^2(\sigma_{uvi}\sigma_{uvj} + \mu_{ui}^2\mu_{vj}^2) \\
\leq & C\sum_{i=1}^{T} p_{ii}^2 + C\sum_{i,j=1}^{T} p_{ij}^2 + tr(P\Gamma)^2 \leq 2C\sum_{i,j=1}^{T} p_{ij}^2 + tr(P_Z\Gamma)^2
\end{aligned}$$

We have $\sum_{j=1}^T p_{ij}^2 = p_{ii}$, so that by equation (7.3),

$$E[(u'Pv - E[u'Pv|Z])^2|Z] \leq C \sum_{i,j=1}^{T} p_{ij}^2 \leq CK.$$

The second conclusion follows by M. Q.E.D.

LEMMA A2: *If i) P is a constant idempotent matrix with rank$(P) = K$; ii) (W_{1T}, V_1, u_1), ..., (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) $K \rightarrow \infty$;then for $\bar{\Sigma} \overset{\text{def}}{=} E[V_tV_t']E[u_t^2] + E[V_tu_t]E[u_tV_t']$, $\kappa_T = \sum_{t=1}^T p_{tt}^2/K$, 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*

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

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 $E[\|W_{1T}\|^4] \leq \sum_{t=1}^T E[\|W_{tT}\|^4] \longrightarrow 0$, 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 $D_T = \sum_{t=1}^T E[W_{tT}W_{tT}']$,

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

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

$$\sum_{t=2}^{T} E[y_{tT}^4] \leq C \sum_{t=2}^{T} E[\|W_{tT}\|^4] + C \sum_{t=2}^{T} E\left[\left(\sum_{j<t}\{v_t p_{tj} u_j + v_j p_{tj} u_t\}\right)^4\right]/K^2$$

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

$$\begin{aligned}
&\sum_{t=2}^{T} E\left[\left(\sum_{j<t} v_t p_{tj} u_j\right)^4\right]/K^2 = \frac{1}{K^2}\sum_{t=2}^{T} \sum_{j,k,\ell,m<t} p_{tj}p_{tk}p_{t\ell}p_{tm} E[v_t^4 u_j u_k u_\ell u_m] \\
= \; & \frac{1}{K^2}\sum_{t=2}^{T} \sum_{j,k,\ell,m<t} E[v_t^4] p_{tj}p_{tk}p_{t\ell}p_{tm} E[u_j u_k u_\ell u_m] \leq \frac{C}{K^2}\sum_{t=2}^{T}\left(\sum_{j<t} p_{tj}^4 + \sum_{j,k<t} p_{tj}^2 p_{tk}^2\right) \\
\leq \; & \frac{C}{K^2}\left(\sum_{t=1}^{T}\sum_{j=1}^{T} p_{tj}^2 + \sum_{t=1}^{T}\left(\sum_{j=1}^{T} p_{tj}^2\right)\left(\sum_{k=1}^{T} p_{tk}^2\right)\right) = \frac{C}{K^2}\left(\sum_{t=1}^{T} p_{tt} + \sum_{t=1}^{T} p_{tt}^2\right) \leq \frac{C}{K} \longrightarrow 0.
\end{aligned}$$

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 $Z_t = (W_{tT}, V_t, u_t)$,

$$\sum_{t=2}^{T} E[y_{tT}^2 \mid Z_1, ..., Z_{t-1}] - s_T^2 \xrightarrow{p} 0 \tag{7.4}$$

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

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

Also

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

and

$$\begin{aligned}
&E\left[w_{tT}\sum_{j<t}(v_t p_{tj} u_j + v_j p_{tj} u_t)/\sqrt{K} \mid Z_1, ..., Z_{t-1}\right] \\
= \; & E[w_{tT} v_t]\sum_{j<t} p_{tj} u_j/\sqrt{K} + E[w_{tT} u_t]\sum_{j<t} p_{tj} v_j/\sqrt{K}.
\end{aligned}$$

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 $\bar{P}_{tj} = 0$, $j \leq t$, and let $\delta = (\delta_1, ..., \delta_T)$. Then $\sum_{t=2}^{T}\sum_{j<t} \delta_t p_{tj} u_j/\sqrt{K} = \delta' \bar{P}' u/\sqrt{K}$. By CS $\delta'\delta = \sum_{t=1}^{T}\left(E\left[w_{tT} v_t\right]\right)^2 \leq$

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

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

so that $\delta'\bar{P}'u/\sqrt{K} \xrightarrow{p} 0$ 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

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

To finish showing that eq. (7.4) is satisfied it only remains to show that for $\bar{y}_{tT} = \sum_{j<t}(v_t p_{tj}u_j + v_j p_{tj}u_t)/\sqrt{K}$,

$$\sum_{t=2}^{T} E\left[\bar{y}_{tT}^2 \mid Z_1, ..., Z_{t-1}\right] - E[\bar{y}_{tT}^2] \xrightarrow{p} 0. \tag{7.5}$$

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

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

Consider the last two terms. Note that

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

Also, we have

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

Note that

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

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

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

It follows similarly that $E\left[\left(\sum_t \sum_{j<k<t} p_{tj} p_{tk} u_k v_j\right)^2\right]/K^2 \longrightarrow 0$.Similar arguments can also be applied to show that each of the other four terms following the equality in eq. (7.6) converges in probability to zero It then follows by T and M that eq. (7.6) is satisfied. By T it then follows that eq. (7.4) is satisfied. Thus all the conditions of the Martingale central limit theorem are satisfied, so that $\sum_{t=2}^T y_{tT} \xrightarrow{d} N(0,1)$. Then by Slutzky theorem the conclusion holds. Q.E.D.

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

LEMMA A3: *If Assumptions 1-4 are satisfied then* $S_T'(\hat{\delta}_{LIML} - \delta_0)/\mu_T \xrightarrow{p} 0$.

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 $\bar{S}_T = diag(0, S_T)$ and $\bar{S}_T^- = diag(0, S_T^{-1})$ where 0 is a scalar, and $\hat{B} = \bar{X}'\bar{X}/T$. Note that $\left\| S_T/\sqrt{T} \right\| \leq C$, so that

$$E[\left\| \bar{\Upsilon}'\bar{V} \right\|^2 / T^2] = tr(S_T z' z S_T')/T^3 \longrightarrow 0,$$

so that $\bar{\Upsilon}'\bar{V}/T \xrightarrow{p} 0$ by M. Also by M,

$$\bar{V}'\bar{V}/T \xrightarrow{p} \bar{\Omega} = E[\bar{V}_t \bar{V}_t'] = diag(\Omega^*, 0) \geq C diag(I_{G-G_2+1}, 0),$$

where G_2 is the number of j with $V_{tj} = 0$. By uniform nonsingularity of $z'z/T$ we have for all T large enough,

$$\bar{S}_T^- \bar{\Upsilon}' \bar{\Upsilon} \bar{S}_T^{-\prime} = diag(0, z'z/T) \geq C diag(0, I_G).$$

Also, by $\mu_{jT} = \sqrt{T}$ for j where $V_{jt} = 0$ we have, for all T large enough,

$$\begin{aligned}\bar{\Upsilon}'\bar{\Upsilon}/T &= \bar{S}_T\bar{S}_T^{-}\bar{\Upsilon}'\bar{\Upsilon}\bar{S}_T^{-\prime}\bar{S}_T'/T \geq C\bar{S}_T diag(0, I_G)\bar{S}_T'/T \\ &\geq C diag(0, \tilde{S}_T) diag(0, I_{G_2}) diag(0, \tilde{S}_T').\end{aligned}$$

Therefore, by D nonsingular and hence $D'D$ positive definite, w.p.a.1 we have

$$\hat{B} \geq C\{diag(0, \tilde{S}_T)diag(0, I_{G_2})diag(0, \tilde{S}_T') + diag(I_{G-G_2+1}, 0)\}.$$

It follows by straightforward arguments from uniform nonsingularity of L_{T22} that the matrix in brackets is unformly nonsingular, so that $\min_{\|\alpha\|=1} \alpha'\hat{B}\alpha \geq C$ w.p.a.1. Also, by similar arguments $\hat{B} = O_p(1)$.

Next, note that $\bar{S}_T^{-\prime}\bar{S}_T^{-} \leq CI/\mu_T^2$, so that

$$E[\left\|\bar{S}_T^{-}\bar{\Upsilon}'\bar{V}\bar{S}_T^{-\prime}\right\|^2] \leq C\,\mathrm{tr}(\bar{S}_T^{-}\bar{\Upsilon}'\bar{\Upsilon}\bar{S}_T^{-\prime})/\mu_T^2 \longrightarrow 0.$$

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

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

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

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

Let $\hat{A} = \mu_T^{-2}(\bar{X}'P\bar{X} - (K/T)\bar{X}'\bar{X})$. By Assumption 2 $\bar{S}_T^{-}\bar{\Upsilon}'\bar{\Upsilon}\bar{S}_T^{-\prime} \geq C\bar{I}$ for all large enough T, where $\bar{I} = diag(0, I_G)$, so that by T w.p.a.1,

$$\begin{aligned}\hat{A} &= \mu_T^{-2}D'\bar{S}_T\bar{S}_T^{-}[\left(1 - \frac{K}{T}\right)\bar{\Upsilon}'\bar{\Upsilon} - \bar{\Upsilon}'(I-P)\bar{\Upsilon} + \bar{\Upsilon}'P\bar{V} + \bar{V}'P\bar{\Upsilon} \\ &\quad - \frac{K}{T}\bar{V}'\bar{\Upsilon} - \frac{K}{T}\bar{\Upsilon}'\bar{V} + \bar{V}'P\bar{V} - \frac{K}{T}\bar{V}'\bar{V}]\bar{S}_T^{-\prime}\bar{S}_T'D \\ &= \mu_T^{-2}D'\bar{S}_T[\left(1 - \frac{K}{T}\right)\bar{S}_T^{-}\bar{\Upsilon}'\bar{\Upsilon}\bar{S}_T^{-\prime} + o_p(1)]\bar{S}_T'D \geq C\mu_T^{-2}D'\bar{S}_T\bar{I}\bar{S}_T'D.\end{aligned}$$

Now partition $\alpha = (\alpha_1, \alpha_2')'$ where α_1 is a scalar. Since $\bar{S}_T = diag(0, S_T)$ we have $\alpha'D'\bar{S}_T\bar{I}\bar{S}_T'D\alpha = [\alpha_2 + \alpha_1\delta_0]' S_T S_T' [\alpha_2 + \alpha_1\delta_0]$. Then from the previous equation and by Q positive definite, w.p.a.1 for all $\|\alpha\| = 1$,

$$\alpha'\hat{A}\alpha \geq C(\alpha_2 + \alpha_1\delta_0)' S_T S_T' (\alpha_2 + \alpha_1\delta_0)/\mu_T^2 = C\left\|S_T'(\alpha_2 + \alpha_1\delta_0)/\mu_T\right\|^2.$$

Now, note that for $c_0 = \|(1, -\delta_0')'\|$ and $\alpha_0 = (1, -\delta_0')'/c_0$ we have $\bar{X}\alpha_0 = u/c_0$, so that

$$\alpha_0'\hat{A}\alpha_0 = (u'Pu - (K/T)u'u)\,/c_0^2\mu_T^2 \xrightarrow{p} 0.$$

For any α with $\|\alpha\| = 1$ let

$$\hat{q}(\alpha) = \frac{T}{\mu^2}\left(\frac{\alpha'\bar{X}'P\bar{X}\alpha}{\alpha'\bar{X}'\bar{X}\alpha} - \frac{K}{T}\right) = \alpha'\hat{A}\alpha/\alpha'\hat{B}\alpha.$$

By $\alpha_0'\hat{B}\alpha_0 \geq \min_{\|\alpha\|=1}\alpha'\hat{B}\alpha/T \geq C$ w.p.a.1. and $\alpha_0'\hat{A}\alpha_0 \xrightarrow{p} 0$ it follows that $\hat{q}(\alpha_0) \leq C^{-1}\alpha_0'\hat{A}\alpha_0 \xrightarrow{p} 0$. Also, for $\hat{\alpha} = \arg\min_{\|\alpha\|=1}\hat{q}(\alpha)$, $\hat{q}(\hat{\alpha}) \leq \hat{q}(\alpha_0)$, so $\hat{q}(\hat{\alpha}) \xrightarrow{p} 0$. Then by $\hat{B} = O_p(1)$ we have $\hat{\alpha}'\hat{A}\hat{\alpha} = \hat{q}(\hat{\alpha})\hat{\alpha}'\hat{B}\hat{\alpha} \xrightarrow{p} 0$, so that

$$\|S_T'(\hat{\alpha}_2 + \hat{\alpha}_1\delta_0)/\mu_T\|^2 \leq C\hat{\alpha}'\hat{A}\hat{\alpha} \xrightarrow{p} 0.$$

Since $S_T S_T'/\mu_T^2 \geq CI$, we have $\|\hat{\alpha}_2 + \hat{\alpha}_1\delta_0\| \xrightarrow{p} 0$. Because α_0 is the unique α with $\|\alpha\| = 1$ satisfying $\|\alpha_2 + \alpha_1\delta_0\| = 0$ it follows by a standard argument that $\hat{\alpha} \xrightarrow{p} (1, -\delta_0')/c_0$. In particular, $\hat{\alpha}_1 \geq C$ w.p.a.1 Then w.p.a.1 $\tilde{\delta} = -\tilde{\alpha}_2/\tilde{\alpha}_1$ exists and

$$\left\|S_T'\left(\tilde{\delta} - \delta_0\right)/\mu_T\right\|^2 = \|S_T'(\hat{\alpha}_2 + \hat{\alpha}_1\delta_0)/\mu_T\|^2/\hat{\alpha}_1^2 \xrightarrow{p} 0.$$

Finally, note that $\hat{q}((1, -\delta')')$ is a monotonic transformation of the LIML objective function $(y - X\delta)'P(y - X\delta)/(y - X\delta)'(y - X\delta)$. Further, since $\hat{\alpha}_1 \neq 0$ w.p.a.1,

$$\min_{\|\alpha\|=1}\hat{q}(\alpha) = \min_{\delta}\hat{q}((1, \delta')')$$

and by invariance to reparameterization, $\tilde{\delta} = \underset{\delta}{\operatorname{argmin}}\ \hat{A}((1, \delta')')$. Q.E.D.

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

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

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*

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

Proof: Note that in Case I, $\sqrt{K}/\mu_T^2 \leq C/\tilde{\mu}_T$ 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,

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

Also, by Lemma A4,

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

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

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

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

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

and $S_T^{-1}\Upsilon'PVS_T^{-1\prime} = O_p(1/\mu_T)$ by CM. Similarly, $S_T^{-1}\Upsilon'VS_T^{-1\prime} = O_p(1/\mu_T)$, 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

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

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,

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

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, $\hat{H}_T = S_T^{-1}\left(X'PX - \hat{\alpha}X'X\right)S_T^{-1\prime} = O_p(1)$. Then the last conclusion follows by T and

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

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 $\varepsilon_T \geq 1/\mu_T$ *then* $\hat{u}'P\hat{u}/\hat{u}'\hat{u} = \breve{\alpha} + O_p(\varepsilon_T^2\mu_T^2/T)$.

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

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

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

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

Also $S_T^{-1}(X'Pu - \hat{\alpha}X'u)/\mu_T \xrightarrow{p} 0$ 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 $S_T'(\hat{\delta}_{LIML}-\delta_0)/\mu_T \xrightarrow{p} 0$ 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,

$$\hat{\alpha}_{FULL} = \hat{\alpha}_{LIML} + O_p(1/T) = \hat{\alpha}_{LIML} + o_p(\mu_T^2/T) = K/T + o_p(\mu_T^2/T). Q.E.D.$$

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 $\bar{\gamma} = X'\bar{u}/\bar{u}'u$. Then differentiating gives

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

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

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

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. Q.E.D.

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*

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

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

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

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

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

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*

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

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

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

It then follows by standard calculations that for $\hat{\gamma} = X'\hat{u}/\hat{u}'\hat{u}$, $\left\|\hat{\gamma}-\gamma\right\|^2 = O_p(\tilde{\mu}_T^{-2})$.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

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

Then it follows that

$$\left\|\hat{V}-\bar{V}\right\|^2/T \le C\left\|u\gamma' - \hat{u}\hat{\gamma}'\right\|^2/T + Ctr[\Upsilon'(I-P)\Upsilon/T].$$

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. Q.E.D.

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

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

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

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 $\left\|\hat{V}-\bar{V}\right\|^2/T = O_p(\Delta_T^2 + \tilde{\mu}_T^{-2})$ by Lemm 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 $d = (d_1, ..., d_T)'$. Note that $\|d\|^2 \le 1$ and $E[\|V'Pd\|^2] \le Cd'd \le C$, so that $V'Pd = O_p(1)$ and $\left\|S_T^{-1}\Upsilon d\right\| \le \left\|z/\sqrt{T}\right\|\|d\| \le C$. 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 $\varepsilon_T = \Delta_T + 1/\mu_T$

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

Lemma A11: *If Assumptions 1-4 are satisfied then*

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

Proof: Let $A=\text{diag}(a_1,...,a_T)$. 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, $\varepsilon'Av/T=\sum_t a_t\varepsilon_t v_t/T=E[a_t\varepsilon_t v_t]+O_p\left(1/\sqrt{T}\right)$. Let $e=(1,...,1)'$ and $\overline{av}=E[a_tv_t]e$. Then

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

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

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

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

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

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

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

$$\begin{aligned}E\left[(\varepsilon'HADv-\varepsilon'HD\overline{av})^2\right]/T^2&=E\left[\left(\sum_{s\neq t}\varepsilon_sp_{st}w_t\right)^2\right]/T^2=\sum_{t\neq s}\sum_{i\neq j}p_{st}p_{ij}E[\varepsilon_sw_t\varepsilon_iw_j]/T^2\\ =\sum_{t\neq s}p_{st}^2\left(E[\varepsilon_s^2]E[w_t^2]+E[\varepsilon_sw_s]E[\varepsilon_tw_t]\right)/T^2&\le C\sum_{s,t}p_{st}^2/T^2=C\sum_t p_{tt}/T^2\le C/T,\end{aligned}$$

so that

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

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

$$E\left[\left(\varepsilon' DADv - \sum_t p_{tt}^2 E[a_t\varepsilon_t v_t]\right)^2 / T^2\right] = \sum_{t=1}^{T} p_{tt}^4 (E[a_t^2\varepsilon_t^2 v_t^2] - E[a_t\varepsilon_t v_t]^2)/T^2 = O(1/T)$$

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

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

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

is an average of a martingale difference. Therefore

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

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

$$\varepsilon' LALv/T = \sum_t a_t \left(\sum_{j<t} p_{jt}\varepsilon_j\right)\left(\sum_{k>t} p_{kt} v_k\right) = \sum_{j<t<k} p_{jt}p_{kt} a_t\varepsilon_j v_k.$$

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 $t = s$, $j = \ell$, $k = m$,

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

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

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

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

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

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 $\left|(y_t - X_t'\tilde{\delta})^2 - (y_t - X_t'\delta)^2\right| \le d_t \left\|\tilde{\delta}-\delta\right\|$ 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, $\sum_t d_t \left\|\bar{V}_t\right\|^2/T = O_p(1)$. In particular, for $\tilde{D} = \text{diag}(d_1,...,d_T)$.

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

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

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

It then follows that

$$\left\|\sum_t \left[\hat{u}_t^2 - \hat{\sigma}_u^2 - (u_t^2-\sigma_u^2)\right]\bar{V}_t\bar{V}_t'/T\right\| \le O_p(1)\left(\left\|\hat{\delta}-\delta_0\right\| + \left\|\hat{\sigma}_u^2 - \sigma_u^2\right\|\right) = O_p\left(1/\tilde{\mu}_T\right)$$

We also have by CS and T,

$$\begin{aligned} \left\|\sum_t \left(\hat{u}_t^2-\hat{\sigma}_u^2\right)(\hat{V}_t\hat{V}_t' - \bar{V}_t\bar{V}_t)/T\right\| &\le \sum_t d_t \left(\left\|\hat{V}_t - \hat{V}_t\right\|^2 + 2\left\|\bar{V}_t\right\|\left\|\hat{V}_t-\bar{V}_t\right\|\right)/T \\ &\le \sum_t d_t\left\|\hat{V}_t-\bar{V}_t\right\|^2/T + 2\left(\sum_t d_t\left\|\bar{V}_t\right\|^2/T\right)^{\frac{1}{2}}\left(\sum_t d_t\left\|\hat{V}_t-\bar{V}_t\right\|^2/T\right)^{\frac{1}{2}} \end{aligned}$$

It follows similarly to previous arguments that

$$\sum_t d_t\left\|\hat{V}_t-\bar{V}_t\right\|^2/T = O_p\left(\Delta_T^2 + \tilde{\mu}_T^{-2}\right).$$

The conclusion then follows by T. Q.E.D.

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 $\varepsilon_T^{\alpha} = \mu_T^2/T\tilde{\mu}_T^2$. Q.E.D.

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

$$S_T^{-1}\tilde{\alpha}(\hat{\delta})\tilde{X}(\hat{\delta})'\tilde{X}(\hat{\delta})S_T^{-1\prime} = \tau_T(1-\tau_T)^{-1}H_T + KS_T^{-1}\tilde{\Omega}S_T^{-1\prime} + O_p(\tilde{\mu}_T^{-1}).$$

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 $\left\|\hat{X} - \tilde{X}\right\| = O_p(\sqrt{T}/\tilde{\mu}_T)$ and $\left\|\tilde{X}\right\| = O_p(\sqrt{T})$, so that $\left\|\hat{X}'\hat{X} - \tilde{X}'\tilde{X}\right\| = O_p(T/\tilde{\mu}_T)$. Therefore we have

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

We also have

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

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, $K\sqrt{T}/T\mu_T^2 = (\sqrt{K}/\mu_T^2)\sqrt{K/T} \leq C/\tilde{\mu}_T$ so that by M

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

It then follows by T that

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

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

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

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 proof of Lemma A5 and A9 that

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

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

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

The conclusion now follows by Lemmas A10 and A11 and T. Q.E.D.

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,

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

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

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

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

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

Proof of Theorem 2: Consider first the case where $\hat{\delta}$ is LIML. Then $\mu_T^{-1}S_T'(\hat{\delta} - \delta_0) \xrightarrow{p} 0$ by Theorem 1, implying $\hat{\delta} \xrightarrow{p} \delta_0$. The first-order conditions for LIML are $\hat{D}(\hat{\delta}) = 0$. Expanding gives

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

where $\bar{\delta}$ lies on the line joining $\hat{\delta}$ and δ_0 and hence $\bar{\beta} = \mu_T^{-1}S_T'(\bar{\delta} - \delta_0) \xrightarrow{p} 0$. 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

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

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,

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

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

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

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 $\Gamma = \text{diag}\left(\bar{\Psi}, \sigma_u^2 \tilde{\Omega}(1-\kappa)\right)$ 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 $c'U_T \xrightarrow{d} N(0, c'\Gamma c)$. Also, if $c'\Gamma c = 0$ then it is straightforward to show that $c'U_T \xrightarrow{p} 0$. Then it follows that

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

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

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

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

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

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

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

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

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

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 $(\mu_T/\sqrt{K})S_T$. Q.E.D.

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,

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

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

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

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

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

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

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

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

For the third conclusion, let $\tilde{t} = c'(\tilde{\delta} - \delta_0) / \left(\hat{\sigma}_u^2 c'\hat{H}c\right)^{1/2}$ and $\hat{\rho} = \left(\hat{\sigma}_u^2 c'\hat{H}c\right)^{1/2} (c'\hat{\Lambda}c)^{-1/2}$, 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

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

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

$$\Pr(\left|\tilde{t}\right| \geq C) \geq \Pr(\left|\hat{t}\right| \geq C\varepsilon, |\hat{\rho}| < \varepsilon) \longrightarrow 1 - \{\Phi(C\varepsilon) - \Phi(-C\varepsilon)\}.$$

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

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

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

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

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

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

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

By $p_{tt} \le 1$ it follows that $p_{tt}^{3/2} \le p_{tt}$. 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

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

Combining the last two equations gives $E\left[\left|\tilde{V}'Z(Z'Z)^{-1}Z_t\right|^3\right] \le Cp_{tt}$. 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. Q.E.D.

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

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

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

In case I we have K/μ_T^2 bounded, giving the conclusion. Q.E.D.

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,*

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

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

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

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

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

Note that

$$\begin{aligned} E[\left\|\hat{J}\right\|^2] &\le CE[\left\|z'P\tilde{V}\right\|^2/T]/\mu_T^2 + CE[\left\|z'\tilde{V}\right\|^2/T]/\mu_T^2. \\ &= Ctr(z'Pz/T)/\mu_T^2 + Ctr(z'z/T)/\mu_T^2 = O(1/\mu_T^2). \end{aligned}$$

Also by Lemma A1 and M we have

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

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

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

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

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

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

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

By an expansion of the Gaussian distribution,

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

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

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

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 $\tau_T\le K/\mu_T^2$,

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

Q.E.D.

Proof of Theorem 5: $S_T=diag(I_{G_1},\sqrt{T}I_{G_2})$. By Lemma A11, when $\hat{\delta}=\delta_0$ we have $\sum_{t=1}^T(\hat{u}_t^2-\hat{\sigma}_n^2)\hat{V}_t\hat{V}_t'/T=O_p(1)$. 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

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

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

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

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

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

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

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

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 $u'Pu = O_p(1)$,

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

We also have by the Lindberg-Feller Central Limit Theorem,

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

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

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

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

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

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 one. 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 $\chi^2(G)$. Since 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 $\hat{\sigma}_u^2 \xrightarrow{p} \sigma_u^2$,

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

Then by the CMT,

$$L\hat{M}(\delta_0) \xrightarrow{d} \tilde{Y}'M^{-1}\tilde{G}(\tilde{G}'M^{-1}\tilde{G})^{-1}\tilde{G}'M^{-1}\tilde{Y}/\sigma_u^2 \overset{d}{\sim} \chi^2(G).Q.E.D.$$

References

Andrews, D.W.K. and J.H. Stock (2006): "Inference with Weak Instruments," in Blundell, R., W. Newey, and T. Persson, eds., *Advances in Economics and Econometrics, Vol. III, Ch. 6,* Cambridge: Cambridge University Press.

Anderson, T.W., N. Kunitomo, and Y. Matsushita (2006): "New Light from Old Wisdoms I: On Asymptotic Properties of LIML Estimator With Possibly Many Instruments," working paper, May 2006.

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

Bekker, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica* 63, 657-681.

Bekker, P.A. and J. van der Ploeg (2005): "Instrumental variable estimation based on grouped data," *Statistica Neerlandica* 59, 239-267.

Bound, J., D. Jaeger, and R. Baker (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.

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

Chao, J. and N. Swanson (2003): "Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments," working paper.

Chao, J. and N. Swanson (2004): "Estimation and Testing Using Jackknife IV in Heteroskedastic Regressions With Many Weak Instruments," working paper.

Chao, J. and N. Swanson (2005): "Consistent Estimation With a Large Number of Weak Instruments," *Econometrica* 73, 1673-1692.

Cruz, L.M. and M.J. Moreira (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 J.G. MacKinnon (2006): "The Case Against Jive," (with discussion and reply), *Journal of Applied Econometrics,* forthcoming.

Donald, S. G. and W. K. Newey (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 J. Hausman (2002) "A New Specification Test for the Validity of Instrumental Variables", *Econometrica 70,* 163-189.

Hahn, J., J.A. Hausman, and G.M. Kuersteiner (2004): "Estimation with Weak Instruments: Accuracy of higher-order bias and MSE approximations," *Econometrics Journal, Volume 7.*

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

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

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.

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

Moriera, M.J. (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 (2005): "GMM with Many Weak Moment Conditions Many Moments," working paper, MIT.

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

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

Stock, J. and M. Yogo (2005): "Testing for Weak Instruments," Chapter 5, in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothen-*

berg, Andrews, D.W.K. and J.H. Stock eds., Cambridge: Cambride University Press.

Stock, J. and M. Yogo (2005): "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," Chapter 6, in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Andrews, D.W.K. and J.H. Stock eds., Cambridge: Cambride University Press.

Table 2. Weak Instrument Limit of LIML and Fuller. $\rho = 0$

ρ	K	μ^2	LIML Median	LIML IQR	LIML Bekker	LIML Standard	Fuller Median	Fuller IQR	Fuller Bekker	Fuller Standard
0.0	1	1	0.001	1.318	0.001	0.001	0.000	0.486	0.001	0.001
0.0	1	2	0.001	1.000	0.001	0.001	0.000	0.502	0.002	0.001
0.0	1	4	0.000	0.708	0.004	0.004	0.000	0.488	0.005	0.003
0.0	1	8	-0.001	0.491	0.010	0.010	-0.001	0.418	0.011	0.009
0.0	1	16	0.000	0.342	0.023	0.023	0.000	0.319	0.024	0.020
0.0	1	32	0.000	0.240	0.035	0.035	0.000	0.232	0.036	0.033
0.0	1	64	0.000	0.169	0.042	0.042	0.000	0.167	0.042	0.041
0.0	2	1	-0.002	1.418	0.001	0.001	-0.001	0.659	0.001	0.001
0.0	2	2	0.000	1.099	0.002	0.002	0.000	0.629	0.002	0.002
0.0	2	4	0.000	0.775	0.004	0.005	0.000	0.560	0.006	0.005
0.0	2	8	0.000	0.525	0.011	0.013	0.000	0.450	0.013	0.012
0.0	2	16	0.000	0.355	0.023	0.026	0.000	0.331	0.024	0.023
0.0	2	32	0.000	0.244	0.036	0.038	0.000	0.236	0.036	0.036
0.0	2	64	0.000	0.171	0.043	0.045	0.000	0.168	0.043	0.043
0.0	4	1	0.002	1.528	0.002	0.003	0.001	0.834	0.003	0.003
0.0	4	2	-0.002	1.227	0.003	0.005	-0.002	0.769	0.004	0.005
0.0	4	4	0.000	0.879	0.006	0.009	0.000	0.656	0.007	0.009
0.0	4	8	-0.001	0.581	0.012	0.018	-0.001	0.500	0.014	0.017
0.0	4	16	0.000	0.377	0.023	0.033	0.000	0.352	0.025	0.030
0.0	4	32	0.000	0.252	0.035	0.044	0.000	0.244	0.036	0.041
0.0	4	64	0.000	0.173	0.043	0.048	0.000	0.171	0.043	0.046
0.0	8	1	-0.001	1.634	0.005	0.012	0.000	1.004	0.006	0.012
0.0	8	2	0.001	1.360	0.006	0.015	0.000	0.917	0.008	0.015
0.0	8	4	0.001	1.011	0.009	0.022	0.001	0.774	0.010	0.021
0.0	8	8	0.001	0.669	0.015	0.034	0.001	0.578	0.017	0.032
0.0	8	16	-0.001	0.419	0.025	0.049	-0.001	0.391	0.026	0.045
0.0	8	32	0.000	0.268	0.036	0.056	0.000	0.260	0.036	0.053
0.0	8	64	0.000	0.179	0.044	0.056	0.000	0.176	0.044	0.054
0.0	16	1	-0.002	1.720	0.010	0.046	-0.002	1.164	0.011	0.047
0.0	16	2	0.000	1.496	0.011	0.051	0.000	1.074	0.013	0.052
0.0	16	4	-0.001	1.170	0.014	0.060	-0.001	0.915	0.015	0.059
0.0	16	8	-0.002	0.793	0.019	0.073	-0.002	0.686	0.020	0.070
0.0	16	16	-0.001	0.486	0.026	0.085	0.000	0.452	0.028	0.080
0.0	16	32	0.000	0.295	0.036	0.082	0.000	0.285	0.036	0.077
0.0	16	64	0.000	0.189	0.043	0.069	0.000	0.186	0.043	0.067
0.0	32	1	0.002	1.795	0.017	0.129	0.001	1.320	0.019	0.134
0.0	32	2	0.001	1.622	0.018	0.134	0.001	1.232	0.020	0.137
0.0	32	4	-0.001	1.334	0.020	0.141	-0.001	1.076	0.022	0.143
0.0	32	8	0.000	0.950	0.023	0.149	0.000	0.826	0.025	0.147
0.0	32	16	0.000	0.590	0.029	0.153	0.000	0.549	0.030	0.149
0.0	32	32	0.000	0.343	0.036	0.133	0.000	0.331	0.037	0.128
0.0	32	64	0.000	0.208	0.043	0.098	0.000	0.204	0.043	0.095

Table 3. Weak Instrument Limit of LIML and Fuller. $\rho = 0.2$

ρ	K	μ^2	LIML Median	LIML IQR	LIML Bekker	LIML Standard	Fuller Median	Fuller IQR	Fuller Bekker	Fuller Standard
0.2	1	1	0.084	1.307	0.002	0.002	0.162	0.484	0.048	0.003
0.2	1	2	0.038	0.989	0.004	0.004	0.119	0.498	0.033	0.005
0.2	1	4	0.010	0.706	0.009	0.009	0.065	0.483	0.021	0.009
0.2	1	8	0.000	0.490	0.016	0.016	0.027	0.414	0.021	0.016
0.2	1	16	-0.001	0.343	0.026	0.026	0.012	0.318	0.029	0.025
0.2	1	32	0.000	0.240	0.036	0.036	0.006	0.232	0.038	0.035
0.2	1	64	0.000	0.169	0.044	0.044	0.003	0.166	0.044	0.042
0.2	2	1	0.096	1.402	0.004	0.005	0.146	0.650	0.009	0.005
0.2	2	2	0.054	1.088	0.006	0.007	0.108	0.619	0.011	0.008
0.2	2	4	0.019	0.771	0.010	0.012	0.062	0.551	0.015	0.013
0.2	2	8	0.003	0.524	0.017	0.020	0.028	0.445	0.022	0.020
0.2	2	16	0.000	0.354	0.027	0.030	0.013	0.328	0.030	0.029
0.2	2	32	0.000	0.244	0.036	0.039	0.006	0.236	0.038	0.038
0.2	2	64	0.000	0.170	0.043	0.044	0.003	0.167	0.043	0.043
0.2	4	1	0.117	1.507	0.008	0.012	0.146	0.821	0.012	0.014
0.2	4	2	0.071	1.209	0.010	0.015	0.108	0.756	0.014	0.017
0.2	4	4	0.030	0.869	0.013	0.021	0.064	0.642	0.018	0.022
0.2	4	8	0.005	0.578	0.019	0.028	0.028	0.492	0.023	0.028
0.2	4	16	0.000	0.376	0.028	0.037	0.013	0.349	0.031	0.036
0.2	4	32	0.000	0.251	0.036	0.044	0.006	0.242	0.038	0.043
0.2	4	64	0.000	0.173	0.043	0.048	0.003	0.170	0.044	0.047
0.2	8	1	0.133	1.609	0.014	0.033	0.151	0.987	0.019	0.037
0.2	8	2	0.091	1.346	0.016	0.037	0.117	0.902	0.021	0.040
0.2	8	4	0.047	1.000	0.019	0.042	0.074	0.758	0.023	0.044
0.2	8	8	0.012	0.661	0.023	0.049	0.034	0.565	0.027	0.049
0.2	8	16	0.002	0.415	0.029	0.054	0.014	0.386	0.032	0.053
0.2	8	32	0.000	0.266	0.037	0.056	0.006	0.257	0.038	0.054
0.2	8	64	0.000	0.178	0.044	0.055	0.003	0.175	0.044	0.054
0.2	16	1	0.149	1.692	0.023	0.082	0.160	1.144	0.028	0.090
0.2	16	2	0.110	1.477	0.024	0.084	0.127	1.057	0.029	0.091
0.2	16	4	0.064	1.154	0.025	0.087	0.085	0.898	0.030	0.092
0.2	16	8	0.022	0.784	0.028	0.089	0.041	0.673	0.031	0.091
0.2	16	16	0.004	0.482	0.031	0.088	0.016	0.446	0.034	0.087
0.2	16	32	0.000	0.293	0.037	0.080	0.006	0.282	0.039	0.078
0.2	16	64	0.000	0.188	0.044	0.068	0.003	0.185	0.044	0.066
0.2	32	1	0.161	1.769	0.032	0.161	0.168	1.295	0.036	0.172
0.2	32	2	0.131	1.594	0.033	0.162	0.142	1.211	0.037	0.171
0.2	32	4	0.087	1.313	0.033	0.163	0.101	1.056	0.037	0.170
0.2	32	8	0.041	0.938	0.034	0.161	0.056	0.812	0.037	0.164
0.2	32	16	0.009	0.583	0.034	0.152	0.020	0.540	0.037	0.150
0.2	32	32	0.001	0.341	0.039	0.129	0.007	0.329	0.040	0.125
0.2	32	64	0.000	0.206	0.043	0.096	0.003	0.202	0.044	0.094

TABLE 4. Weak Instrument Limit of LIML and Fuller. $\rho = 0.5$

			LIML				Fuller			
ρ	K	μ^2	Median	IQR	Bekker	Standard	Median	IQR	Bekker	Standard
0.5	1	1	0.200	1.221	0.024	0.024	0.380	0.470	0.182	0.032
0.5	1	2	0.091	0.952	0.031	0.031	0.268	0.494	0.132	0.039
0.5	1	4	0.024	0.700	0.038	0.038	0.149	0.463	0.085	0.046
0.5	1	8	0.003	0.493	0.042	0.042	0.068	0.395	0.061	0.050
0.5	1	16	-0.001	0.344	0.043	0.043	0.031	0.311	0.053	0.049
0.5	1	32	0.000	0.240	0.042	0.042	0.016	0.229	0.049	0.046
0.5	1	64	0.000	0.170	0.044	0.044	0.008	0.166	0.047	0.046
0.5	2	1	0.239	1.308	0.036	0.042	0.360	0.601	0.096	0.057
0.5	2	2	0.123	1.028	0.040	0.046	0.260	0.569	0.084	0.059
0.5	2	4	0.040	0.754	0.044	0.051	0.150	0.503	0.072	0.062
0.5	2	8	0.005	0.516	0.045	0.051	0.068	0.413	0.061	0.059
0.5	2	16	0.000	0.351	0.043	0.047	0.031	0.318	0.053	0.053
0.5	2	32	0.000	0.243	0.042	0.044	0.015	0.232	0.048	0.048
0.5	2	64	0.000	0.171	0.045	0.046	0.008	0.167	0.048	0.048
0.5	4	1	0.283	1.392	0.055	0.081	0.361	0.745	0.093	0.105
0.5	4	2	0.167	1.134	0.054	0.080	0.267	0.683	0.085	0.100
0.5	4	4	0.064	0.830	0.052	0.076	0.157	0.572	0.075	0.091
0.5	4	8	0.012	0.564	0.050	0.068	0.073	0.453	0.065	0.078
0.5	4	16	0.001	0.369	0.045	0.055	0.032	0.333	0.054	0.062
0.5	4	32	0.000	0.250	0.043	0.049	0.016	0.238	0.049	0.053
0.5	4	64	0.000	0.173	0.044	0.048	0.008	0.168	0.047	0.049
0.5	8	1	0.325	1.478	0.078	0.146	0.375	0.890	0.109	0.179
0.5	8	2	0.218	1.245	0.073	0.137	0.287	0.813	0.099	0.165
0.5	8	4	0.099	0.937	0.065	0.121	0.175	0.673	0.085	0.141
0.5	8	8	0.024	0.632	0.056	0.097	0.081	0.510	0.070	0.110
0.5	8	16	0.001	0.401	0.048	0.071	0.032	0.362	0.057	0.079
0.5	8	32	0.000	0.261	0.043	0.057	0.016	0.248	0.049	0.061
0.5	8	64	0.000	0.176	0.045	0.053	0.008	0.172	0.048	0.055
0.5	16	1	0.367	1.553	0.097	0.223	0.397	1.035	0.120	0.259
0.5	16	2	0.271	1.365	0.091	0.208	0.316	0.954	0.111	0.240
0.5	16	4	0.147	1.076	0.080	0.183	0.204	0.805	0.097	0.208
0.5	16	8	0.048	0.733	0.065	0.145	0.098	0.599	0.078	0.161
0.5	16	16	0.005	0.457	0.052	0.101	0.035	0.411	0.061	0.111
0.5	16	32	0.000	0.282	0.045	0.074	0.015	0.269	0.050	0.078
0.5	16	64	0.000	0.185	0.045	0.063	0.007	0.180	0.048	0.064
0.5	32	1	0.400	1.603	0.113	0.296	0.417	1.165	0.130	0.331
0.5	32	2	0.316	1.461	0.106	0.281	0.345	1.093	0.122	0.313
0.5	32	4	0.204	1.218	0.094	0.253	0.244	0.955	0.108	0.280
0.5	32	8	0.085	0.871	0.077	0.209	0.127	0.725	0.088	0.228
0.5	32	16	0.016	0.546	0.060	0.152	0.045	0.490	0.067	0.162
0.5	32	32	0.001	0.323	0.047	0.107	0.016	0.307	0.052	0.112
0.5	32	64	0.000	0.199	0.045	0.083	0.008	0.194	0.048	0.084

Table 5. Weak Instrument Limit of LIML and Fuller. $\rho = 0.8$

			LIML				Fuller			
ρ	K	μ^2	Median	IQR	Bekker	Standard	Median	IQR	Bekker	Standard
0.8	1	1	0.290	1.044	0.113	0.113	0.556	0.429	0.495	0.203
0.8	1	2	0.124	0.891	0.102	0.102	0.390	0.400	0.321	0.164
0.8	1	4	0.027	0.699	0.087	0.087	0.220	0.370	0.175	0.126
0.8	1	8	0.001	0.498	0.074	0.074	0.100	0.355	0.108	0.100
0.8	1	16	-0.001	0.346	0.062	0.062	0.048	0.297	0.082	0.079
0.8	1	32	0.000	0.242	0.053	0.053	0.025	0.225	0.066	0.065
0.8	1	64	0.000	0.170	0.049	0.049	0.012	0.164	0.056	0.055
0.8	2	1	0.347	1.088	0.147	0.166	0.554	0.480	0.397	0.275
0.8	2	2	0.160	0.922	0.118	0.133	0.394	0.434	0.268	0.207
0.8	2	4	0.041	0.723	0.093	0.102	0.225	0.389	0.162	0.146
0.8	2	8	0.003	0.512	0.076	0.080	0.102	0.364	0.108	0.107
0.8	2	16	0.000	0.350	0.063	0.065	0.048	0.301	0.083	0.083
0.8	2	32	0.000	0.243	0.054	0.055	0.025	0.226	0.066	0.066
0.8	2	64	0.000	0.170	0.048	0.049	0.013	0.164	0.055	0.055
0.8	4	1	0.414	1.163	0.183	0.238	0.562	0.574	0.316	0.351
0.8	4	2	0.214	0.970	0.141	0.183	0.405	0.496	0.230	0.265
0.8	4	4	0.062	0.763	0.102	0.127	0.234	0.420	0.152	0.176
0.8	4	8	0.007	0.533	0.079	0.091	0.106	0.378	0.108	0.119
0.8	4	16	0.000	0.358	0.064	0.069	0.048	0.307	0.083	0.087
0.8	4	32	0.000	0.245	0.054	0.056	0.025	0.228	0.066	0.068
0.8	4	64	0.000	0.171	0.048	0.050	0.012	0.165	0.055	0.056
0.8	8	1	0.489	1.213	0.219	0.316	0.586	0.679	0.299	0.422
0.8	8	2	0.286	1.043	0.168	0.248	0.431	0.594	0.230	0.331
0.8	8	4	0.101	0.819	0.119	0.171	0.253	0.475	0.160	0.226
0.8	8	8	0.013	0.572	0.085	0.111	0.111	0.404	0.111	0.142
0.8	8	16	0.000	0.373	0.065	0.078	0.049	0.320	0.084	0.097
0.8	8	32	0.001	0.250	0.055	0.060	0.025	0.232	0.067	0.073
0.8	8	64	0.000	0.173	0.049	0.052	0.012	0.167	0.056	0.059
0.8	16	1	0.561	1.245	0.249	0.390	0.620	0.781	0.305	0.483
0.8	16	2	0.373	1.127	0.201	0.323	0.473	0.713	0.248	0.403
0.8	16	4	0.162	0.900	0.142	0.232	0.287	0.564	0.176	0.291
0.8	16	8	0.029	0.640	0.094	0.146	0.125	0.451	0.118	0.181
0.8	16	16	0.000	0.405	0.069	0.095	0.049	0.346	0.085	0.115
0.8	16	32	0.000	0.262	0.056	0.068	0.024	0.243	0.068	0.081
0.8	16	64	0.000	0.177	0.049	0.056	0.013	0.171	0.056	0.063
0.8	32	1	0.624	1.250	0.277	0.457	0.656	0.873	0.316	0.534
0.8	32	2	0.469	1.201	0.234	0.401	0.530	0.836	0.270	0.473
0.8	32	4	0.248	0.998	0.172	0.309	0.341	0.689	0.201	0.367
0.8	32	8	0.062	0.731	0.111	0.203	0.152	0.523	0.132	0.240
0.8	32	16	0.004	0.458	0.076	0.126	0.053	0.388	0.091	0.148
0.8	32	32	0.001	0.282	0.059	0.084	0.025	0.262	0.069	0.098
0.8	32	64	0.000	0.185	0.049	0.065	0.012	0.178	0.056	0.072

Figure 1: 3 Instruments: K and 2SLS

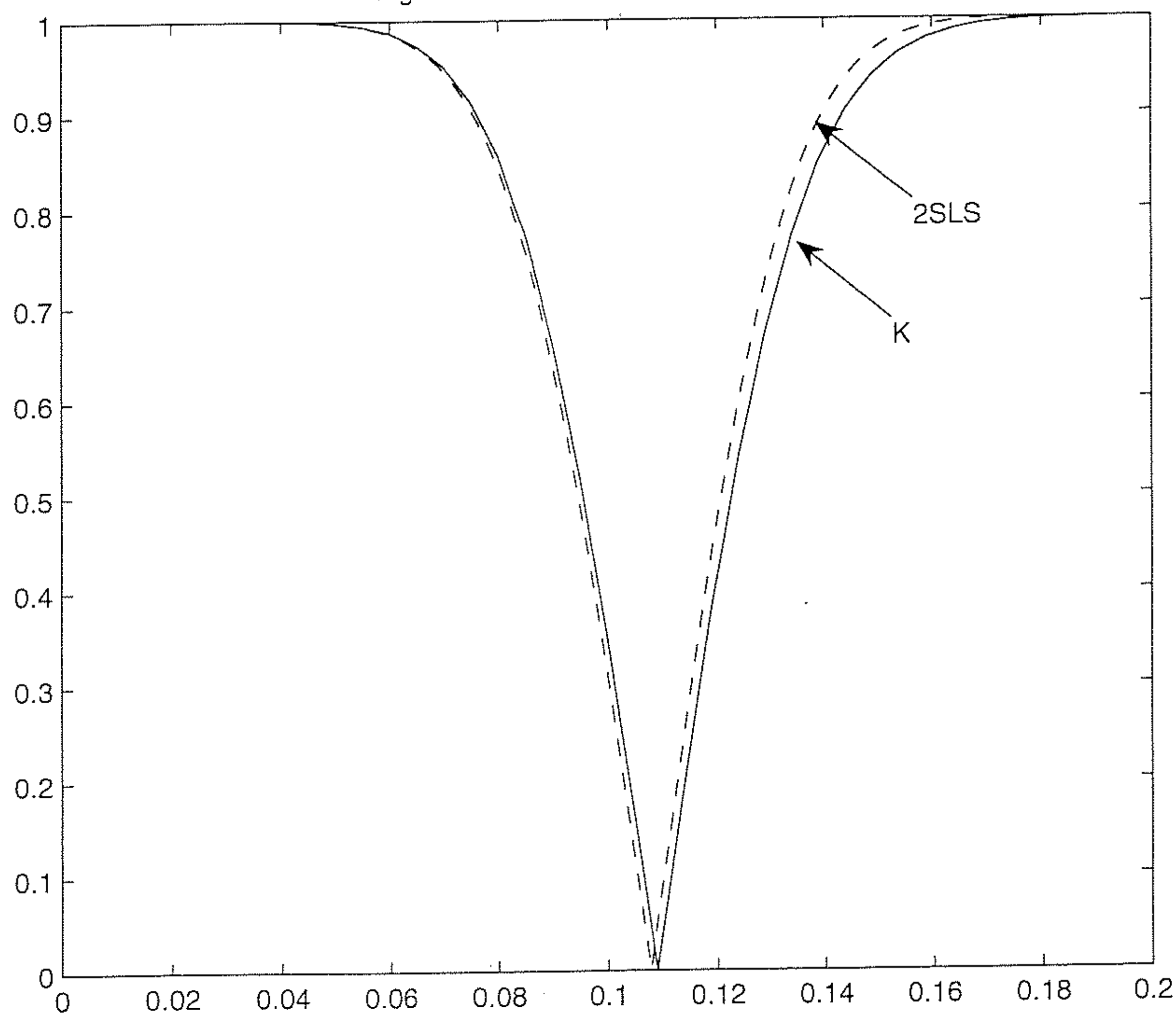

Figure 2: 180 Instruments: K and 2SLS

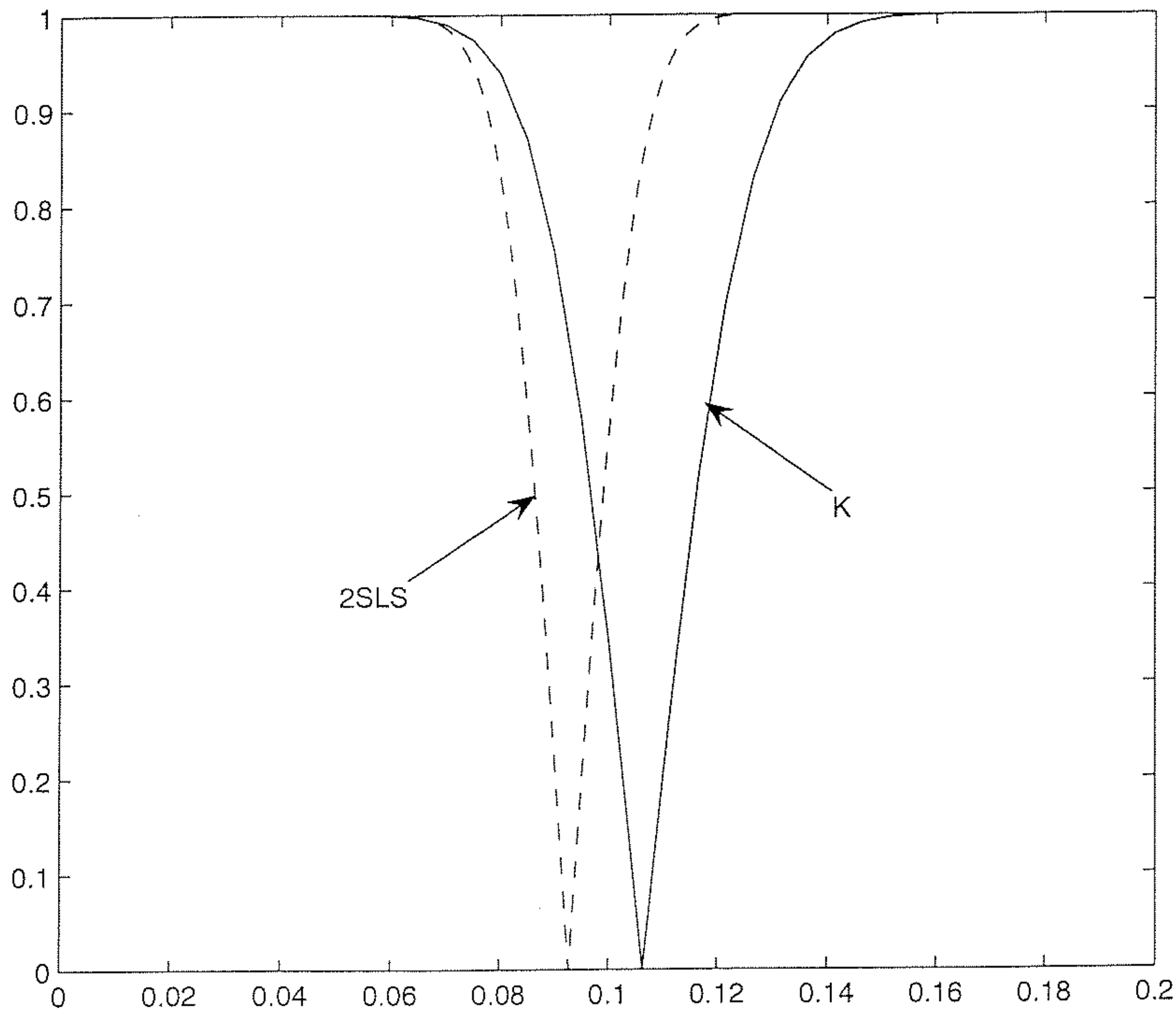

Figure 3: 180 Instruments: K and Fuller

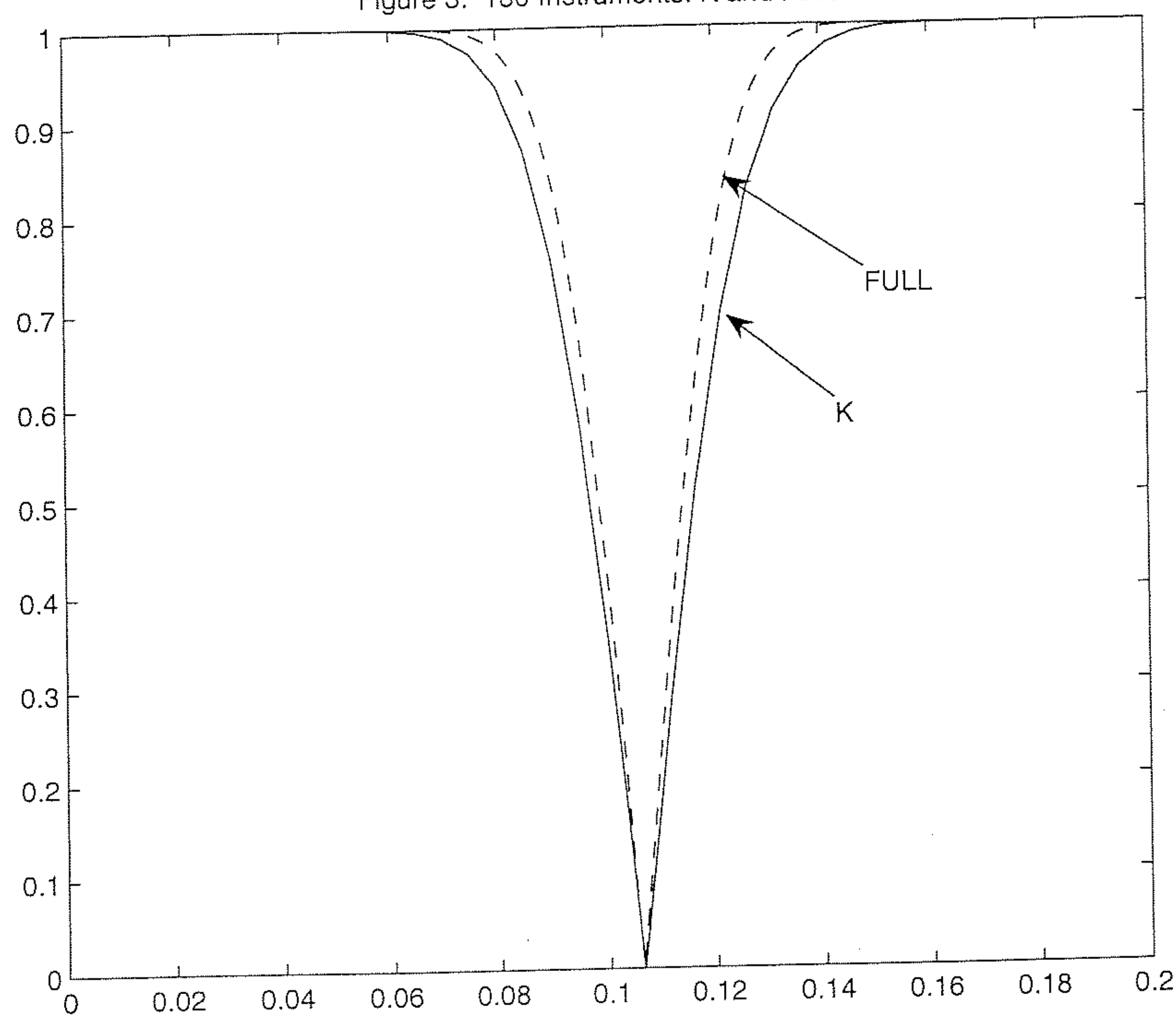

Figure 4: 180 Instruments: K and Fuller with Bekker Standard Errors