

Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments*

John C. Chao¹, Norman R. Swanson², Jerry A. Hausman³, Whitney K. Newey⁴, and Tiemen Woutersen⁵

¹University of Maryland, ²Rutgers University, ³MIT, ⁴MIT, ⁵Johns Hopkins

Preliminary and Incomplete

Abstract

This paper derives the limiting distributions of alternative jackknife IV (*JIV*) estimators and gives consistent standard errors with heteroskedasticity and many instruments. The asymptotic framework includes the many instrument sequence of Bekker (1994) and the many weak instrument sequence of Chao and Swanson (2005). We show that *JIV* estimators are asymptotically normal and standard errors are consistent provided that $\frac{\sqrt{K_n}}{r_n} \rightarrow 0$, as $n \rightarrow \infty$, where K_n and r_n denote, respectively, the number of instruments and the rate of growth of the concentration parameter. This is in contrast to the asymptotic behavior of such classical *IV* estimators as *LIML*, *B2SLS*, and *2SLS*, all of which are inconsistent in the presence of heteroskedasticity, unless $\frac{K_n}{r_n} \rightarrow 0$. We also show that the rate of convergence and the form of the asymptotic covariance matrix of the *JIV* estimators will in general depend on strength of the instruments as measured by the relative order of magnitude of r_n and K_n .

JEL classification: C13, C31.

Keywords: heteroskedasticity, instrumental variables, Jackknife estimation, many instruments, weak instruments.

* John C. Chao: Department of Economics, University of Maryland, College Park, MD, USA 20742, chao@econ.umd.edu. Whitney K. Newey: Department of Economics, MIT, Cambridge MA 02142-1347, wnewey@mit.edu. Norman R. Swanson: Department of Economics, Rutgers University, New Brunswick, NJ, USA 80901, nswanson@econ.rutgers.edu. Earlier versions of this paper were presented at the NSF/NBER conference on weak and/or many instruments at MIT in 2003, and at the 2004 winter meetings of the Econometric Society in San Diego, where conference participants provided many useful comments and suggestions. Particular thanks are owed to D.Ackerberg, D.Andrews, J.Angrist, M.Caner, M. Carrasco, P. Guggenberger, J.Hahn, G.Imbens, R.Klein, M.Moriera, G.D.A. Phillips, P.C.B. Phillips, and J.Stock for helpful discussions.

1 Introduction

It has long been known that the two-stage least squares (*2SLS*) estimator is biased with many instruments (see e.g. Sawa (1969), Phillips (1983), and the references cited therein). Due in large part to this problem, various approaches have been proposed in the literature for reducing the bias of the *2SLS* estimator. In recent years there has been interest in developing procedures based on using “delete-one” fitted values in lieu of the usual first-stage *OLS* fitted values, as the instruments employed in second stage estimation. A number of different versions of these estimators, referred to as jackknife instrumental variables (*JIV*) estimators, have been proposed and analyzed by Phillips and Hale (1977), Angrist, Imbens, and Krueger (1999), Blomquist and Dahlberg (1999), Akerberg and Deveraux (2003), Davidson and MacKinnon (2006), and Hausman, Newey, Woutersen, Chao, and Swanson (2007).

The *JIV* estimators are consistent with many instruments and heteroskedasticity of unknown form while other estimators, including the limited information maximum likelihood (*LIML*) and bias corrected *2SLS* (*B2SLS*) estimators are not; see Bekker and van der Ploeg (2005), Akerberg and Deveraux (2003), Chao and Swanson (2003), and Hasuman et. al. (2007). The main objective of this paper is to develop asymptotic theory for the *JIV* estimators in a setting that includes the many instrument sequence of Kunitomo (1980) and Bekker (1994) and the many instrument sequence of Chao and Swanson (2005). To be precise we show that *JIV* estimators are consistent and asymptotically normal when $\frac{\sqrt{K_n}}{r_n} \rightarrow 0$, as $n \rightarrow \infty$, where K_n and r_n denote, respectively, the number of instruments and the rate of growth of the so-called concentration parameter. In contrast, consistency of *LIML* and *B2SLS* generally requires $\frac{K_n}{r_n} \rightarrow 0$, as $n \rightarrow \infty$, meaning the number of instruments is small relative to identification strength. We show that both the rate of convergence of the *JIV* estimator and the form of its asymptotic covariance matrix depends on how weak the available instruments are, as measured by the relative order of magnitude of r_n vis-à-vis K_n . We also show consistency of standard errors under heteroskedasticity and many instruments.

In the process of showing asymptotic normality of *JIV* this paper gives a central limit for quadratic forms with an idempotent matrix. This theorem that can be used to study estimators other than *JIV*. For example, it has already been used in Hausman et. al. (2007) to derive the asymptotic properties of jackknife versions of *LIML* and of the Fuller (1977) estimator that are robust to heteroskedasticity and are as efficient as *LIML* under homoskedasticity.

This paper is a revised version of Chao and Swanson (2004), where a central limit theorem was first proved and consistent standard errors given. This version adds to Chao and Swanson (2004) by allowing for the many instrument sequence of Kunitomo (1980), Morimune (1983) and Bekker (1994) and giving a refined version of the central limit theorem.

The rest of the paper is organized as follows. Section 2 sets up the model and describes the estimators and standard errors. Section 3 lays out the framework for the asymptotic theory and presents the main results of our paper. Section 4 comments on the implications of these results. and concludes. All proofs are gathered in an appendix.

2 The Model and Estimators

The model we consider is given by

$$\begin{aligned} y_{n \times 1} &= X_{n \times G} \delta_0 + \varepsilon_{n \times 1}, \\ X &= \Upsilon + U, \end{aligned}$$

where n is the number of observations, G the number of right-hand side variables, Υ is a matrix of observations on the reduced form, and U is the matrix of disturbance observations. For the asymptotic approximations, the elements of Υ will be implicitly allowed to depend on n , although we suppress dependence of Υ on n for notational convenience. Estimation of δ_0 will be based on a $n \times K$ matrix Z of instrumental variable observations with $\text{rank}(Z) = K$. We will assume that $E[\varepsilon] = 0$ and $E[U] = 0$ conditional on (Υ, Z) .

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, some columns of X may be exogenous, with the corresponding column of U being zero. The model also allows for Z to approximate the reduced form. For example, let X'_i , Υ'_i , and Z'_i denote the i^{th} row (observation) for X , Υ , and Z respectively. We could have $\Upsilon_i = f_0(w_i)$ be an vector of unknown functions of a vector w_i of underlying instruments and $Z_i = (p_{1K}(w_i), \dots, p_{KK}(w_i))'$ for approximating functions $p_{kK}(w)$, such as power series or splines. In this case linear combinations of Z_i may approximate the unknown reduced form, e.g. as in Donald and Newey (2001).

To describe the estimators, let $P = Z(Z'Z)^{-1}Z'$. Let $\bar{\Pi}_{-i} = (Z'Z - Z_iZ'_i)^{-1}(Z'X - Z_iX'_i)$ be the reduced form coefficients obtained by regressing X on Z using all observations except the i^{th} .

The JIV estimator of Phillips and Hale (1977) is obtained as

$$\tilde{\delta} = \left(\sum_{i=1}^n \bar{\Pi}'_{-i} Z_i X_i' \right)^{-1} \sum_{i=1}^n \bar{\Pi}'_{-i} Z_i y_i.$$

Using standard results on recursive residuals it follows that

$$\bar{\Pi}'_{-i} Z_i = \left(X' Z (Z' Z)^{-1} Z_i - P_{ii} X_i' \right) / (1 - P_{ii}) = \sum_{j \neq i} P_{ij} X_j' / (1 - P_{ii}).$$

Then we have

$$\tilde{\delta} = \tilde{H}^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} y_j, \quad \tilde{H} = \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} X_j'.$$

The JIV estimator proposed by Angrist and Imbens (1999), their JIVE2, has a similar form except that $\Pi_{-i} = (Z' Z)^{-1} (Z' X - Z_i X_i')$ is used in place $\bar{\Pi}_{-i}$. It is given by

$$\hat{\delta} = \hat{H}^{-1} \sum_{i \neq j} X_i P_{ij} y_j, \quad \hat{H} = \sum_{i \neq j} X_i P_{ij} X_j'.$$

As we will show, these estimators are consistent and asymptotically normal under heteroskedasticity when $\sqrt{K}/r_n \rightarrow 0$, where r_n is proportional to the concentration parameter. In contrast, consistency of LIML and Fuller (1977) requires $K/r_n \rightarrow 0$ when P_{ii} is asymptotically correlated with $E[X_i \varepsilon_i] / E[\varepsilon_i^2]$, as discussed in Chao and Swanson (2004) and Hausman et. al. (2007), as does consistency of the bias corrected 2SLS estimator of Donald and Newey (2001), when P_{ii} is asymptotically correlated with $E[X_i \varepsilon_i]$, as discussed in Akerberg and Devereaux (2003). Thus, JIV estimators are robust to heteroskedasticity and many instruments (K growing as fast as r_n) while LIML, Fuller (1977), or bias corrected 2SLS estimators are not. The JIV estimators also have a closed form and are thus computationally simple relative to the jackknife versions of LIML and Fuller (1977) given in Hausman et. al. (2007), though they are not as efficient under homoskedasticity and many weak instruments.

The form of variance estimator can be motivated by noting that, for $\xi_i = (1 - P_{ii})^{-1} \varepsilon_i$, substituting $y_i = X_i' \delta_0 + \varepsilon_i$ in the equation for $\tilde{\delta}$ gives

$$\tilde{\delta} = \delta_0 + \tilde{H}^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j. \tag{1}$$

After appropriate normalization the matrix \tilde{H}^{-1} will converge and a central limit theorem will apply to $\sum_{i \neq j} X_i P_{ij} \xi_j$. This will lead to a sandwich form for the asymptotic variance. Here \tilde{H}^{-1} can be used to estimate the outside terms in the sandwich. The inside term, which is the variance

of $\sum_{i \neq j} X_i P_{ij} \xi_j$, can be estimated by dropping terms that are zero from the variance, removing the expectation, and replacing ξ_i by an estimate $\tilde{\xi}_i = (1 - P_{ii})^{-1} (y_i - X_i' \tilde{\delta})$. Using the independence of the observations, $E[\varepsilon_i] = 0$, and the exclusion of the $i = j$ terms in the double sums, it follows that

$$E\left[\sum_{i \neq j} X_i P_{ij} \xi_j \left(\sum_{i \neq j} X_i P_{ij} \xi_j\right)'\right] = E\left[\sum_{i,j} \sum_{k \notin \{i,j\}} P_{ik} P_{jk} X_i X_j' \xi_k^2 + \sum_{i \neq j} P_{ij}^2 X_i \xi_i X_j' \xi_j\right].$$

By removing the expectation and replacing ξ_i by $\tilde{\xi}_i$ we obtain

$$\tilde{\Sigma} = \sum_{i,j} \sum_{k \notin \{i,j\}} P_{ik} P_{jk} X_i X_j' \tilde{\xi}_k^2 + \sum_{i \neq j} P_{ij}^2 X_i \tilde{\xi}_i X_j' \tilde{\xi}_j.$$

The estimator of the asymptotic variance of $\tilde{\delta}$ is then given by

$$\tilde{V} = \tilde{H}^{-1} \tilde{\Sigma} \tilde{H}^{-1'}$$

This estimator is robust to heteroskedasticity in allowing $Var(\xi_i)$ to vary over i and $E[X_i \varepsilon_i]$ to vary over i .

An asymptotic variance estimator for $\hat{\delta}$ can be formed in an analogous way. Let $\hat{\varepsilon}_i = y_i - X_i' \hat{\delta}$ and

$$\hat{\Sigma} = \sum_{i,j} \sum_{k \notin \{i,j\}} P_{ik} P_{jk} X_i X_j' \hat{\varepsilon}_k^2 + \sum_{i \neq j} P_{ij}^2 X_i \hat{\varepsilon}_i X_j' \hat{\varepsilon}_j.$$

The variance estimator for $\hat{\delta}$ is given by

$$\hat{V} = \hat{H}^{-1} \hat{\Sigma} \hat{H}^{-1}$$

Here \hat{H} is symmetric because P is symmetric, so that a transpose is not needed for the third matrix in \hat{V} .

3 Many Instrument Asymptotics

The asymptotic theory we give combines the many instrument asymptotics of Kunitomo (1980), Mormune (1983), and Bekker (1994) with the many weak instrument asymptotics of Chao and Swanson (2005). All our regularity conditions are conditional on (Υ, Z) . To state the regularity conditions let $Z_i', \varepsilon_i, U_i'$, and Υ_i' denote the i^{th} row of Z, ε, U , and Υ respectively.

Assumption 1: $K = K_n \rightarrow \infty$, Z includes among its columns a vector of ones, and there is a positive constant $C < 1$ such that for all n large enough with probability one $\text{rank}(Z) = K$ and $P_{ii} \leq C$, ($i = 1, \dots, n$).

The restriction that $\text{rank}(Z) = K$ is a normalization that requires excluding redundant columns from Z . It can be verified in particular cases. For instance, when w_i is a continuously distributed scalar, $Z_i = p^K(w_i)$, and $p_{kK}(w) = w^{k-1}$ it can be shown that $Z'Z$ is nonsingular with probability one for $K < n$.¹ The condition $P_{ii} \leq C < 1$ implies that $K/n \leq C$, because $K/n = \sum_{i=1}^n P_{ii}/n \leq C$.

Let $\lambda_{\min}(A)$ denote the smallest eigenvalue of a symmetric matrix A .

Assumption 2: There is a $G \times G$ matrix $S_n = \tilde{S}_n \text{diag}(\mu_{1n}, \dots, \mu_{Gn})$ and z_i such that $\Upsilon_i = S_n z_i / \sqrt{n}$, \tilde{S}_n is bounded and the smallest eigenvalue of $\tilde{S}_n \tilde{S}_n'$ is bounded away from zero, for each j either $\mu_{jn} = \sqrt{n}$ or $\mu_{jn} / \sqrt{n} \rightarrow 0$, $\mu_n = \min_{1 \leq j \leq G} \mu_{jn} \rightarrow \infty$, and $\sqrt{K} / \mu_n^2 \rightarrow 0$. Also, there is $C > 0$ such that with probability one $\|\sum_{i=1}^n z_i z_i' / n\| \leq C$ and $\lambda_{\min}(\sum_{i=1}^n z_i z_i' / n) \geq C$.

Here $r_n = \mu_n^2$ can be interpreted as being proportional to the concentration parameter. For instance, note that in the $G = 1$ case we have $\sum_{i=1}^n \Upsilon_i^2 = \mu_n^2 \sum_{i=1}^n z_i^2 / n$, so by $\sum_{i=1}^n z_i^2 / n$ bounded and bounded away from zero, $\sum_{i=1}^n \Upsilon_i^2$ will be proportional to μ_n^2 . For $r_n = n$ we have asymptotic theory like Kunitomo (1980), Morimune (1984), and Bekker (1994), where the number of instruments K can grow as fast as the sample size. Allowing for K to grow and for r_n to grow slower than n models having many instruments without strong identification, as in Chao and Swanson (2005). The general form of Assumption 2 allows for some components of the reduced form to give only weak identification (corresponding to $\mu_{jn} / \sqrt{n} \rightarrow 0$) and other components (corresponding to $\mu_{jn} = \sqrt{n}$) to give strong identification. In particular, this condition allows for fixed constant coefficients in the reduced form.

It is sensible to have conditions on the rate of growth of the concentration parameter since the concentration parameter is a natural measure of instrument strength, as has been pointed out by numerous authors, including Phillips (1983), Rothenberg (1983), and Stock and Yogo (2005a). Assumption 2 stipulates that r_n must grow no faster than n but allows for r_n to grow much more slowly than n , as seems appropriate for modeling weak instruments.

¹The observations w_1, \dots, w_T are distinct with probability one and therefore, by $K < n$, 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 $a'Z'Za > 0$.

A fundamental rate condition is $\sqrt{K}/r_n \rightarrow 0$. This condition, formulated in Chao and Swanson (2005), ensures that random part of \tilde{H} does not dominate the nonrandom part is, it ensures that noise does not dominate signal for JIV estimators.

Assumption 3: There is constant C such that conditional on (Υ, Z) the observations $(\varepsilon_1, U_1), \dots, (\varepsilon_n, U_n)$ are independent with $E[\varepsilon_i] = 0$, $E[U_i] = 0$, $E[\varepsilon_i^2] < C$, $E[\|U_i\|^2] \leq C$.

This hypothesis requires second conditional moments of disturbances to be bounded and uniform nonsingularity of the variance of reduced form disturbances.

Assumption 4: There is π_{K_n} such that with probability one $\sum_{i=1}^n \|z_i - \pi_{K_n} Z_i\|^2 / n \rightarrow 0$.

This condition allows an unknown reduced form that is approximated by a linear combination of the instrumental variables.

We can easily interpret all these conditions in an important example of a linear model with exogenous covariates and a possibly unknown reduced form. This example is given by

$$X_i = \begin{pmatrix} \pi_{11} Z_{1i} + \mu_n f_0(w_i) / \sqrt{n} \\ Z_{1i} \end{pmatrix} + \begin{pmatrix} v_i \\ 0 \end{pmatrix}, Z_i = \begin{pmatrix} Z_{1i} \\ p^K(w_i) \end{pmatrix},$$

where Z_{1i} 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 $p^K(w) \stackrel{def}{=} (p_{1K}(w), \dots, p_{K-G_2, K}(w))'$. The variables in X_i other than Z_{1i} are endogenous with reduced form $\pi_{11} Z_{1i} + \mu_n f_0(w_i) / \sqrt{n}$. The function $f_0(w)$ may be a linear combination of a subvector of $p^K(w)$, in which case $z_i = \pi_{K_n} Z_i$ for some π_{K_n} in Assumption 4 or it may be an unknown function that can be approximated by a linear combination of $p^K(w)$. For $\mu_n = \sqrt{n}$ this example is like the model in Donald and Newey (2001) where Z_i includes approximating functions for the optimal (asymptotic variance minimizing) instruments Υ_i , but the number of instruments can grow as fast as the sample size. When $\mu_n^2/n \rightarrow 0$, it is a modified version where the model is more weakly identified.

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

$$z_i = \begin{pmatrix} f_0(w_i) \\ Z_{1i} \end{pmatrix}, S_n = \tilde{S}_n \text{diag}(\mu_n, \dots, \mu_n, \sqrt{n}, \dots, \sqrt{n}), \tilde{S}_n = \begin{pmatrix} I & \pi_{11} \\ 0 & I \end{pmatrix}.$$

By construction we have $\Upsilon_i = S_n z_i / \sqrt{n}$. Assumption 2 imposes the requirements that

$$\sum_{i=1}^n z_i z_i' / n \text{ is bounded and uniformly nonsingular.}$$

The other requirements of Assumption 2 are satisfied by construction. Assumption 3 requires conditional mean zero and bounded second moment for disturbances. For Assumption 4, let $\pi_{Kn} = [\tilde{\pi}'_{Kn}, [I_{G_2}, 0]]'$. Then Assumption 4 will be satisfied if there exists $\tilde{\pi}_{Kn}$ such that with probability one

$$\sum_{i=1}^n \|z_i - \pi_{Kn} Z_i\|^2/n = \sum_{i=1}^n \|f_0(w_i) - \tilde{\pi}'_{Kn} Z_i\|^2/n \longrightarrow 0.$$

THEOREM 1: *If Assumptions 1-4 are satisfied then $\mu_n^{-1} S'_n(\tilde{\delta} - \delta_0) \xrightarrow{p} 0$, $\hat{\delta} \xrightarrow{p} \delta_0$, $\mu_n^{-1} S'_n(\hat{\delta} - \delta_0) \xrightarrow{p} 0$, and $\hat{\delta} \xrightarrow{p} \delta_0$.*

The following additional condition is useful for asymptotic normality and consistency of the asymptotic variance.

Assumption 5: There is $C > 0$ such that with probability one, $\sum_{i=1}^n \|z_i\|^4/n^2 \longrightarrow 0$, $E[\varepsilon_i^4] < C$, $E[\|U_i\|^4] \leq C$.

To give asymptotic normality results we need to describe the asymptotic variances. We will give results that do not depend on convergence of various moment matrices, so we give the asymptotic variances as a function of n (rather than as a limit). Let

$$\begin{aligned} \bar{H}_n &= \sum_{i=1}^n z_i z_i' / n, \bar{\Omega}_n = \sum_{i=1}^n z_i z_i' \sigma_i^2 / n, \\ \bar{\Psi}_n &= S_n^{-1} \sum_{i \neq j} P_{ij}^2 \left(E[U_i U_i'] \sigma_j^2 (1 - P_{jj})^{-2} + E[U_i \varepsilon_i] (1 - P_{ii})^{-1} E[\varepsilon_j U_j'] (1 - P_{jj})^{-1} \right) S_n^{-1'}, \\ H_n &= \sum_{i=1}^n (1 - P_{ii}) z_i z_i' / n, \Omega_n = \sum_{i=1}^n (1 - P_{ii})^2 z_i z_i' \sigma_i^2 / n, \\ \Psi_n &= S_n^{-1} \sum_{i \neq j} P_{ij}^2 \left(E[U_i U_i'] \sigma_j^2 + E[U_i \varepsilon_i] E[\varepsilon_j U_j'] \right) S_n^{-1'}. \end{aligned}$$

When K/r_n is bounded the asymptotic variance of $S'_n(\tilde{\delta} - \delta_0)$ will be

$$\bar{V}_n = \bar{H}_n^{-1} (\bar{\Omega}_n + \bar{\Psi}_n) \bar{H}_n^{-1},$$

and the asymptotic variance of $S'_n(\hat{\delta} - \delta_0)$ will be

$$V_n = H_n^{-1} (\Omega_n + \Psi_n) H_n^{-1}.$$

To state the asymptotic normality results, let $A^{1/2}$ denote a square root matrix for a positive semi-definite matrix A , satisfying $A^{1/2} A^{1/2'} = A$. Also, for A nonsingular let $A^{-1/2} = (A^{1/2})^{-1}$.

THEOREM 2: *If Assumptions 1-5 are satisfied, $\sigma_i^2 \geq C > 0$ with probability one, and K/r_n is bounded, then with probability one \bar{V}_n and V_n are nonsingular for large enough n and*

$$\bar{V}_n^{-1/2} S'_n(\tilde{\delta} - \delta_0) \xrightarrow{d} N(0, I_G), V_n^{-1/2} S'_n(\hat{\delta} - \delta_0) \xrightarrow{d} N(0, I_G).$$

The convergence rate of the estimator is related to the size of S_n . In a simple case where δ is a scalar we can take $S_n = \mu_n = \sqrt{r_n}$. In this case the convergence rate of the estimator will simply be $1/\sqrt{r_n}$ when K/r_n is bounded. This rate, the inverse square root of the concentration parameter, is the usual one for instrumental variables. This rate changes when K grows faster than r_n .

The rate of convergence in Theorem 2 corresponds to the rate found by Stock and Yogo (2005) under homoskedasticity for LIML, Fuller's modified LIML, and B2SLS, when r_n grows at the same rate as K and slower than n .

The term $\bar{\Psi}_n$ in the asymptotic variance of $\tilde{\delta}$ and the term Ψ_n in the asymptotic variance of $\hat{\delta}$ account for the presence of many instruments. The order of these terms is K/r_n , so that if $K/r_n \rightarrow 0$ these terms can be dropped without affecting the asymptotic variance. When K/r_n is bounded but does not go to zero these terms have the same order as other terms and it is important to account for their presence in the standard errors. If $K/r_n \rightarrow \infty$ then these terms will dominate and slow down the convergence rate of the estimators. When $K/r_n \rightarrow \infty$ the asymptotic variance of $\sqrt{r_n/K} S'_n(\tilde{\delta} - \delta_0)$ will be

$$\bar{V}_n^* = \bar{H}_n^{-1}(r_n/K) \bar{\Psi}_n \bar{H}_n^{-1},$$

and the asymptotic variance of $\sqrt{r_n/K} S'_n(\hat{\delta} - \delta_0)$ will be

$$V_n^* = H_n^{-1}(r_n/K) \Psi_n H_n^{-1}.$$

When $K/r_n \rightarrow \infty$ the asymptotic variance matrices \bar{V}_n^* and V_n^* may be singular, especially when some components of X_i are exogenous or when different identification strengths are present. In order to allow for this singularity the asymptotic normality results are stated in terms of a linear combination of the estimator. Let L_n be a sequence of $\ell \times G$ matrices.

THEOREM 3: *Suppose that Assumptions 1-5 are satisfied and $K/r_n \rightarrow \infty$. If L_n is bounded and with probability one there is $C > 0$ such that $\lambda_{\min}(L_n \bar{V}_n^* L_n') \geq C$ then*

$$(L_n \bar{V}_n^* L_n')^{-1/2} L_n \sqrt{r_n/K} S'_n(\tilde{\delta} - \delta_0) \xrightarrow{d} N(0, I_\ell).$$

Also, if with probability one there is $C > 0$ such that $\lambda_{\min}(L_n V_n^* L_n') \geq C$ then

$$(L_n V_n^* L_n')^{-1/2} L_n \sqrt{r_n/K} S_n' (\hat{\delta} - \delta_0) \xrightarrow{d} N(0, I_\ell).$$

Here the convergence rate is related to the size of $(\sqrt{r_n/K})S_n$. In a simple case where δ is a scalar we can take $S_n = \sqrt{r_n}$, giving a convergence rate of \sqrt{K}/r_n . Here the conclusion of the theorem will be that $(r_n/\sqrt{K})(\tilde{\delta} - \delta_0)$ is asymptotically normal. It is interesting that $\sqrt{K}/r_n \rightarrow 0$ is a condition for consistency in this setting as well as in the context of Theorem 1 above.

From Theorems 2 and 3 it is clear that the rates of convergence of both JIV estimators depend in general on the strength of the available instruments as reflected in the relative orders of magnitude of r_n vis-à-vis K . Note also that, whenever r_n grows at a slower rate than n , the rate of convergence is slower than the conventional \sqrt{n} rate of convergence, since in this case the available instruments are weaker than that assumed in the conventional strongly identified case, where the concentration parameter is taken to grow at the rate n .

When $P_{ii} = Z_i'(Z'Z)^{-1}Z_i$ goes to zero uniformly in i , the asymptotic variances of the two JIV estimators will get close in large sample. Since $\sum_{i=1}^n P_{ii} = \text{tr}(P) = K$, P_{ii} going to zero will occur when K grows more slowly than n , though precise conditions for this depend on the nature of Z_i . As a practical matter, P_{ii} will generally be very close to zero in applications where K is very small relative to n , making the jackknife estimators very close to each other.

Under homoskedasticity we can compare the asymptotic variances of the two JIV estimators. With homoskedasticity the asymptotic variance of $\tilde{\delta}$ is

$$\begin{aligned} \bar{V}_n &= \bar{V}_n^1 + \bar{V}_n^2, \bar{V}_n^1 = \sigma^2 \bar{H}_n^{-1}, \bar{V}_n^2 = S_n^{-1} \sigma^2 E[U_i U_i'] \sum_{i \neq j} P_{ij}^2 / (1 - P_{jj})^2 S_n^{-1} \\ &+ S_n^{-1} E[U_i \varepsilon_i] E[U_i' \varepsilon_i] S_n^{-1'} \sum_{i \neq j} P_{ij}^2 (1 - P_{jj})^{-1} (1 - P_{jj})^{-1}. \end{aligned}$$

Also, the asymptotic variance of $\hat{\delta}$ is

$$\begin{aligned} V_n &= V_n^1 + V_n^2, V_n^1 = \sigma^2 H_n^{-1} \left[\sum_{i=1}^n (1 - P_{ii})^2 z_i z_i' / n \right] H_n^{-1}, \\ V_n^2 &= S_n^{-1} \left(\sigma^2 E[U_i U_i'] + E[U_i \varepsilon_i] E[U_i' \varepsilon_i] \right) S_n^{-1'} \sum_{i \neq j} P_{ij}^2. \end{aligned}$$

By the fact that $(1 - P_{ii})^{-1} > 1$ we have $\bar{V}_n^2 \geq V_n^2$ in the positive semi-definite sense. Also, note that \bar{V}_n^1 is the variance of an IV estimator with instruments $z_i(1 - P_{ii})$ while V_n^2 is the variance of

the corresponding least squares estimator, so that $\bar{V}_n^{-1} \leq V_n^{-1}$. Thus, it appears that in general we cannot rank the asymptotic variances of the two estimators.

Turning now to the consistency of the asymptotic variance estimators, we impose the following additional condition.

Assumption 6: *There is C with $\|z_i\| \leq C$ for all i , and there exists π_n such that $\max_{i \leq n} \|z_i - \pi_n Z_i\| \rightarrow 0$*

The next result shows that the estimators of the asymptotic variance we have given are consistent after normalization.

THEOREM 4: *Suppose that Assumptions 1-6 are satisfied. If K/r_n is bounded then $S_n' \tilde{V} S_n - \bar{V}_n \xrightarrow{p} 0$ and $S_n' \hat{V} S_n - V_n \xrightarrow{p} 0$. Also, if $K/r_n \rightarrow \infty$ then $r_n S_n' \tilde{V} S_n / K - \bar{V}_n^* \xrightarrow{p} 0$ and $r_n S_n' \hat{V} S_n / K - V_n^* \xrightarrow{p} 0$.*

A primary use of asymptotic variance estimators is in conducting approximate inference concerning coefficients.

THEOREM 5: *Suppose Assumptions 1-6 are satisfied and that $a(\delta)$ is an $\ell \times 1$ vector of functions such i) $a(\delta)$ is continuously differentiable in a neighborhood of δ_0 , ii) there is a square matrix B_n such that for $A = \partial a(\delta_0) / \partial \delta$, $B_n A S_n^{-1'}$ is bounded; iii) for any $\bar{\delta}_k \xrightarrow{p} \delta_0$, ($k = 1, \dots, \ell$) and $\bar{A} = [\partial a_1(\bar{\delta}_1) / \partial \delta, \dots, \partial a_\ell(\bar{\delta}_\ell) / \partial \delta]$, we have $B_n(\bar{A} - A) S_n^{-1'}$ $\xrightarrow{p} 0$. If K/r_n is bounded and $\lambda_{\min}(B_n A S_n^{-1'} \bar{V}_n S_n^{-1} A' B_n') \geq C$ or $K/r_n \rightarrow \infty$ and $\lambda_{\min}(B_n A S_n^{-1'} \bar{V}_n^* S_n^{-1} A' B_n') \geq C$ then for $\tilde{A} = \partial a(\tilde{\delta}) / \partial \delta$,*

$$(\tilde{A} \tilde{V} \tilde{A}')^{-1/2} [a(\tilde{\delta}) - a(\delta_0)] \xrightarrow{d} N(0, I).$$

If K/r_n is bounded and $\lambda_{\min}(B_n A S_n^{-1'} V_n S_n^{-1} A' B_n') \geq C$ or $K/r_n \rightarrow \infty$ and $\lambda_{\min}(B_n A S_n^{-1'} V_n^ S_n^{-1} A' B_n') \geq C$ then for $\hat{A} = \partial a(\hat{\delta}) / \partial \delta$,*

$$(\hat{A} \hat{V} \hat{A}')^{-1/2} [a(\hat{\delta}) - a(\delta_0)] \xrightarrow{d} N(0, I).$$

Perhaps the most important special case of this result is that for a single linear combination. For example, that case will lead to t-statistics based on the consistent variance estimator having the usual standard normal limiting distribution. The following result considers that case.

COROLLARY 6: *Suppose Assumptions 1-6 are satisfied and that c and b_n are such that $b_n c' S_n^{-1}$ is bounded. Then if K/r_n is bounded and $b_n^2 c' S_n^{-1} \bar{V}_n S_n^{-1} c \geq C$ or $K/r_n \rightarrow \infty$ and $b_n^2 c' S_n^{-1} \bar{V}_n^* S_n^{-1} c \geq C$ then*

$$\frac{c'(\tilde{\delta} - \delta_0)}{\sqrt{c' \tilde{V} c}} \xrightarrow{d} N(0, 1).$$

Also, if K/r_n is bounded and $c' S_n^{-1} V_n S_n^{-1} c / b_n^2 \geq C$ or $K/r_n \rightarrow \infty$ and $c' S_n^{-1} V_n^ S_n^{-1} c / b_n^2 \geq C$ then*

$$\frac{c'(\hat{\delta} - \delta_0)}{\sqrt{c' \hat{V} c}} \xrightarrow{d} N(0, 1).$$

4 Concluding Remarks

In this paper we have derived limiting distribution results for two alternative JIV estimators. These estimators are both seen to be consistent and asymptotically normal in the presence of many instruments and heteroskedasticity that precludes the consistency of LIML, 2SLS, and B2SLS in the context of heteroskedasticity of unknown form.

5 Appendix A - 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. We first give four lemmas that are useful in the proof of consistency, asymptotic normality, and consistency of the asymptotic variance estimator. We group them together here for ease of reference, because they are also used in Hausman et. al. (2007).

LEMMA A1: *If $(W_i, Y_i), (i = 1, \dots, n)$ are independent, W_i and Y_i are scalars, and P is symmetric, idempotent of rank K then for $\bar{w} = E[(W_1, \dots, W_n)']$, $\bar{y} = E[(Y_1, \dots, Y_n)']$, $\bar{\sigma}_{Wn} = \max_{i \leq n} \text{Var}(W_i)^{1/2}$, $\bar{\sigma}_{Yn}^2 = \max_{i \leq n} \text{Var}(Y_i)^{1/2}$,*

$$\sum_{i \neq j} P_{ij} W_i Y_j = \sum_{i \neq j} P_{ij} \bar{w}_i \bar{y}_j + O_p(K^{1/2} \bar{\sigma}_{Wn} \bar{\sigma}_{Yn} + \bar{\sigma}_{Wn} \sqrt{\bar{y}' \bar{y}} + \bar{\sigma}_{Yn} \sqrt{\bar{w}' \bar{w}}).$$

Proof. Let $\tilde{w}_i = W_i - \bar{w}_i$ and $\tilde{y}_i = Y_i - \bar{y}_i$. Note that

$$\sum_{i \neq j} P_{ij} W_i Y_j - \sum_{i \neq j} P_{ij} \bar{w}_i \bar{y}_j = \sum_{i \neq j} P_{ij} \tilde{w}_i \tilde{y}_j + \sum_{i \neq j} P_{ij} \tilde{w}_i \bar{y}_j + \sum_{i \neq j} P_{ij} \bar{w}_i \tilde{y}_j.$$

Let $D_n = \bar{\sigma}_{Wn}^2 \bar{\sigma}_{Yn}^2$. Note that for $i \neq j$ and $k \neq \ell$, $E[\tilde{w}_i \tilde{y}_j \tilde{w}_k \tilde{y}_\ell]$ is zero unless $i = k$ and $j = \ell$ or $i = \ell$ and $j = k$. Then by CS and $\sum_j P_{ij} = P_{ii}$,

$$\begin{aligned} E \left[\left(\sum_{i \neq j} P_{ij} \tilde{w}_i \tilde{y}_j \right)^2 \right] &= \sum_{i \neq j} \sum_{k \neq \ell} P_{ij} P_{k\ell} E[\tilde{w}_i \tilde{y}_j \tilde{w}_k \tilde{y}_\ell] \\ &= \sum_{i \neq j} P_{ij}^2 \left(E[\tilde{w}_i^2] E[\tilde{y}_j^2] + E[\tilde{w}_i \tilde{y}_i] E[\tilde{w}_j \tilde{y}_j] \right) \\ &\leq 2D_n \sum_{i \neq j} P_{ij}^2 \leq 2D_n \sum_i P_{ii} = 2D_n K. \end{aligned}$$

Then by M, $\sum_{i \neq j} P_{ij} \tilde{w}_i \tilde{y}_j = O_p(D_n^{1/2} K^{1/2})$. Also, for $\tilde{w} = (\tilde{w}_1, \dots, \tilde{w}_n)'$, we have $\sum_{i \neq j} P_{ij} \tilde{w}_i \bar{y}_j = \tilde{w} P \bar{y} - \sum_i P_{ii} \bar{y}_i \tilde{w}_i$. By independence across i we have $E[\tilde{w} \tilde{w}'] \leq \bar{\sigma}_{Wn}^2 I_n$, so that

$$\begin{aligned} E[(\bar{y}' P \tilde{w})^2] &= \bar{y}' P E[\tilde{w} \tilde{w}'] P \bar{y} \leq \bar{\sigma}_{Wn}^2 \bar{y}' P \bar{y} \leq \bar{\sigma}_{Wn}^2 \bar{y}' \bar{y}, \\ E \left[\left(\sum_i P_{ii} \bar{y}_i \tilde{w}_i \right)^2 \right] &= \sum_i P_{ii}^2 E[\tilde{w}_i^2] \bar{y}_i^2 \leq \bar{\sigma}_{Wn}^2 \bar{y}' \bar{y}. \end{aligned}$$

Then by M and T we have $\sum_{i \neq j} P_{ij} \tilde{w}_i \bar{y}_j = O_p(\bar{\sigma}_{Wn} \sqrt{\bar{y}' \bar{y}})$. Interchanging the roles of Y_i and W_i gives $\sum_{i \neq j} P_{ij} \bar{w}_i \tilde{y}_j = O_p(\bar{\sigma}_{Yn} \sqrt{\bar{w}' \bar{w}})$. The conclusion then follows by T. Q.E.D.

LEMMA A2: If *i*) P is a symmetric, idempotent matrix with $\text{rank}(P) = K$, $P_{ii} \leq C < 1$; *ii*) $(W_{1n}, U_1, \varepsilon_1), \dots, (W_{nn}, U_n, \varepsilon_n)$ are independent and $D_n = \sum_{i=1}^n E[W_{in}W'_{in}]$ is bounded; *iii*) $E[W'_{in}] = 0$, $E[U_i] = 0$, $E[\varepsilon_i] = 0$ and there exists a constant C such that $E[\|U_i\|^4] \leq C$, $E[\varepsilon_i^4] \leq C$; *iv*) $\sum_{i=1}^n E[\|W_{in}\|^4] \rightarrow 0$; *v*) $K \rightarrow \infty$; then for $\bar{\Sigma}_n \stackrel{\text{def}}{=} \sum_{i \neq j} P_{ij}^2 (E[U_i U'_i] E[\varepsilon_j^2] + E[U_i \varepsilon_i] E[\varepsilon_j U'_j]) / K$ and any bounded sequences c_{1n} and c_{2n} of conformable vectors with $\Xi_n = c'_{1n} D_n c_{1n} + c'_{2n} \bar{\Sigma}_n c_{2n} > C$ it follows that

$$Y_n = \Xi_n^{-1/2} \left(\sum_{i=1}^n c'_{1n} W_{in} + c'_{2n} \sum_{i \neq j} U_i P_{ij} \varepsilon_j / \sqrt{K} \right) \xrightarrow{d} N(0, 1).$$

Proof: The proof of Lemma A2 is long and is deferred to Appendix B.

The next two results are helpful in proving consistency of the variance estimator. They use the same notation as Lemma A1 and $\bar{\mu}_W = \max_{i \leq n} |E[W_i]|$ and $\bar{\mu}_Y = \max_{i \leq n} |E[Y_i]|$.

LEMMA A3: If $(W_i, Y_i), (i = 1, \dots, n)$ are independent, W_i and Y_i are scalars then

$$\sum_{i \neq j} P_{ij}^2 W_i Y_j = E \left[\sum_{i \neq j} P_{ij}^2 W_i Y_j \right] + O_p \left(\{K(\bar{\sigma}_W^2 \bar{\sigma}_Y^2 + \bar{\sigma}_W^2 \bar{\mu}_Y^2 + \bar{\mu}_W^2 \bar{\sigma}_Y^2)\}^{1/2} \right).$$

Proof: Using the notation of the proof of Lemma A1, we have

$$\sum_{i \neq j} P_{ij}^2 W_i Y_j - \sum_{i \neq j} P_{ij}^2 \bar{w}_i \bar{y}_j = \sum_{i \neq j} P_{ij}^2 \tilde{w}_i \tilde{y}_j + \sum_{i \neq j} P_{ij}^2 \tilde{w}_i \bar{y}_j + \sum_{i \neq j} P_{ij}^2 \bar{w}_i \tilde{y}_j.$$

As before, for $i \neq j$ and $k \neq \ell$, $E[\tilde{w}_i \tilde{y}_j \tilde{w}_k \tilde{y}_\ell]$ is zero unless $i = k$ and $j = \ell$ or $i = \ell$ and $j = k$. Then by CS, $P_{ij}^4 \leq P_{ij}^2$, and $\sum_j P_{ij} = P_{ii}$,

$$\begin{aligned} E \left[\left(\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \tilde{y}_j \right)^2 \right] &= \sum_{i \neq j} \sum_{k \neq \ell} P_{ij}^2 P_{kl}^2 E[\tilde{w}_i \tilde{y}_j \tilde{w}_k \tilde{y}_\ell] \\ &= \sum_{i \neq j} P_{ij}^4 \left(E[\tilde{w}_i^2] E[\tilde{y}_j^2] + E[\tilde{w}_i \tilde{y}_i] E[\tilde{w}_j \tilde{y}_j] \right) \\ &\leq 2\bar{\sigma}_W^2 \bar{\sigma}_Y^2 \sum_{i \neq j} P_{ij}^4 \leq 2\bar{\sigma}_W^2 \bar{\sigma}_Y^2 K. \end{aligned}$$

Then by M, $\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \tilde{y}_j = O_p((K\bar{\sigma}_W^2 \bar{\sigma}_Y^2)^{1/2})$. Also, $\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \bar{y}_j = \tilde{w} \tilde{P} \bar{y} - \sum_i P_{ii}^2 \bar{y}_i \tilde{w}_i$ where

$\tilde{P}_{ij} = P_{ij}^2$. By independence across i we have $E[\tilde{w}\tilde{w}'] \leq \bar{\sigma}_{Wn}^2 I_n$, so that

$$\begin{aligned} E[(\tilde{y}'\tilde{P}\tilde{w})^2] &= \tilde{y}'\tilde{P}E[\tilde{w}\tilde{w}']\tilde{P}\tilde{y} \leq \bar{\sigma}_{Wn}^2 \tilde{y}'\tilde{P}^2\tilde{y} = \bar{\sigma}_{Wn}^2 \sum_{i,j,k} \tilde{y}_i P_{ik}^2 P_{kj}^2 \tilde{y}_j \\ &\leq \bar{\sigma}_W^2 \bar{\mu}_Y^2 \sum_{i,j,k} P_{ik}^2 P_{kj}^2 = \bar{\sigma}_W^2 \bar{\mu}_Y^2 \sum_k (\sum_i P_{ik}^2) (\sum_j P_{kj}^2) \\ &= \bar{\sigma}_W^2 \bar{\mu}_Y^2 \sum_k P_{kk}^2 \leq K \bar{\sigma}_W^2 \bar{\mu}_Y^2. \\ E\left[\left(\sum_i P_{ii}^2 \tilde{y}_i \tilde{w}_i\right)^2\right] &= \sum_i P_{ii}^4 E[\tilde{w}_i^2] \tilde{y}_i^2 \leq K \bar{\sigma}_W^2 \bar{\mu}_Y^2. \end{aligned}$$

Then by M and T we have $\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \tilde{y}_j = O_p((K \bar{\sigma}_W^2 \bar{\mu}_Y^2)^{1/2})$. Interchanging the roles of Y_i and W_i gives $\sum_{i \neq j} P_{ij}^2 \tilde{w}_i \tilde{y}_j = O_p((K \bar{\sigma}_Y^2 \bar{\mu}_W^2)^{1/2})$. The conclusion then follows by T. Q.E.D.

As a notational convention we let $\sum_{i \neq j \neq k}$ denote $\sum_i \sum_{j \neq i} \sum_{k \notin \{i,j\}}$.

LEMMA A4: *If W_i, Y_i, η_i , are independent across i with $E[W_i] = a_i/\sqrt{n}$, $E[Y_i] = b_i/\sqrt{n}$, $|a_i| \leq C$, $|b_i| \leq C$, $E[\eta_i^2] \leq C$, $\text{Var}(W_i) \leq C/r_n$, $\text{Var}(Y_i) \leq C/r_n$, there exists π_n such that $\max_{i \leq n} |a_i - Z_i' \pi_n| \rightarrow 0$, and $\sqrt{K}/r_n \rightarrow 0$ then*

$$A_n = E\left[\sum_{i \neq j \neq k} W_i P_{ik} \eta_k P_{kj} Y_j\right] = O(1), \quad \sum_{i \neq j \neq k} W_i P_{ik} \eta_k P_{kj} Y_j - A_n \xrightarrow{p} 0.$$

Proof: Given in Appendix B.

LEMMA A5: *If Assumptions 1-3 are satisfied then*

$$\begin{aligned} S_n^{-1} \tilde{H} S_n^{-1'} &= \sum_{i \neq j} z_i P_{ij} (1 - P_{jj})^{-1} z_j' / n + o_p(1), \quad S_n^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} \varepsilon_j / \mu_n \xrightarrow{p} 0, \\ S_n^{-1} \hat{H} S_n^{-1'} &= \sum_{i \neq j} z_i P_{ij} z_j' / n + o_p(1), \quad S_n^{-1} \sum_{i \neq j} X_i P_{ij} \varepsilon_j / \mu_n \xrightarrow{p} 0. \end{aligned}$$

Proof: Apply Lemma A1 with $Y_i = e_k' S_n^{-1} X_i = z_{ik} / \sqrt{n} + e_k' S_n^{-1} U_i$ and $W_i = e_\ell' S_n^{-1} X_i (1 - P_{ii})^{-1}$ for some k and ℓ . Note that since $\|S_n^{-1}\| \leq C/\sqrt{r_n}$,

$$E[Y_i] = z_{ik} / \sqrt{n}, \quad \text{Var}(Y_i) \leq C/r_n, \quad E[W_i] = z_{i\ell} / \sqrt{n} (1 - P_{ii}), \quad \text{Var}(W_i) \leq C/r_n.$$

Note that

$$\begin{aligned} \sqrt{K} \bar{\sigma}_{Wn} \bar{\sigma}_{Yn} &\leq C \sqrt{K} / r_n \rightarrow 0, \quad \bar{\sigma}_{Wn} \sqrt{\tilde{y}'\tilde{y}} \leq C r_n^{-1/2} \sqrt{\sum z_{ik}^2 / n} \rightarrow 0, \\ \bar{\sigma}_{Yn} \sqrt{\tilde{w}'\tilde{w}} &\leq C r_n^{-1/2} \sqrt{\sum_i z_{i\ell}^2 (1 - P_{ii})^{-2} / n} \leq C r_n^{-1/2} \sqrt{\sum_i z_{i\ell}^2 / n} \rightarrow 0. \end{aligned}$$

Since $e'_k S_n^{-1} \tilde{H} S_n^{-1'} e_\ell = e'_k S_n^{-1} \sum_{i \neq j} X_i P_{ij} X'_j S_n^{-1'} e_\ell / (1 - P_{jj}) = \sum_{i \neq j} W_i P_{ij} Y_j$ and $P_{ij} \bar{w}_i \bar{y}_j = P_{ij} z_{ik} z_{j\ell} / n (1 - P_{jj})$, by Lemma A1 the first conclusion holds for the (k, ℓ) element. The first conclusion then follows by applying this same argument to each element.

For the second conclusion, apply Lemma A1 with $Y_i = e'_k S_n^{-1} X_i$ as before and $W_i = \varepsilon_i / \sqrt{r_n} (1 - P_{ii})$.

For the third conclusion, apply Lemma A1 with $W_i = e'_k S_n^{-1} X_i$ as before and $Y_i = e'_\ell S_n^{-1} X_i$, so that

$$\sqrt{K} \bar{\sigma}_{Wn} \bar{\sigma}_{Yn} \leq C \sqrt{K} / r_n \longrightarrow 0, \bar{\sigma}_{Wn} \sqrt{\bar{y}' \bar{y}} \leq C r_n^{-1/2} \sqrt{\sum z_{ik}^2 / n} \longrightarrow 0, \bar{\sigma}_{Yn} \sqrt{\bar{w}' \bar{w}} \longrightarrow 0.$$

The fourth conclusion follows similarly. Q.E.D.

$$\text{Let } \bar{H}_n = \sum_i z_i z'_i / n \text{ and } H_n = \sum_i (1 - P_{ii}) z_i z'_i / n.$$

LEMMA A6: *If Assumptions 1-4 are satisfied then*

$$S_n^{-1} \tilde{H} S_n^{-1'} = \bar{H}_n + o_p(1), S_n^{-1} \hat{H} S_n^{-1'} = H_n + o_p(1).$$

Proof: Let $\bar{z}_i = \sum_{j=1}^n P_{ij} z_j$ be the i^{th} element of Pz and note that

$$\begin{aligned} \sum_{i=1}^n \|z_i - \bar{z}_i\|^2 / n &= \|(I - P)z\|^2 / n = \text{tr}(z'(I - P)z/n) = \text{tr}[(z - Z\pi'_{K_n})'(I - P)(z - Z\pi'_{K_n})/n] \\ &\leq \text{tr}[(z - Z\pi'_{K_n})'(z - Z\pi'_{K_n})/n] = \sum_{i=1}^n \|z_i - \pi_{K_n} z_i\|^2 / n \longrightarrow 0. \end{aligned}$$

It follows that

$$\begin{aligned} \left\| \sum_i (\bar{z}_i - z_i) (1 - P_{ii})^{-1} z'_i / n \right\| &\leq \sum_i \|\bar{z}_i - z_i\| \left\| (1 - P_{ii})^{-1} z'_i \right\| / n \\ &\leq \sqrt{\sum_i \|\bar{z}_i - z_i\|^2 / n} \sqrt{\sum_i \|(1 - P_{ii})^{-1} z_i\|^2 / n} \longrightarrow 0. \end{aligned}$$

Then

$$\begin{aligned} \sum_{i \neq j} z_i P_{ij} (1 - P_{jj})^{-1} z'_j / n &= \sum_{i,j} z_i P_{ij} (1 - P_{jj})^{-1} z'_j / n - \sum_i z_i P_{ii} (1 - P_{ii})^{-1} z'_i / n \\ &= \sum_i \bar{z}_i (1 - P_{ii})^{-1} z'_i / n - \sum_i z_i P_{ii} (1 - P_{ii})^{-1} z'_i / n \\ &= \bar{H}_n + \sum_i (\bar{z}_i - z_i) (1 - P_{ii})^{-1} z'_i / n = \bar{H}_n + o(1). \end{aligned}$$

The first conclusion then follows from Lemma A5 and the triangle inequality. Also, as in the last equation we have

$$\begin{aligned} \sum_{i \neq j} z_i P_{ij} z'_j / n &= \sum_{i,j} z_i P_{ij} z'_j / n - \sum_i P_{ii} z_i z'_i / n = \sum_i \bar{z}_i z'_i / n - \sum_i P_{ii} z_i z'_i / n \\ &= H_n + \sum_i (\bar{z}_i - z_i) z'_i / n = H_n + o(1), \end{aligned}$$

so the second conclusion follows similarly to the first. Q.E.D.

Proof of Theorem 1: First, note that by $\lambda_{\min}(S_n S'_n / r_n) \geq \lambda_{\min}(\tilde{S}_n \tilde{S}'_n) \geq C$ we have

$$\left\| S'_n (\tilde{\delta} - \delta_0) / \mu_n \right\| \geq \lambda_{\min}(S_n S'_n / r_n)^{1/2} \left\| \tilde{\delta} - \delta_0 \right\| \geq C \left\| \tilde{\delta} - \delta_0 \right\|.$$

Therefore, $S'_n (\tilde{\delta} - \delta_0) / \mu_n \xrightarrow{p} 0$ will imply $\tilde{\delta} \xrightarrow{p} \delta_0$. Note that by Assumption 2, \bar{H}_n is bounded and $\lambda_{\min}(\bar{H}_n) \geq C$ for large enough n , with probability one. For \tilde{H} from Section 2, it follows from Lemma A6 and Assumption 2 that with probability one $\lambda_{\min}(S_n^{-1} \tilde{H} S_n^{-1'}) \geq C$ as the sample size grows. Hence $(S_n^{-1} \tilde{H} S_n^{-1'})^{-1} = O_p(1)$. By eq. (1),

$$\mu_n^{-1} S'_n (\tilde{\delta} - \delta_0) = (S_n^{-1} \tilde{H} S_n^{-1'})^{-1} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j / \mu_n = O_p(1) o_p(1) \xrightarrow{p} 0.$$

All of the previous statements are conditional on conditional on (Υ, Z) for a given sample size n , so that for the random variable $R_n = \mu_n^{-1} S'_n (\tilde{\delta} - \delta_0)$ we have shown that for any constant $v > 0$, with probability one

$$\Pr(\|R_n\| \geq v | \Upsilon, Z) \longrightarrow 0.$$

Then by the dominated convergence theorem,

$$\Pr(\|R_n\| \geq v) = E[\Pr(\|R_n\| \geq v | \Upsilon, Z)] \longrightarrow 0.$$

Therefore, since v is arbitrary, it follows that $R_n = \mu_n^{-1} S'_n (\tilde{\delta} - \delta_0) \xrightarrow{p} 0$.

Next note that $P_{ii} \leq C < 1$, so in the positive semi-definite sense in large enough samples with probability one,

$$H_n = \sum (1 - P_{ii}) z_i z'_i / n \geq C \bar{H}_n,$$

Thus, H_n is bounded and bounded away from singularity for large enough n with probability one. Then the rest of the conclusion follow analogously with $\hat{\delta}$ replacing $\tilde{\delta}$ and H_n replacing \bar{H}_n . Q.E.D.

We now turn to the asymptotic normality results. In what follows let $\xi_i = \varepsilon_i$ when considering the JIV2 estimator and let $\xi_i = \varepsilon_i/(1 - P_{ii})$ when considering JIV1.

Proof of Theorem 2: Note that $E[\xi_i^2] \leq C$ so that

$$E\left[\left\|\sum_{i=1}^n (z_i - \bar{z}_i) \xi_i / \sqrt{n}\right\|^2\right] = \sum_{i=1}^n \|z_i - \bar{z}_i\|^2 E[\xi_i^2]/n \leq C \sum_{i=1}^n \|z_i - \bar{z}_i\|^2 / n \longrightarrow 0.$$

Then by M,

$$S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j - \sum_i z_i (1 - P_{ii}) \xi_i / \sqrt{n} - S_n^{-1} \sum_{i \neq j} U_i P_{ij} \xi_j = \sum_{i=1}^n (z_i - \bar{z}_i) \xi_i / \sqrt{n} \xrightarrow{p} 0.$$

Let $W_{in} = z_i(1 - P_{ii})\xi_i/\sqrt{n}$ and

$$\begin{aligned} \Gamma_n &= \text{Var}\left(\sum_{i=1}^n W_{in} + \sum_{i \neq j} S_n^{-1} U_i P_{ij} \xi_j\right) \\ &= \sum_{i=1}^n z_i z_i' (1 - P_{ii})^2 E[\xi_i^2]/n + S_n^{-1} \sum_{i \neq j} P_{ij}^2 \left(E[U_i U_i'] E[\xi_j^2] + E[U_i \xi_i] E[U_j' \xi_j]\right) S_n^{-1'} \end{aligned}$$

Note that $\mu_n S_n^{-1}$ is bounded by Assumption 2 and that $\sum_{i \neq j} P_{ij}^2 / K \leq 1$, so by z_i bounded, Assumption 3, and K/μ_n^2 bounded it follows that Γ_n is bounded. Also, $E[\xi_i^2] \geq C > 0$, so that

$$\Gamma_n \geq \sum_{i=1}^n z_i z_i' (1 - P_{ii})^2 E[\xi_i^2]/n \geq C \sum_{i=1}^n z_i z_i' / n,$$

so by Assumption 2 $\lambda_{\min}(\Gamma_n) \geq C > 0$ for all n large enough. It follows that Γ_n^{-1} exists and is bounded in n for n large enough.

Let α be a $G \times 1$ nonzero vector. Now apply Lemma A2 with U_i there equal to U_i here, ε_i there equal to ξ_i here, $W_{in} = z_i(1 - P_{ii})\xi_i/\sqrt{n}$, $c_{1n} = \alpha' \Gamma_n^{-1}$, and $c_{2n} = \alpha' \Gamma_n^{-1} S_n^{-1} \sqrt{K}$. Note that condition i) of Lemma A2 is satisfied. Also, by z_i and $E[\xi_i^2]$ bounded, condition ii) of Lemma A2 is satisfied and condition iii) is satisfied by Assumptions 3 and 5. Also, by $(1 - P_{ii})^{-1} \leq C$ and Assumption 5,

$$\sum_{i=1}^n E[\|W_{in}\|^4] \leq C \sum_{i=1}^n \|z_i\|^4 / n^2 \longrightarrow 0.$$

so condition iv) is satisfied. Finally, condition v) is satisfied by hypothesis. Note also that c_{1n} is bounded, $c_{2n} = \alpha' \Gamma_n^{-1} S_n^{-1} \mu_n \sqrt{K} / \mu_n$ is bounded by hypothesis, and $\Xi_n = \alpha' \alpha$ by construction.

Then by the conclusion of Lemma A2,

$$\begin{aligned} (\alpha'\alpha)^{-1/2} \alpha'\Gamma_n^{-1/2} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j &= (\alpha'\alpha)^{-1/2} \alpha'\Gamma_n^{-1/2} \left(\sum_{i=1}^n W_{in} + \sum_{i \neq j} S_n^{-1} U_i P_{ij} \xi_j + o_p(1) \right) \\ &= \Xi_n^{-1/2} \left(\sum_{i=1}^n c'_{1n} W_{in} + c'_{2n} \sum_{i \neq j} U_i P_{ij} \varepsilon_j / \sqrt{K} \right) \xrightarrow{d} N(0, 1). \end{aligned}$$

It follow that $\alpha'\Gamma_n^{-1/2} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j \xrightarrow{d} N(0, \alpha'\alpha)$, so by the Cramer-Wold device,

$$\Gamma_n^{-1/2} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j \xrightarrow{d} N(0, I_G).$$

Consider now the JIV1 estimator, where $\xi_i = \varepsilon_i / (1 - P_{ii})$ and $\Gamma_n = \bar{\Omega}_n + \bar{\Psi}_n$, so that

$$(\bar{\Omega}_n + \bar{\Psi}_n)^{-1/2} S_n^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} \varepsilon_j \xrightarrow{d} N(0, I_G).$$

Note that $B_n = \bar{V}_n^{-1/2} \bar{H}_n^{-1} \Gamma_n^{1/2}$ is an orthogonal matrix, since $B_n B_n' = \bar{V}_n^{-1/2} \bar{V}_n \bar{V}_n^{-1/2} = I$. Also, $\bar{V}_n^{-1/2}$ is bounded by $\lambda_{\min}(\bar{V}_n) \geq C$, and $\Gamma_n^{1/2}$ is also bounded by Γ_n bounded. By Lemma A6, $(S_n^{-1} \tilde{H} S_n^{-1'})^{-1} = \bar{H}_n^{-1} + o_p(1)$. Therefore, we have

$$\begin{aligned} \bar{V}_n^{-1/2} (S_n^{-1} \tilde{H} S_n^{-1'})^{-1} \Gamma_n^{1/2} &= \bar{V}_n^{-1/2} (\bar{H}_n^{-1} + o_p(1)) \Gamma_n^{1/2} \\ &= B_n + o_p(1). \end{aligned}$$

Note also that if $Y_n \xrightarrow{d} N(0, I_G)$ then for any orthogonal matrix B_n , $B_n Y_n \xrightarrow{d} N(0, I_G)$. Then by the Slutsky lemma and $\tilde{\delta} = \delta_0 + \tilde{H}^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j$ for $\xi_j = (1 - P_{jj})^{-1} \varepsilon_j$ we have

$$\begin{aligned} \bar{V}_n^{-1/2} S_n' (\tilde{\delta} - \delta_0) &= \bar{V}_n^{-1/2} S_n' \tilde{H}^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j = \bar{V}_n^{-1/2} (S_n^{-1} \tilde{H}^{-1} S_n^{-1'})^{-1} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j \\ &= \bar{V}_n^{-1/2} (S_n^{-1} \tilde{H} S_n^{-1'})^{-1} \Gamma_n^{1/2} \Gamma_n^{-1/2} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j \\ &= B_n \Gamma_n^{-1/2} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j + o_p(1) \xrightarrow{d} N(0, I_G). \end{aligned}$$

The conclusion for JIV2 follows by a similar argument for $\xi_i = \varepsilon_i$. Q.E.D.

Proof of Theorem 3: Under the hypotheses of Theorem 3, $r_n/K \rightarrow 0$. Similarly to the proof of Theorem 2 we have,

$$\begin{aligned} \sqrt{r_n/K} S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j &= \sqrt{r_n/K} \sum_{i=1}^n z_i (1 - P_{ii}) \xi_i / \sqrt{n} + \mu_n S_n^{-1} \sum_{i \neq j} U_i P_{ij} \xi_j / \sqrt{K} + o_p(1) \\ &= \mu_n S_n^{-1} \sum_{i \neq j} U_i P_{ij} \xi_j / \sqrt{K} + o_p(1). \end{aligned}$$

Here let

$$\begin{aligned}\Gamma_n &= \text{Var}(\mu_n S_n^{-1} \sum_{i \neq j} U_i P_{ij} \xi_j / \sqrt{K}) \\ &= \mu_n^2 S_n^{-1} \sum_{i \neq j} P_{ij}^2 \left(E[U_i U_i'] E[\xi_j^2] + E[U_i \xi_i] E[U_j' \xi_j] \right) S_n^{-1} / K.\end{aligned}$$

Note that Γ_n is bounded by $\mu_n S_n^{-1}$ bounded, $E[\|U_i\|^2] \leq C$, and $E[\xi_j^2] \leq C$. Let \bar{L}_n be any sequence of bounded matrices with $\lambda_{\min}(\bar{L}_n \Gamma_n \bar{L}_n') \geq C$ and let

$$\bar{Y}_n = (\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} \bar{L}_n \sqrt{r_n/K} S_n^{-1} \sum_{i \neq j} U_i P_{ij} \xi_j.$$

Now let α be a nonzero vector and apply Lemma A2 with $W_{in} = 0$, $\varepsilon_i = \xi_i$, $c_{1n} = 0$, and $c_{2n} = \alpha' (\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} \bar{L}_n \mu_n S_n^{-1}$. We have $c_{2n}' \text{Var} \left(\sum_{i \neq j} U_i P_{ij} \xi_j / \sqrt{K} \right) c_{2n} = \alpha' \alpha > 0$ by construction. Then by the conclusion of Lemma A2 it follows that $\alpha' \bar{Y}_n \xrightarrow{d} N(0, \alpha' \alpha)$. Then by the Cramer-Wold device we have

$$\bar{Y}_n \xrightarrow{d} N(0, I).$$

Consider now the JIV1 estimator, where $\xi_i = \varepsilon_i / (1 - P_{ii})$ and $\Gamma_n = (r_n/K) \bar{\Psi}_n$, so that

$$(\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} \bar{L}_n \sqrt{r_n/K} S_n^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} \varepsilon_j = \bar{Y}_n + o_p(1) \xrightarrow{d} N(0, I_\ell).$$

Let L_n be as specified in the statement of the result such that $\lambda_{\min}(L_n \bar{V}_n^* L_n') \geq C$ and let $\bar{L}_n = L_n \bar{H}_n^{-1}$, so that $L_n \bar{V}_n^* L_n' = \bar{L}_n \Gamma_n \bar{L}_n'$. Note that, $(\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2}$ and $\Gamma_n^{1/2}$ are bounded. By Lemma A6, $(S_n^{-1} \tilde{H} S_n^{-1'})^{-1} = \bar{H}_n^{-1} + o_p(1)$. Therefore, we have

$$(\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} L_n (S_n^{-1} \tilde{H} S_n^{-1'})^{-1} = (\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} L_n (\bar{H}_n^{-1} + o_p(1)) = (\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} \bar{L}_n + o_p(1).$$

Note also that $\sqrt{r_n/K} S_n^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} \varepsilon_j = O_p(1)$. Then we have

$$\begin{aligned}& \left(L_n \bar{H}_n^{-1} \bar{\Psi}_n \bar{H}_n^{-1} L_n' \right)^{-1/2} L_n S_n' (\tilde{\delta} - \delta_0) \\ &= (\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} L_n (S_n^{-1} \tilde{H}^{-1} S_n^{-1'}) \sqrt{r_n/K} S_n^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} \varepsilon_j \\ &= \left[(\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} \bar{L}_n + o_p(1) \right] \sqrt{r_n/K} S_n^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} \varepsilon_j \\ &= (\bar{L}_n \Gamma_n \bar{L}_n')^{-1/2} \bar{L}_n \sqrt{r_n/K} S_n^{-1} \sum_{i \neq j} X_i P_{ij} (1 - P_{jj})^{-1} \varepsilon_j + o_p(1) \\ &= \bar{Y}_n + o_p(1) \xrightarrow{d} N(0, I_\ell).\end{aligned}$$

The conclusion for JIV2 follows by a similar argument for $\xi_i = \varepsilon_i$. Q.E.D.

Next, we turn to the proof of Theorem 4. Let $\tilde{\xi}_i = (y_i - X_i' \tilde{\delta}) / (1 - P_{ii})$ and $\xi_i = \varepsilon_i / (1 - P_{ii})$ for JIV1 and $\hat{\xi}_i = y_i - X_i' \hat{\delta}$ and $\xi_i = \varepsilon_i$ for JIV2. Also, let

$$\begin{aligned}\dot{X}_i &= S_n^{-1} X_i, \hat{\Sigma}_1 = \sum_{i \neq j \neq k} \dot{X}_i P_{ik} \hat{\xi}_k^2 P_{kj} \dot{X}_j', \hat{\Sigma}_2 = \sum_{i \neq j} P_{ij}^2 \left(\dot{X}_i \dot{X}_i' \hat{\xi}_j^2 + \dot{X}_i \hat{\xi}_i \hat{\xi}_j \dot{X}_j' \right), \\ \dot{\Sigma}_1 &= \sum_{i \neq j \neq k} \dot{X}_i P_{ik} \xi_k^2 P_{kj} \dot{X}_j', \dot{\Sigma}_2 = \sum_{i \neq j} P_{ij}^2 \left(\dot{X}_i \dot{X}_i' \xi_j^2 + \dot{X}_i \xi_i \xi_j \dot{X}_j' \right).\end{aligned}$$

LEMMA A7: *If Assumptions 1-6 are satisfied then $\hat{\Sigma}_1 - \dot{\Sigma}_1 = o_p(1)$ and $\hat{\Sigma}_2 - \dot{\Sigma}_2 = o_p(K/r_n)$.*

Proof: To show the first conclusion note that for $\dot{\delta} = \hat{\delta}$ and $X_i^P = X_i / (1 - P_{ii})$ for JIV1 and $\dot{\delta} = \tilde{\delta}$ and $X_i^P = X_i$ for JIV2, $\dot{\delta} \xrightarrow{p} \delta_0$ and

$$\hat{\xi}_i^2 - \xi_i^2 = -2\xi_i X_i^{P'} (\dot{\delta} - \delta_0) + \left[X_i^{P'} (\dot{\delta} - \delta_0) \right]^2.$$

Let η_i be any element $-2\xi_i X_i^{P'}$ or of $X_i^P X_i^{P'}$. Note that S_n / \sqrt{n} is bounded so by the Cauchy-Schwartz inequality, $\|\Upsilon_i\| = \|S_n z_i / \sqrt{n}\| \leq C$. Then

$$E[\eta_i^2] \leq CE[\xi_i^2] + CE[\|X_i\|^2] \leq C + CE[\|\Upsilon_i\|^2] + CE[\|U_i\|^2] \leq C.$$

Let $\hat{\Delta}_n$ denote a sequence of random variables converging to zero in probability. By Lemma A4,

$$\hat{\Delta} \sum_{i \neq j \neq k} \dot{X}_i P_{ik} \eta_k P_{kj} \dot{X}_j' = o_p(1) O_p(1) \xrightarrow{p} 0.$$

From the above expression for $\hat{\xi}_i^2 - \xi_i^2$ we see that $\hat{\Sigma}_1 - \dot{\Sigma}_1$ is a sum of terms of the form $\hat{\Delta} \sum_{i \neq j \neq k} \dot{X}_i P_{ik} \eta_k P_{kj} \dot{X}_j'$, so by the triangle inequality $\hat{\Sigma}_1 - \dot{\Sigma}_1 \xrightarrow{p} 0$.

Let $d_i = C + |\varepsilon_i| + \|U_i\|$, $\hat{A} = (1 + \|\hat{\delta}\|)$ for JIV1, $\hat{A} = (1 + \|\tilde{\delta}\|)$ for JIV2, $\hat{B} = \|\hat{\delta} - \delta_0\|$ for JIV1, and $\hat{B} = \|\tilde{\delta} - \delta_0\|$ for JIV2. By the conclusion of Theorem 1 we have $\hat{A} = O_p(1)$ and $\hat{B} \xrightarrow{p} 0$. Also, by P_{ii} bounded away from 1, $(1 - P_{ii})^{-1} \leq C$, so for both JIV1 and JIV2,

$$\begin{aligned}\|X_i\| &\leq C + \|U_i\| \leq d_i, \|\dot{X}_i\| \leq C \mu_n^{-1} d_i, \\ |\hat{\xi}_i - \xi_i| &\leq C |X_i' (\hat{\delta} - \delta_0)| \leq C d_i \hat{B}, |\hat{\xi}_i| \leq C |X_i' (\delta_0 - \hat{\delta})| + |\xi_i| \leq C d_i \hat{A}, \\ |\hat{\xi}_i^2 - \xi_i^2| &\leq (|\xi_i| + |\hat{\xi}_i|) |\hat{\xi}_i - \xi_i| \leq C d_i (1 + \hat{A}) d_i \hat{B} \leq C d_i^2 \hat{A} \hat{B}, \\ \|\dot{X}_i (\hat{\xi}_i - \xi_i)\| &\leq C \mu_n^{-1} d_i^2 \hat{B}, \|\dot{X}_i \hat{\xi}_i\| \leq C \mu_n^{-1} d_i^2 \hat{A}, \|\dot{X}_i \xi_i\| \leq C \mu_n^{-1} d_i^2.\end{aligned}$$

Also note that

$$E \left[\sum_{i \neq j} P_{ij}^2 d_i^2 d_j^2 r_n^{-1} \right] \leq Cr_n^{-1} \sum_{i,j} P_{ij}^2 = Cr_n^{-1} \sum_i P_{ii} = CK/r_n.$$

so that $\sum_{i \neq j} P_{ij}^2 d_i^2 d_j^2 r_n^{-1} = O_p(K/r_n)$ by the Markov inequality. Then it follows that

$$\left\| \sum_{i \neq j} P_{ij}^2 \left(\dot{X}_i \dot{X}_i' (\hat{\xi}_j^2 - \xi_j^2) \right) \right\| \leq \sum_{i \neq j} P_{ij}^2 \|\dot{X}_i\|^2 |\hat{\xi}_j^2 - \xi_j^2| \leq C\mu_n^{-2} \sum_{i \neq j} P_{ij}^2 d_i^2 d_j^2 \hat{A} \hat{B} = o_p(K/r_n).$$

We also have

$$\begin{aligned} \left\| \sum_{i \neq j} P_{ij}^2 \left(\dot{X}_i \hat{\xi}_i \hat{\xi}_j \dot{X}_j' - \dot{X}_i \xi_i \xi_j \dot{X}_j' \right) \right\| &\leq \sum_{i \neq j} P_{ij}^2 \left(\|\dot{X}_i \hat{\xi}_i\| \|\dot{X}_j (\hat{\xi}_j - \xi_j)\| + \|\dot{X}_j \xi_j\| \|\dot{X}_i (\hat{\xi}_i - \xi_i)\| \right) \\ &\leq Cr_n^{-1} \sum_{i \neq j} P_{ij}^2 d_i^2 d_j^2 \hat{A} \hat{B} = o_p\left(\frac{K}{r_n}\right). \end{aligned}$$

The second conclusion then follows by the triangle inequality. Q.E.D.

LEMMA A8: *If Assumptions 1-6 are satisfied then*

$$\begin{aligned} \dot{\Sigma}_1 &= \sum_{i \neq j \neq k} z_i P_{ik} E[\xi_k^2] P_{kj} z_j' / n + o_p(1), \\ \dot{\Sigma}_2 &= \sum_{i \neq j} P_{ij}^2 z_i z_i' E[\xi_j^2] / n \\ &\quad + S_n^{-1} \sum_{i \neq j} P_{ij}^2 \left(E[U_i U_i'] E[\xi_j^2] + E[U_i \xi_i] E[\xi_j U_j'] \right) S_n^{-1'} + o_p(K/r_n), \end{aligned}$$

Proof: To prove the first conclusion apply Lemma A4 with W_i equal to an element of \dot{X}_i , Y_j equal to an element of \dot{X}_j , and $\eta_k = \xi_k^2$.

Next, note that $\text{Var}(\xi_i^2) \leq C$ and $r_n \leq Cn$, so that for $u_{ki} = e_k' S_n^{-1} U_i$,

$$\begin{aligned} E[(\dot{X}_{ik} \dot{X}_{il})^2] &\leq CE[\dot{X}_{ik}^4 + \dot{X}_{il}^4] \leq C \left\{ z_{ik}^4 / n^2 + E[u_k^4] + z_{il}^4 / n^2 + E[u_l^4] \right\} \leq Cr_n^{-2}, \\ E[(\dot{X}_{ik} \xi_i)^2] &\leq CE[(z_{ik}^2 \xi_i^2 / n + u_{ki}^2 \xi_i^2)] \leq Cn^{-1} + C\mu_n^{-2} \leq Cr_n^{-1}. \end{aligned}$$

Also, we have, for $\Omega_i = E[U_i U_i']$,

$$E[\dot{X}_i \dot{X}_i'] = z_i z_i' / n + S_n^{-1} \Omega_i S_n^{-1'}, E[\dot{X}_i \xi_i] = S_n^{-1} E[U_i \xi_i].$$

Next let W_i be $e_j' \dot{X}_i \dot{X}_i' e_k$ for some j and k , so that

$$\begin{aligned} E[W_i] &= e_j' S_n^{-1} E[U_i U_i'] S_n^{-1'} e_k + z_{ij} z_{ik} / n, |E[W_i]| \leq Cr_n^{-1}. \\ \text{Var}(W_i) &= \text{Var} \left\{ \left(e_j' S_n^{-1} U_i + z_{ij} / \sqrt{n} \right) \left(e_k' S_n^{-1} U_i + z_{ik} / \sqrt{n} \right) \right\} \\ &\leq C/r_n^2 + C/nr_n \leq C/r_n^2. \end{aligned}$$

Also let $Y_i = \xi_i^2$. Then $\sqrt{K}(\bar{\sigma}_{Wn}\bar{\sigma}_{Yn} + \bar{\sigma}_{Wn}\bar{\mu}_{Yn} + \bar{\mu}_{Wn}\bar{\sigma}_{Yn}) \leq C\sqrt{K}/r_n$, so applying Lemma A3 for this W_i and Y_i gives

$$\sum_{i \neq j} P_{ij}^2 \dot{X}_i \dot{X}_i' \xi_j^2 = \sum_{i \neq j} P_{ij}^2 \left(z_i z_i' / n + S_n^{-1} \Omega_i S_n^{-1'} \right) E[\xi_j^2] + O_p(\sqrt{K}/r_n).$$

It follows similarly from Lemma A3 with W_i and Y_i equal to elements of $\dot{X}_i \xi_i$ that

$$\sum_{i \neq j} P_{ij}^2 \dot{X}_i \xi_i \xi_j \dot{X}_j' = S_n^{-1} \sum_{i \neq j} P_{ij}^2 E[U_i \xi_i] E[\xi_j U_j'] S_n^{-1'} + O_p(\sqrt{K}/r_n).$$

Also, by $K \rightarrow \infty$ we have $O_p(\sqrt{K}/r_n) = o_p(K/r_n)$. The second conclusion then follows by T. Q.E.D.

Proof of Theorem 4: Let $\dot{\sigma}_i^2 = E[\xi_i^2]$ and $\bar{z}_i = \sum_j P_{ij} z_j = e_i' P z$. Then

$$\begin{aligned} \sum_{i \neq j \neq k} z_i P_{ik} \dot{\sigma}_k^2 P_{kj} z_j' / n &= \sum_i \sum_{j \neq i} \sum_{k \notin \{i, j\}} z_i P_{ik} \dot{\sigma}_k^2 P_{kj} z_j' / n \\ &= \sum_i \sum_{j \neq i} \left(\sum_k z_i P_{ik} \dot{\sigma}_k^2 P_{kj} z_j' - z_i P_{ii} \dot{\sigma}_i^2 P_{ij} z_j' - z_i P_{ij} \dot{\sigma}_j^2 P_{jj} z_j' \right) / n \\ &= \left(\sum_k \bar{z}_k \dot{\sigma}_k^2 z_k' - \sum_{i, k} P_{ik}^2 z_i z_i' \dot{\sigma}_k^2 - \sum_i z_i P_{ii} \dot{\sigma}_i^2 z_i' + \sum_i z_i P_{ii} \dot{\sigma}_i^2 P_{ii} z_i' \right. \\ &\quad \left. - \sum_j \bar{z}_j \dot{\sigma}_j^2 P_{jj} z_j' + \sum_i z_j P_{jj} \dot{\sigma}_j^2 P_{jj} z_j' \right) / n \\ &= \sum_i \dot{\sigma}_i^2 \left(\bar{z}_i z_i' - P_{ii} z_i z_i' - P_{ii} \bar{z}_i z_i' + P_{ii}^2 z_i z_i' \right) / n - \sum_{i \neq j} P_{ij}^2 z_i z_i' \dot{\sigma}_j^2 / n. \end{aligned}$$

Also, as shown above, Assumption 4 implies that $\sum_i \|z_i - \bar{z}_i\|^2 / n \leq z'(I - P)z / n \rightarrow 0$. Then by $\dot{\sigma}_i^2$ and P_{ii} bounded we have

$$\begin{aligned} \left\| \sum_i \dot{\sigma}_i^2 (\bar{z}_i z_i' - z_i z_i') / n \right\| &\leq \sum_i \dot{\sigma}_i^2 (2 \|z_i\| \|z_i - \bar{z}_i\| + \|z_i - \bar{z}_i\|^2) / n \\ &\leq C (\sum_i \|z_i\|^2 / n)^{1/2} (\sum_i \|z_i - \bar{z}_i\|^2 / n)^{1/2} + C \sum_i \|z_i - \bar{z}_i\|^2 / n \rightarrow 0, \\ \left\| \sum_i \dot{\sigma}_i^2 P_{ii} (z_i z_i' - z_i z_i') / n \right\| &\leq (\sum_i \dot{\sigma}_i^4 P_{ii}^2 \|z_i\|^2 / n)^{1/2} (\sum_i \|z_i - \bar{z}_i\|^2 / n)^{1/2} \rightarrow 0. \end{aligned}$$

It follows that

$$\sum_{i \neq j \neq k} z_i P_{ik} \dot{\sigma}_k^2 P_{kj} z_j' / n = \sum_i \dot{\sigma}_i^2 (1 - P_{ii})^2 z_i z_i' / n + o(1) - \sum_{i \neq j} P_{ij}^2 z_i z_i' \dot{\sigma}_j^2 / n.$$

It then follows by Lemmas A7 and A8 and the triangle inequality that

$$\begin{aligned}
\hat{\Sigma}_1 + \hat{\Sigma}_2 &= \sum_{i \neq j \neq k} z_i P_{ik} \dot{\sigma}_k^2 P_{kj} z'_j / n + \sum_{i \neq j} P_{ij}^2 z_i z'_i \dot{\sigma}_j^2 / n \\
&\quad + S_n^{-1} \sum_{i \neq j} P_{ij}^2 \left(E[U_i U_i'] \dot{\sigma}_j^2 + E[U_i \xi_i] E[\xi_j U_j'] \right) S_n^{-1'} + o_p(1) + o_p(K/r_n) \\
&= \sum_i \dot{\sigma}_i^2 (1 - P_{ii})^2 z_i z'_i / n \\
&\quad + S_n^{-1} \sum_{i \neq j} P_{ij}^2 \left(E[U_i U_i'] \dot{\sigma}_j^2 + E[U_i \xi_i] E[\xi_j U_j'] \right) S_n^{-1'} + o_p(1) + o_p(K/r_n).
\end{aligned}$$

Then for JIV1, where $\xi_i = \varepsilon_i / (1 - P_{ii})$ and $\dot{\sigma}_i^2 = \sigma_i^2 / (1 - P_{ii})^2$, we have

$$\hat{\Sigma}_1 + \hat{\Sigma}_2 = \bar{\Omega}_n + \bar{\Psi}_n + o_p(1) + o_p(K/r_n).$$

For JIV2, where $\xi_i = \varepsilon_i$ and $\dot{\sigma}_i^2 = \sigma_i^2$, we have

$$\hat{\Sigma}_1 + \hat{\Sigma}_2 = \Omega_n + \Psi_n + o_p(1) + o_p(K/r_n).$$

Consider the case where K/r_n is bounded, implying $o_p(K/r_n) = o_p(1)$. Then by boundedness of \bar{H}_n^{-1} , $\bar{\Omega}_n + \bar{\Psi}_n$, H_n^{-1} , and $\Omega_n + \Psi_n$,

$$\begin{aligned}
S_n' \tilde{V} S_n &= \left(S_n^{-1} \tilde{H} S_n^{-1'} \right)^{-1} \left(\hat{\Sigma}_1 + \hat{\Sigma}_2 \right) \left(S_n^{-1} \tilde{H}' S_n^{-1'} \right)^{-1} \\
&= \left(\bar{H}_n^{-1} + o_p(1) \right) \left(\bar{\Omega}_n + \bar{\Psi}_n + o_p(1) \right) \left(\bar{H}_n^{-1} + o_p(1) \right) = \bar{V}_n + o_p(1). \\
S_n' \hat{V} S_n &= V_n + o_p(1),
\end{aligned}$$

giving the first conclusion.

Next, consider the case where $K/r_n \rightarrow \infty$. Then for JIV1, where $\xi_i = \varepsilon_i / (1 - P_{ii})$ and $\dot{\sigma}_i^2 = \sigma_i^2 / (1 - P_{ii})^2$, boundedness of $\bar{\Omega}_n$ in n implies we have

$$\begin{aligned}
(r_n/K) \left(\hat{\Sigma}_1 + \hat{\Sigma}_2 \right) &= (r_n/K) \bar{\Omega}_n + (r_n/K) \bar{\Psi}_n + (r_n/K) o_p(1) + o_p(1) \\
&= (r_n/K) \bar{\Psi}_n + o_p(1).
\end{aligned}$$

For JIV2, where $\xi_i = \varepsilon_i$ and $\dot{\sigma}_i^2 = \sigma_i^2$, we have

$$\begin{aligned}
(r_n/K) \left(\hat{\Sigma}_1 + \hat{\Sigma}_2 \right) &= (r_n/K) \Omega_n + (r_n/K) \Psi_n + (r_n/K) o_p(1) + o_p(1) \\
&= (r_n/K) \Psi_n + o_p(1).
\end{aligned}$$

Then by boundedness of \bar{H}_n^{-1} , $(r/K_n)\bar{\Psi}_n$, H_n^{-1} , and $(r/K_n)\Psi_n$,

$$\begin{aligned} S'_n \tilde{V} S_n &= \left(S_n^{-1} \tilde{H} S_n^{-1'} \right)^{-1} \left(\hat{\Sigma}_1 + \hat{\Sigma}_2 \right) \left(S_n^{-1} \tilde{H}' S_n^{-1'} \right)^{-1} \\ &= \left(\bar{H}_n^{-1} + o_p(1) \right) \left(r_n \bar{\Psi}_n / K_n + o_p(1) \right) \left(\bar{H}_n^{-1} + o_p(1) \right) = \bar{V}_n^* + o_p(1). \\ S'_n \hat{V} S_n &= V_n^* + o_p(1), \end{aligned}$$

giving the second conclusion. Q.E.D.

Proof of Theorem 5: An expansion gives

$$a(\hat{\delta}) - a(\delta_0) = \bar{A}(\hat{\delta} - \delta_0)$$

for $\bar{A} = \partial a(\bar{\delta}) / \partial \delta$ where $\bar{\delta}$ lies on the line joining $\hat{\delta}$ and δ_0 and actually differs from element to element of $a(\delta)$. It follows by $\hat{\delta} \xrightarrow{p} \delta_0$ that $\bar{\delta} \xrightarrow{p} \delta_0$, so that by condition iii), $B_n \hat{A} S_n^{-1'} = B_n A S_n^{-1'} + o_p(1)$. Then multiplying by B_n and using the conclusion of Theorem 4 we have

$$\begin{aligned} & \left(\hat{A} \hat{V} \hat{A}' \right)^{-1/2} \left[a(\hat{\delta}) - a(\delta_0) \right] \\ &= \left(B_n \hat{A} S_n^{-1'} S'_n \hat{V} S_n S_n^{-1} \hat{A}' B_n' \right)^{-1/2} B_n \bar{A} S_n^{-1'} S'_n (\hat{\delta} - \delta_0) \\ &= \left[\left(B_n A S_n^{-1} + o_p(1) \right) \left(\bar{V}_n + o_p(1) \right) \left(S_n^{-1'} A B_n' + o_p(1) \right) \right]^{-1/2} \\ & \quad \times \left(B_n A S_n^{-1'} + o_p(1) \right) S'_n (\hat{\delta} - \delta_0) \\ &= \left(B_n A S_n^{-1} \bar{V}_n S_n^{-1'} A' B_n' \right)^{-1/2} B_n A S_n^{-1'} S'_n (\hat{\delta} - \delta_0) + o_p(1) \\ &= \left(B_n A S_n^{-1} \bar{V}_n S_n^{-1'} A' B_n' \right)^{-1/2} B_n A S_n^{-1} \bar{V}_n^{1/2} \bar{V}_n^{-1/2} S'_n (\hat{\delta} - \delta_0) + o_p(1) \\ &= \left(F_n F_n' \right)^{-1/2} F_n \bar{Y}_n + o_p(1) \end{aligned}$$

for $F_n = B_n A S_n^{-1} \bar{V}_n^{1/2}$ and $\bar{Y}_n = \bar{V}_n^{-1/2} S'_n (\hat{\delta} - \delta_0)$. By Theorem 2, $\bar{Y}_n \xrightarrow{d} N(0, I_G)$. Then since $L_n = (F_n F_n')^{-1/2} F_n$ satisfies $L_n L_n' = I$, it follows from the Slutsky Theorem and standard convergence in distribution results that

$$\left(\hat{A} \hat{V} \hat{A}' \right)^{-1/2} \left[a(\hat{\delta}) - a(\delta_0) \right] = L_n \bar{Y}_n + o_p(1) \xrightarrow{d} N(0, I),$$

giving the conclusion. Q.E.D.

Proof of Corollary 6: Let $a(\delta) = c'\delta$, so that $\bar{A} = A = c'$. Note that condition i) of Theorem 5 is satisfied. Let $B_n = b_n$. Then $B_n A S_n^{-1'} = b_n c' S_n^{-1'}$ is bounded by hypothesis so condition ii)

of Theorem 5 is satisfied. Also, $B_n(\bar{A} - A)S_n^{-1'} = 0$ so condition iii) of Theorem 5 is satisfied. If K/r_n is bounded then by hypothesis, $\lambda_{\min}(B_nAS_n^{-1'}\bar{V}_nS_n^{-1}A'B'_n) = b_n^2c'S_n^{-1'}\bar{V}_nS_n^{-1}c \geq C$ or if $K/r_n \rightarrow \infty$ then $\lambda_{\min}(B_nAS_n^{-1'}\bar{V}_n^*S_n^{-1}A'B'_n) = b_n^2c'S_n^{-1'}\bar{V}_n^*S_n^{-1}c \geq C$, giving the first conclusion. The second conclusion follows similarly. Q.E.D.

6 Appendix B - Proofs of Lemmas A2 and A4

We first give a series of Lemmas that will be useful for the proofs of Lemmas A2 and A4.

LEMMA B1: For any subset \mathcal{I}_2 of the set $\{(i, j)_{i,j=1}^n\}$ and any subset \mathcal{I}_3 of $\{(i, j, k)_{i,j,k=1}^n\}$, for all n with probability one as $n \rightarrow \infty$. (a) $\sum_{\mathcal{I}_2} P_{ij}^4 \leq K$; (b) $\sum_{\mathcal{I}_3} P_{ij}^2 P_{jk}^2 \leq K$; and (c) for constants $\{\bar{a}_i, \bar{b}_i, \bar{c}_i\}_{i=1}^n$, $\left| \sum_{\mathcal{I}_3} P_{ij}^2 P_{ik} P_{jk} \bar{a}_i \bar{b}_j \bar{c}_k \right| \leq K \max_{i \leq n} |\bar{a}_i| \max_{i \leq n} |\bar{b}_j| \max_{i \leq n} |\bar{c}_i|$.

Proof: By Assumption 2, $Z'Z$ is nonsingular with probability one for n sufficiently large. Also, by P idempotent, $\text{rank}(P) = \text{tr}(P) = K$, $0 \leq P_{ii} \leq 1$, and $\sum_{j=1}^n P_{ij}^2 = P_{ii}$. Therefore, for $D_n = \max_{i \leq n} |\bar{a}_i| \max_{i \leq n} |\bar{b}_i| \max_{i \leq n} |\bar{c}_i|$,

$$\begin{aligned} \sum_{\mathcal{I}_2} P_{ij}^4 &\leq \sum_{i,j=1}^n P_{ij}^4 \leq \sum_{i,j=1}^n P_{ij}^2 = \sum_{i=1}^n P_{ii} = K, \\ \sum_{\mathcal{I}_3} P_{ij}^2 P_{jk}^2 &\leq \sum_{i,j,k=1}^n P_{ij}^2 P_{jk}^2 = \sum_{j=1}^n \left(\sum_{i=1}^n P_{ij}^2 \right) \left(\sum_{k=1}^n P_{jk}^2 \right) \\ &= \sum_{j=1}^n P_{jj}^2 \leq \sum_{j=1}^n P_{jj} = K, \\ \left| \sum_{\mathcal{I}_3} P_{ij}^2 P_{ik} P_{jk} \bar{a}_i \bar{b}_j \bar{c}_k \right| &\leq D_n \sum_{i,j} P_{ij}^2 \sum_k |P_{ik} P_{jk}| \leq D_n \sum_{i,j} P_{ij}^2 \sqrt{\sum_k P_{ik}^2} \sqrt{\sum_k P_{jk}^2} \\ &\leq D_n \sum_{i,j} P_{ij}^2 \sqrt{P_{ii} P_{jj}} \leq D_n \sum_{i,j} P_{ij}^2 = D_n K. \text{Q.E.D.} \end{aligned}$$

For the next result let $S_n = \sum_{i < j < k < l} (P_{ik} P_{jk} P_{il} P_{jl} + P_{ij} P_{jk} P_{il} P_{kl} + P_{ij} P_{ik} P_{jl} P_{kl})$.

LEMMA B2: If Assumption 2 is satisfied then for $D = \text{diag}(P_{11}, \dots, P_{nn})$, a) $\text{tr}[(P - D)^4] \leq CK$; b) $\left| \sum_{i < j < k < l} P_{ik} P_{jk} P_{il} P_{jl} \right| \leq CK$, and c) $|S_n| \leq CK$.

Proof: To show part (a), note that

$$(P - D)^4 = (P - PD - DP + D^2)^2 = P - PD - PDP + PD^2 - PDP + PDPD + PD^2P - PD^3 - DP + DPD + DPDP - DPD^2 + D^2P - D^2PD - D^3P + D^4.$$

Note that $\text{tr}(A') = \text{tr}(A)$ and $\text{tr}(AB) = \text{tr}(BA)$ for any square matrices A and B . Then

$$\text{tr} \left[(P - D)^4 \right] = \text{tr}(P) - 4\text{tr}(PD) + 4\text{tr}(PD^2) + 2\text{tr}(PDPD) - 4\text{tr}(PD^3) + \text{tr}(D^4).$$

By $0 \leq P_{ii} \leq 1$ we have $D^j \leq I$ for any integer and $\text{tr}(PD^j) = \text{tr}(PD^jP) \leq \text{tr}(P) = K$. Also, $\text{tr}(PDPD) = \text{tr}(PDPDP) \leq \text{tr}(PD^2P) \leq \text{tr}(P) = K$ and $\text{tr}(D^4) = \sum_i P_{ii}^4 \leq K$. Therefore, by T we have

$$\left| \text{tr} \left[(P - D)^4 \right] \right| \leq 16K,$$

giving conclusion a).

Next, let $L_{ij} = P_{ij}1(i > j)$ be the matrix that is the lower triangle of P and zeros elsewhere.

Then $P = L + L' + D$ so

$$\begin{aligned} (P - D)^4 &= (L + L')^4 = (L^2 + LL' + L'L + L'^2)^2 \\ &= L^4 + L^2LL' + L^2L'L + L^2L'^2 + LL'L^2 + LL'LL' + LL'L'L + LL'^3 \\ &\quad + L'LL^2 + L'LLL' + L'LL'L + L'LL'^2 + L'^2L^2 + L'^2LL' + L'^2L'L + L'^4 \end{aligned}$$

Note that for an integer j ; $[(L')^j]' = L^j$. Then using $\text{tr}(AB) = \text{tr}(BA)$ and $\text{tr}(A') = \text{tr}(A)$,

$$\begin{aligned} \text{tr}((P - D)^4) &= 2\text{tr}(L^4) + 8\text{tr}(L^3L') + 4\text{tr}(L^2L'^2) \\ &\quad + 2\text{tr}(L'LL'L) \end{aligned}$$

Next, compute each of the terms. Note that

$$\text{tr}(L^4) = \sum_{i,j,k,\ell} P_{ij}1(i > j)P_{jk}1(j > k)P_{k,\ell}1(k > \ell)P_{\ell,i}1(\ell > i) = 0.$$

$$\begin{aligned} \text{tr}(L^3L') &= \sum_{i,j,k,\ell} P_{ij}1(i > j)P_{jk}1(j > k)P_{k\ell}1(k > \ell)P_{\ell i}1(i > \ell) = \sum_{i>j>k>\ell} P_{ij}P_{jk}P_{k\ell}P_{\ell i} \\ &= \sum_{\ell < k < j < i} P_{ij}P_{jk}P_{k\ell}P_{\ell i} = \sum_{i < j < k < \ell} P_{\ell k}P_{kj}P_{ji}P_{i\ell} = \sum_{i < j < k < \ell} P_{ij}P_{jk}P_{k\ell}P_{\ell i} \end{aligned}$$

$$\begin{aligned}
\text{tr}(L^2 L'^2) &= \sum_{i,j,k,\ell} P_{ij}1(i > j)P_{jk}1(j > k)P_{k\ell}1(\ell > k)P_{\ell i}1(i > \ell) = \sum_{i>j>k, i>\ell>k} P_{ij}P_{jk}P_{k\ell}P_{\ell i} \\
&= \sum_{i>j=\ell>k} P_{ij}P_{jk}P_{k\ell}P_{\ell i} + \sum_{i>j>\ell>k} P_{ij}P_{jk}P_{k\ell}P_{\ell i} + \sum_{i>\ell>j>k} P_{ij}P_{jk}P_{k\ell}P_{\ell i} \\
&= \sum_{i>j>k} P_{ij}P_{jk}P_{kj}P_{ji} + \sum_{i<j<k<\ell} (P_{\ell k}P_{ki}P_{ij}P_{j\ell} + P_{\ell j}P_{ji}P_{ik}P_{k\ell}) \\
&= \sum_{i<j<k} P_{ij}^2 P_{jk}^2 + 2 \sum_{i<j<k<\ell} P_{ik}P_{k\ell}P_{\ell j}P_{ji}
\end{aligned}$$

$$\begin{aligned}
\text{tr}(LL'LL') &= \sum_{i,j,k,\ell} P_{ij}1(i > j)P_{jk}(k > j)P_{k\ell}1(k > \ell)P_{\ell i}1(i > \ell) \\
&= \sum_{j<i} P_{ij}P_{ji}P_{ij}P_{ji} + \sum_{j<k<i} P_{ij}P_{jk}P_{kj}P_{ji} + \sum_{j<i<k} P_{ij}P_{jk}P_{kj}P_{ji} + \sum_{j<\ell<i} P_{ij}P_{ji}P_{i\ell}P_{\ell i} \\
&\quad + \sum_{\ell<j<i} P_{ij}P_{ji}P_{i\ell}P_{\ell i} + \left(\sum_{\ell<j<k<i} + \sum_{j<\ell<k<i} + \sum_{\ell<j<i<k} + \sum_{j<\ell<i<k} \right) P_{ij}P_{jk}P_{k\ell}P_{\ell i} \\
&= \sum_{i<j} P_{ij}^4 + 2 \sum_{i<j<k} (P_{ij}^2 P_{ik}^2 + P_{ik}^2 P_{jk}^2) + 4 \sum_{i<j<k<\ell} P_{ik}P_{kj}P_{j\ell}P_{\ell i}
\end{aligned}$$

Summing up gives the result

$$\text{tr}((P - D)^4) = 2 \sum_{i<j} P_{ij}^4 + 4 \sum_{i<j<k} (P_{ij}^2 P_{jk}^2 + P_{ik}^2 P_{jk}^2 + P_{ij}^2 P_{ik}^2) + 8S_n.$$

Then by the triangle inequality and Lemma B1 we have

$$|S_n| \leq (1/4) \sum_{i<j} P_{ij}^4 + 1/2 \sum_{i<j<k} (P_{ij}^2 P_{jk}^2 + P_{ik}^2 P_{jk}^2 + P_{ij}^2 P_{ik}^2) + (1/8) \text{tr}((P - D)^4) \leq CK,$$

giving part c). That is, $S_n = O(K)$.

To show part (b), take $\{\varepsilon_i\}$ to be a sequence of i.i.d. random variables with mean 0 and variance 1 and where ε_i and Z are independent for all i and n . Define the random quantities

$$\begin{aligned}
\Delta_1 &= \sum_{i < j < k} [P_{ij}P_{ik}\varepsilon_j\varepsilon_k + P_{ij}P_{jk}\varepsilon_i\varepsilon_k + P_{ik}P_{jk}\varepsilon_i\varepsilon_j], \\
\Delta_2 &= \sum_{i < j < k} [P_{ij}P_{ik}\varepsilon_j\varepsilon_k + P_{ij}P_{jk}\varepsilon_i\varepsilon_k], \Delta_3 = \sum_{i < j < k} P_{ik}P_{jk}\varepsilon_i\varepsilon_j.
\end{aligned}$$

Note that by Lemma A1,

$$\begin{aligned}
E[\Delta_3^2] &= E\left[\sum_{i<j<k} P_{ik}P_{jk}\varepsilon_i\varepsilon_j \sum_{\ell<m<q} P_{\ell q}P_{mq}\varepsilon_\ell\varepsilon_m \right] \\
&= \sum_{i<j<\{k,\ell\}} P_{ik}P_{jk}P_{i\ell}P_{j\ell} = \sum_{i<j<k} (P_{ik})^2 (P_{jk})^2 + 2 \sum_{i<j<k<\ell} P_{ik}P_{jk}P_{i\ell}P_{j\ell} \\
&= O(K) + 2 \sum_{i<j<k<\ell} P_{ik}P_{jk}P_{i\ell}P_{j\ell}.
\end{aligned}$$

Also, note that

$$\begin{aligned} E[\Delta_2\Delta_3] &= E\left[\sum_{i<j<k} (P_{ij}P_{ik}\varepsilon_j\varepsilon_k + P_{ij}P_{jk}\varepsilon_i\varepsilon_k) \sum_{\ell<m<q} P_{\ell q}P_{mq}\varepsilon_\ell\varepsilon_m\right] \\ &= \sum_{i<j<k<\ell} P_{ij}P_{ik}P_{j\ell}P_{k\ell} + \sum_{i<j<k<\ell} P_{ij}P_{jk}P_{i\ell}P_{k\ell}, \end{aligned}$$

and

$$\begin{aligned} E(\Delta_2^2) &= E\left[\left(\sum_{i<j<k} P_{ij}P_{ik}\varepsilon_j\varepsilon_k + P_{ij}P_{jk}\varepsilon_i\varepsilon_k\right) \times \left(\sum_{\ell<m<q} P_{\ell m}P_{\ell q}\varepsilon_m\varepsilon_q + P_{\ell m}P_{mq}\varepsilon_\ell\varepsilon_q\right)\right] \\ &= \sum_{\{i,\ell\}<j<k} P_{ij}P_{ik}P_{\ell j}P_{\ell k} + \sum_{i<\{j,m\}<k} P_{ij}P_{jk}P_{im}P_{mk} \\ &\quad + \sum_{i<j<m<k} P_{ij}P_{ik}P_{jm}P_{mk} + \sum_{\ell<i<j<k} P_{ij}P_{jk}P_{i\ell}P_{\ell k} \\ &= \sum_{i<j<k} P_{ij}^2P_{ik}^2 + \sum_{i<j<k} P_{ij}^2P_{jk}^2 + 2 \sum_{i<\ell<j<k} P_{ij}P_{ik}P_{\ell j}P_{\ell k} + 2 \sum_{i<j<m<k} P_{ij}P_{jk}P_{im}P_{mk} \\ &\quad + \sum_{i<j<k<\ell} P_{ij}P_{i\ell}P_{jk}P_{k\ell} + \sum_{i<j<k<\ell} P_{jk}P_{k\ell}P_{ij}P_{i\ell} \\ &= \sum_{i<j<k} P_{ij}^2P_{ik}^2 + \sum_{i<j<k} P_{ij}^2P_{jk}^2 + 2S_n = O(K). \end{aligned}$$

Since $\Delta_1 = \Delta_2 + \Delta_3$, it follows that

$$E[\Delta_1^2] = E[\Delta_2^2] + E[\Delta_3^2] + 2E[\Delta_2\Delta_3] = O(K) + 2S_n = O(K).$$

Therefore, by T, the expression for $E[\Delta_3^2]$ given above, and $\Delta_3 = \Delta_1 - \Delta_2$,

$$\begin{aligned} \left| \sum_{i<j<k<\ell} P_{ik}P_{jk}P_{i\ell}P_{j\ell} \right| &\leq E[\Delta_3^2] + O(K) \leq E[(\Delta_1 - \Delta_2)^2] + O(K) \\ &\leq 2E[\Delta_1^2] + 2E[\Delta_2^2] + O(K) \leq O(K). \end{aligned} \text{Q.E.D.}$$

LEMMA B3: Let L be the lower triangular matrix with $L_{ij} = P_{ij}1(i > j)$. Then, under Assumption 2, $\|LL'\| \leq C\sqrt{K}$ as $n \rightarrow \infty$, where $\|A\| = [\text{Tr}(A'A)]^{\frac{1}{2}}$.

Proof: From the proof of Lemma B2 and by Lemma B1 and Lemma B2 b) we have

$$\begin{aligned} \|LL'\|^2 &= \text{tr}(LL'LL') = \sum_{i<j} P_{ij}^4 + 2 \sum_{i<j<k} (P_{ij}^2P_{ik}^2 + P_{ik}^2P_{jk}^2) + 4 \sum_{i<j<k<\ell} P_{ik}P_{kj}P_{j\ell}P_{i\ell} \\ &\leq C(K + \left| \sum_{i<j<k<\ell} P_{ik}P_{kj}P_{j\ell}P_{i\ell} \right|) \leq CK. \end{aligned}$$

Taking square roots gives the answer. Q.E.D.

Proof of Lemma A2: Let $b_{1n} = c_{1n}\Xi_n^{-1/2}$ and $b_{2n} = c_{2n}\Xi_n^{-1/2}$, and note that these are bounded in n since Ξ_n is bounded away from zero by hypothesis. Let $w_{1n} = b'_{1n}W_{1n}$ and $u_i = b'_{2n}U_i$, where we suppress the n subscript on u_i for notational convenience. Then

$$Y_n = w_{1n} + \sum_{i=2}^n y_{in}, y_{in} = w_{in} + \bar{y}_{in}, \bar{y}_{in} = \sum_{j<i} (u_j P_{ij} \varepsilon_i + u_i P_{ij} \varepsilon_j) / \sqrt{K}.$$

Also, $E[\|w_{1n}\|^4] \leq \sum_i E[\|w_{in}\|^4] \leq C \sum_i E[\|W_{in}\|^4] \rightarrow 0$, so that $w_{1n} \xrightarrow{p} 0$. Hence

$$Y_n = \sum_{i=2}^n y_{in} + o_p(1).$$

Note that y_{in} is a martingale difference, so that we can apply the martingale central limit theorem. Note also that $E[w_{in}\bar{y}_{jn}] = 0$, for all i and j and that

$$\begin{aligned} E[(\bar{y}_{in})^2] &= \sum_{j<i} \sum_{k<i} E[(u_j P_{ij} \varepsilon_i + u_i P_{ij} \varepsilon_j)(u_k P_{ik} \varepsilon_i + u_i P_{ik} \varepsilon_k)] / K \\ &= \sum_{j<i} P_{ij}^2 (E[u_j^2] E[\varepsilon_i^2] + E[u_i^2] E[\varepsilon_j^2] + 2E[u_i \varepsilon_i] E[u_j \varepsilon_j]) / K. \end{aligned}$$

Thus,

$$\begin{aligned} s_n^2 &= E \left[\left(\sum_{i=2}^n y_{in} \right)^2 \right] = \sum_{i=2}^n (E[w_{in}^2] + E[\bar{y}_{in}^2]) \\ &= b'_{1n} D_n b_{1n} - E[w_{1n}^2] + \sum_{i \neq j} P_{ij}^2 (E[u_j^2] E[\varepsilon_i^2] + E[u_i \varepsilon_i] E[u_j \varepsilon_j]) / K \\ &= b'_{1n} D_n b_{1n} + b'_{2n} \bar{\Sigma}_n b_{2n} + o(1) = 1 + o(1) \rightarrow 1. \end{aligned}$$

Thus, s_n^2 is bounded and bounded away from zero. Also,

$$\sum_{i=2}^n E[y_{in}^4] \leq C \sum_{i=2}^n E[\|W_{in}\|^4] + C \sum_{i=2}^n E[\bar{y}_{in}^4].$$

By condition iv), $\sum_{i=2}^n E[\|W_{in}\|^4] \rightarrow 0$. Let $\bar{y}_{in}^\varepsilon = \sum_{j<i} u_j P_{ij} \varepsilon_i / \sqrt{n}$ and $\bar{y}_{in}^u = \sum_{j<i} u_i P_{ij} \varepsilon_j / \sqrt{K}$.

By $|P_{ij}| < 1$ and $\sum_j P_{ij}^2 = P_{ii}$,

$$\begin{aligned} \sum_{i=2}^n E[(\bar{y}_{in}^\varepsilon)^4] &\leq \frac{C}{K^2} \sum_{i=2}^n \sum_{j,k,\ell,m<i} P_{ij} P_{ik} P_{i\ell} P_{im} E[\varepsilon_i^4] E[u_j u_k u_\ell u_m] \\ &\leq \frac{C}{K^2} \sum_{i=2}^n \left(\sum_{j<i} P_{ij}^4 + \sum_{j,k<i} P_{ij}^2 P_{ik}^2 \right) \leq CK/K^2 \rightarrow 0. \end{aligned}$$

Similarly we have $\sum_{i=2}^n E[(\bar{y}_{in}^u)^4] \rightarrow 0$, so that

$$\sum_{i=2}^n E[\bar{y}_{in}^4] \leq C \sum_{i=2}^n \left\{ E[(\bar{y}_{in}^\varepsilon)^4] + E[(\bar{y}_{in}^u)^4] \right\} \rightarrow 0.$$

Then by T we have

$$\sum_{i=2}^n E[y_{in}^4] \rightarrow 0.$$

To apply the martingale central limit theorem it now suffices to show that for $Z_i = (W_{in}, U_i, \varepsilon_i)$,

$$\sum_{i=2}^n E[y_{in}^2 | Z_1, \dots, Z_{i-1}] - s_n^2 \xrightarrow{p} 0. \quad (2)$$

Note first that by independence of W_{1n}, \dots, W_{nn} ,

$$E[w_{in}^2 | Z_1, \dots, Z_{i-1}] = E[w_{in}^2].$$

Also, $E[w_{in}\bar{y}_{in}] = 0$ and

$$E[w_{in}\bar{y}_{in} | Z_1, \dots, Z_{i-1}] = E[w_{in}u_i] \sum_{j<i} P_{ij}\varepsilon_j/\sqrt{K} + E[w_{in}\varepsilon_i] \sum_{j<i} P_{ij}u_j/\sqrt{K}.$$

Let $\delta_i = E[w_{in}u_i]$ and consider the first term, $\delta_i \sum_{j<i} P_{ij}\varepsilon_j/\sqrt{K}$. Let \bar{P} be the upper triangular matrix with $\bar{P}_{ij} = P_{ij}$ for $j > i$ and $\bar{P}_{ij} = 0$, $j \leq i$, and let $\delta = (\delta_1, \dots, \delta_n)$. Then, $\sum_{i=2}^n \sum_{j<i} \delta_i P_{ij}\varepsilon_j/\sqrt{K} = \delta' \bar{P}' \varepsilon/\sqrt{K}$. By CS $\delta' \delta = \sum_{i=1}^n (E[w_{in}u_i])^2 \leq \sum_{i=1}^n E[w_{in}^2] E[u_i^2] \leq C$. By Lemma B3, $\|\bar{P}' \bar{P}\| \leq \sqrt{K}$. It then follows given $E[u_j^2] \leq C$ that

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

so that $\delta' \bar{P}' \varepsilon/\sqrt{K} \xrightarrow{p} 0$ by M. Similarly, we have $\sum_{i=2}^n E[w_{in}u_i] \sum_{j<i} P_{ij}\varepsilon_j/\sqrt{K} \rightarrow 0$. Therefore, it follows by T that $\sum_{i=2}^n E[w_{in}\bar{y}_{in} | Z_1, \dots, Z_{i-1}] \xrightarrow{p} 0$. To finish showing that eq. (2) is satisfied it only remains to show that

$$\sum_{i=2}^n \left(E[\bar{y}_{in}^2 | Z_1, \dots, Z_{i-1}] - E[\bar{y}_{in}^2] \right) \xrightarrow{p} 0. \quad (3)$$

Now, let $\sigma_i^2 = E[\varepsilon_i^2]$, $\omega_i^2 = \omega_{in}^2 = E[u_i^2]$, and $\gamma_i = \gamma_{in} = E[u_i \varepsilon_i]$, where again we suppress the

subscript n on ω_i^2 and γ_i when doing so is not likely to lead to confusion. We can write

$$\begin{aligned}
& \sum_{i=2}^n \left(E \left[\bar{y}_{in}^2 \mid Z_1, \dots, Z_{i-1} \right] - E[\bar{y}_{in}^2] \right) \\
&= \sum_{j < i} \omega_i^2 P_{ij}^2 (\varepsilon_j^2 - \sigma_j^2) / K + 2 \sum_{j < k < i} \omega_i^2 P_{ij} P_{ik} \varepsilon_j \varepsilon_k / K \\
&\quad + \sum_{j < i} \sigma_i^2 P_{ij}^2 (u_j^2 - \omega_j^2) / K + 2 \sum_{j < k < i} \sigma_i^2 P_{ij} P_{ik} u_j u_k / K \\
&\quad + 2 \sum_{j < i} \gamma_i P_{ij}^2 (u_j \varepsilon_j - \gamma_j) / K + 2 \sum_{j < k < i} \gamma_i P_{ij} P_{ik} (u_j \varepsilon_k + u_k \varepsilon_j) / K.
\end{aligned}$$

Note first that by Lemma B1,

$$\begin{aligned}
& E \left[\left(\sum_{j < i} \gamma_i P_{ij}^2 (u_j \varepsilon_j - \gamma_j) / K \right)^2 \right] = \sum_{j < i} \sum_{\ell < k} \gamma_i P_{ij}^2 \gamma_k P_{k\ell}^2 E [(u_j \varepsilon_j - \gamma_j) (u_\ell \varepsilon_\ell - \gamma_\ell)] / K^2 \\
&= \sum_{j < i, j < k} \gamma_i P_{ij}^2 \gamma_k P_{kj}^2 E [(u_j \varepsilon_j - \gamma_j)^2] / K^2 \leq C \sum_{j < i, j < k} P_{ij}^2 P_{kj}^2 / K^2 \leq \frac{CK}{K^2} \rightarrow 0.
\end{aligned}$$

Similar arguments can also be applied to show that

$$E \left[\left(\sum_{j < i} \omega_i^2 P_{ij}^2 (\varepsilon_j^2 - \sigma_j^2) / K \right)^2 \right] \rightarrow 0, E \left[\left(\sum_{j < i} \sigma_i^2 P_{ij}^2 (u_j^2 - \omega_j^2) / K \right)^2 \right] \rightarrow 0.$$

Next, it follows by Lemma ?? that

$$E \left[\left(\sum_{j < k < i} \gamma_i P_{ij} P_{ik} u_j \varepsilon_k / K \right)^2 \right] \rightarrow 0,$$

and similarly that all of the remaining terms in eq. (4) converge to zero in mean square, so that the conclusion follows by the martingale central limit theorem. Q.E.D.

Proof of Lemma A4: Let $\bar{w}_i = E[W_i]$, $\tilde{W}_i = W_i - \bar{w}_i$, $\bar{y}_i = E[Y_i]$, $\tilde{Y}_i = Y_i - \bar{y}_i$, $\bar{\eta}_i = E[\eta_i]$, $\tilde{\eta}_i = \eta_i - \bar{\eta}_i$,

$$\begin{aligned}\bar{\mu}_W^2 &= \max_{i \leq n} \bar{w}_i^2 \leq C/n, \bar{\mu}_Y^2 = \max_{i \leq n} \bar{y}_i^2 \leq C/n, \bar{\mu}_\eta^2 = \max_{i \leq n} \bar{\eta}_i^2 \leq C, \\ \bar{\sigma}_W^2 &= \max_{i \leq n} \text{Var}(W_i) \leq C/r_n, \bar{\sigma}_Y^2 = \max_{i \leq n} \text{Var}(Y_i) \leq C/r_n, \bar{\sigma}_\eta^2 = \max_{i \leq n} \text{Var}(\eta_i) \leq C.\end{aligned}$$

Also, let

$$\check{y}_i = \sum_j P_{ij} \bar{y}_j, \check{w}_i = \sum_j P_{ij} \bar{w}_j,$$

be predicted values from projecting \bar{y} and \bar{w} on P , and note that

$$\sum_i \check{y}_i^2 \leq \sum_i \bar{y}_i^2 \leq C, \sum_i \check{w}_i^2 \leq \sum_i \bar{w}_i^2 \leq C.$$

By independent observations we have

$$\begin{aligned}A_n &= \sum_{i \neq j} \sum_{k \notin \{i, j\}} \bar{w}_i P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j = \sum_{i \neq j} \sum_k \bar{w}_i P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j - \sum_{i \neq j} \bar{w}_i P_{ii} \bar{\eta}_i P_{ij} \bar{y}_j - \sum_{i \neq j} \bar{w}_i P_{ij} \bar{\eta}_j P_{jj} \bar{y}_j \\ &= \sum_k \check{w}_k \bar{\eta}_k \check{y}_k - \sum_{i, k} \bar{w}_i \bar{y}_i P_{ik}^2 \bar{\eta}_k - \sum_i \bar{w}_i P_{ii} \bar{\eta}_i \check{y}_i - \sum_i \check{w}_i P_{ii} \bar{\eta}_i \bar{y}_i + 2 \sum_i \bar{w}_i \bar{y}_i P_{ii}^2 \bar{\eta}_i.\end{aligned}$$

By CS,

$$\begin{aligned}\left| \sum_k \check{w}_k \bar{\eta}_k \check{y}_k \right| &\leq \sqrt{\sum_k \check{w}_k^2} \sqrt{\sum_k \bar{\eta}_k^2 \check{y}_k^2} \leq C \sqrt{\sum_k \check{w}_k^2} \sqrt{\sum_k \check{y}_k^2} \leq C, \\ \left| \sum_i \bar{w}_i P_{ii} \bar{\eta}_i \check{y}_i \right| &\leq \sqrt{\sum_i \bar{w}_i^2 P_{ii}^2 \bar{\eta}_i^2} \sqrt{\sum_i \check{y}_i^2} \leq C \sqrt{\sum_i \bar{w}_i^2} \sqrt{\sum_i \check{y}_i^2} \leq C,\end{aligned}$$

and it follows similarly that $\sum_i \check{w}_i P_{ii} \bar{\eta}_i \bar{y}_i$ is bounded. By Lemma B1,

$$\left| \sum_{i, k} \bar{w}_i \bar{y}_i P_{ik}^2 \bar{\eta}_k \right| \leq Cn^{-1} \left| \sum_{i, k} P_{ik}^2 \right| \leq CK/n \leq C.$$

Also, $|\sum_i \bar{w}_i \bar{y}_i P_{ii}^2 \bar{\eta}_i| \leq Cn/n = C$. Thus, $|A_n| \leq C$ holds by T.

Next, note that

$$\begin{aligned}W_i P_{ik} \eta_k P_{kj} Y_j &= \tilde{W}_i P_{ik} \eta_k P_{kj} Y_j + \bar{w}_i P_{ik} \eta_k P_{kj} Y_j \\ &= \tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} Y_j + \tilde{W}_i P_{ik} \bar{\eta}_k P_{kj} Y_j + \bar{w}_i P_{ik} \tilde{\eta}_k P_{kj} Y_j + \bar{w}_i P_{ik} \bar{\eta}_k P_{kj} Y_j \\ &= \tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j + \tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \bar{y}_j + \tilde{W}_i P_{ik} \bar{\eta}_k P_{kj} \tilde{Y}_j + \tilde{W}_i P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j \\ &\quad + \bar{w}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j + \bar{w}_i P_{ik} \tilde{\eta}_k P_{kj} \bar{y}_j + \bar{w}_i P_{ik} \bar{\eta}_k P_{kj} \tilde{Y}_j + \bar{w}_i P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j.\end{aligned}$$

Summing and subtracting the last term gives

$$\sum_{i \neq j \neq k} W_i P_{ik} \eta_k P_{kj} Y_j - A_n = \sum_{r=1}^7 \hat{\psi}_r,$$

wherer

$$\begin{aligned} \hat{\psi}_1 &= \sum_{i \neq j \neq k} \tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j, \hat{\psi}_2 = \sum_{i \neq j \neq k} \tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \bar{y}_j, \hat{\psi}_3 = \sum_{i \neq j \neq k} \tilde{W}_i P_{ik} \bar{\eta}_k P_{kj} \tilde{Y}_j, \\ \hat{\psi}_4 &= \sum_{i \neq j \neq k} \tilde{W}_i P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j, \hat{\psi}_5 = \sum_{i \neq j \neq k} \bar{w}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j, \hat{\psi}_6 = \sum_{i \neq j \neq k} \bar{w}_i P_{ik} \tilde{\eta}_k P_{kj} \bar{y}_j, \end{aligned}$$

and $\hat{\psi}_7 = \sum_{i \neq j \neq k} \bar{w}_i P_{ik} \bar{\eta}_k P_{kj} \tilde{Y}_j$. By T the second conclusion will follow from $\hat{\psi}_r \xrightarrow{p} 0$, ($r = 1, \dots, 7$). Also, note that $\hat{\psi}_7$ is the same as $\hat{\psi}_4$ and $\hat{\psi}_5$ is the same as $\hat{\psi}_2$ with the random variables W and Y interchanged. Since the conditions on W and Y are symmetric it suffices to show that $\hat{\psi}_r \xrightarrow{p} 0$, $r \in \{1, 2, 3, 4, 6\}$.

Consider now $\hat{\psi}_1$. Note that for $i \neq j \neq k$ and $r \neq s \neq t$ we have $E[\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j \tilde{W}_r P_{rs} \tilde{\eta}_s P_{st} \tilde{Y}_t] = 0$ except when the each of the three indices i, j, k is equal to one of the three indices r, s, t . There are six ways this can happen leading to six terms in

$$E[\hat{\psi}_1^2] = \sum_{i \neq j \neq k} \sum_{r \neq s \neq t} E[\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j \tilde{a}_r P_{rs} \tilde{\eta}_s P_{st} \tilde{Y}_t] = \sum_{q=1}^6 \hat{\tau}_q.$$

Note that by hypothesis, $\bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \bar{\sigma}_Y^2 K \leq C r_n^{-2} K \rightarrow 0$. By Lemma B1, we have

$$|\hat{\tau}_1| = \sum_{i \neq j \neq k} E[(\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j)^2] = \sum_{i \neq j \neq k} E[\tilde{W}_i^2] P_{ik}^2 E[\tilde{\eta}_k^2] P_{kj}^2 E[\tilde{Y}_j^2] \leq \bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \bar{\sigma}_Y^2 K \rightarrow 0.$$

Similarly, by CS,

$$\begin{aligned} |\hat{\tau}_3| &= \left| \sum_{i \neq j \neq k} E[(\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j)(\tilde{W}_j P_{jk} \tilde{\eta}_k P_{ki} \tilde{Y}_i)] \right| = \left| \sum_{i \neq j \neq k} E[\tilde{W}_i \tilde{Y}_i] E[\tilde{W}_j \tilde{Y}_j] E[\tilde{\eta}_k^2] P_{ik}^2 P_{kj}^2 \right| \\ &\leq \bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \bar{\sigma}_Y^2 K \rightarrow 0. \end{aligned}$$

Next, by Lemma B1 and CS

$$\begin{aligned} |\hat{\tau}_2| &= \left| \sum_{i \neq j \neq k} E[(\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j)(\tilde{W}_i P_{ij} \tilde{\eta}_j P_{jk} \tilde{Y}_k)] \right| = \left| \sum_{i \neq j \neq k} E[\tilde{W}_i^2] E[\tilde{\eta}_k \tilde{Y}_k] E[\tilde{\eta}_j \tilde{Y}_j] P_{ik} P_{ij} P_{jk}^2 \right| \\ &\leq \bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \bar{\sigma}_Y^2 K \rightarrow 0. \end{aligned}$$

Similarly,

$$\begin{aligned} |\hat{\tau}_4| &= \left| \sum_{i \neq j \neq k} E[(\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j)(\tilde{W}_j P_{ji} \tilde{\eta}_i P_{ik} \tilde{Y}_k)] \right| = \left| \sum_{i \neq j \neq k} E[\tilde{W}_i \tilde{\eta}_i] E[\tilde{W}_j \tilde{Y}_j] E[\tilde{\eta}_k \tilde{Y}_k] P_{ik}^2 P_{kj} P_{ji} \right| \\ &\leq \bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \bar{\sigma}_Y^2 K \longrightarrow 0, \end{aligned}$$

$$\begin{aligned} |\hat{\tau}_5| &= \left| \sum_{i \neq j \neq k} E[(\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j)(\tilde{W}_k P_{ki} \tilde{\eta}_i P_{ij} \tilde{Y}_j)] \right| = \left| \sum_{i \neq j \neq k} E[\tilde{W}_i \tilde{\eta}_i] E[\tilde{Y}_j^2] E[\tilde{W}_k \tilde{\eta}_k] P_{ik}^2 P_{kj} P_{ji} \right| \\ &\leq \bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \bar{\sigma}_Y^2 K \longrightarrow 0, \end{aligned}$$

$$\begin{aligned} |\hat{\tau}_6| &= \left| \sum_{i \neq j \neq k} E[(\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \tilde{Y}_j)(\tilde{W}_k P_{kj} \tilde{\eta}_j P_{ji} \tilde{Y}_i)] \right| = \left| \sum_{i \neq j \neq k} E[\tilde{W}_i \tilde{Y}_i] E[\tilde{\eta}_j \tilde{Y}_j] E[\tilde{W}_k \tilde{\eta}_k] P_{jk}^2 P_{ij} P_{ik} \right| \\ &\leq \bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \bar{\sigma}_Y^2 K \longrightarrow 0. \end{aligned}$$

The triangle inequality then gives $E[\hat{\psi}_1^2] \longrightarrow 0$, so $\hat{\psi}_1 \xrightarrow{p} 0$ holds by M.

Consider now $\hat{\psi}_2$. Note that for $i \neq j \neq k$ and $r \neq s \neq t$ we have $E[\tilde{W}_i P_{ik} \tilde{\eta}_k P_{kj} \bar{y}_j \tilde{W}_r P_{rs} \tilde{\eta}_s P_{st} \bar{y}_t] = 0$ except when $i = r$ and $j = s$ or $i = s$ and $j = r$. Then by $(A + B + C)^2 \leq 3(A^2 + B^2 + C^2)$ and for fixed k , $\sum_{i \neq k} P_{ik}^2 \leq P_{kk}$, $\sum_{i \neq k} P_{ik}^4 \leq P_{kk}$, it follows that

$$\begin{aligned} \sum_{i \neq k} P_{ik}^2 \left(\sum_{j \notin \{i, k\}} P_{kj} \bar{y}_j \right)^2 &\leq 3 \sum_{i \neq k} P_{ik}^2 \left(\bar{y}_k^2 + P_{ki}^2 \bar{y}_i^2 + P_{kk}^2 \bar{y}_k^2 \right) \\ &\leq 3 \left(\sum_k P_{kk} \left(\bar{y}_k^2 + 2\bar{y}_k^2 \right) \right) \leq 3 \left(\sum_k \bar{y}_k^2 + 2 \sum_k \bar{y}_k^2 \right) \leq 9n\bar{\mu}_Y^2 \leq C. \end{aligned}$$

It follows by $|AB| \leq (A^2 + B^2)/2$, CS, and $P_{ik} = P_{ki}$ that

$$\begin{aligned} E[\hat{\psi}_2^2] &= \sum_{i \neq k} E[\tilde{W}_i^2] P_{ik}^2 E[\tilde{\eta}_k^2] \left(\sum_{j \notin \{i, k\}} P_{kj} \bar{y}_j \right)^2 \\ &\quad + \sum_{i \neq k} E[\tilde{W}_i \tilde{\eta}_i] P_{ik}^2 E[\tilde{W}_k \tilde{\eta}_k] \left(\sum_{j \notin \{i, k\}} P_{kj} \bar{y}_j \right) \left(\sum_{j \notin \{i, k\}} P_{ij} \bar{y}_j \right) \\ &\leq 2\bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \sum_{i \neq k} P_{ik}^2 \left(\sum_{j \notin \{i, k\}} P_{kj} \bar{y}_j \right)^2 \leq C/r_n \longrightarrow 0. \end{aligned}$$

Then $\hat{\psi}_2 \xrightarrow{p} 0$ holds by M.

Consider $\hat{\psi}_3$. Note that for $i \neq j \neq k$ and $r \neq s \neq t$ we have $E[\tilde{W}_i P_{ik} \bar{\eta}_k P_{kj} \tilde{Y}_j \tilde{W}_r P_{rs} \bar{\eta}_s P_{st} \tilde{Y}_t] = 0$ except when $i = r$ and $j = t$ or $i = t$ and $j = r$. Thus,

$$\begin{aligned} E[\hat{\psi}_3^2] &= \sum_{i \neq j} \left(E[\tilde{W}_i^2] E[\tilde{Y}_j^2] + E[\tilde{W}_i \tilde{Y}_i] E[\tilde{W}_j \tilde{Y}_j] \right) \left(\sum_{k \notin \{i, j\}} P_{ik} \bar{\eta}_k P_{kj} \right)^2 \\ &\leq 2\bar{\sigma}_W^2 \bar{\sigma}_\eta^2 \sum_{i \neq j} \left(\sum_{k \notin \{i, j\}} P_{ik} \bar{\eta}_k P_{kj} \right)^2. \end{aligned}$$

Note that

$$\sum_{k \notin \{i, j\}} P_{ik} P_{kj} \bar{\eta}_k = \sum_k P_{ik} P_{kj} \bar{\eta}_k - P_{ij} P_{ii} \bar{\eta}_i - P_{ij} P_{jj} \bar{\eta}_j.$$

Note also that

$$\begin{aligned} \sum_i \left(\sum_k P_{ik} \bar{\eta}_k P_{ki} \right)^2 &= \sum_{i, k, \ell} P_{ik}^2 P_{i\ell}^2 \bar{\eta}_k \bar{\eta}_\ell \leq \bar{\mu}_\eta^2 \sum_{i, k, \ell} P_{ik}^2 P_{i\ell}^2 = \bar{\mu}_\eta^2 \sum_i P_{ii}^2 \leq \bar{\mu}_\eta^2 K, \\ \sum_{i, j} \left(\sum_k P_{ik} \bar{\eta}_k P_{kj} \right)^2 &= \sum_{i, j, k, \ell} P_{ik} \bar{\eta}_k P_{jk} P_{i\ell} \bar{\eta}_\ell P_{j\ell} = \sum_{k, \ell} \bar{\eta}_k \bar{\eta}_\ell \left(\sum_i P_{ik} P_{i\ell} \right) \left(\sum_j P_{jk} P_{j\ell} \right) \\ &= \sum_{k, \ell} \bar{\eta}_k \bar{\eta}_\ell P_{k\ell}^2 \leq \bar{\mu}_\eta^2 \sum_{k, \ell} P_{k\ell}^2 = \bar{\mu}_\eta^2 K. \end{aligned}$$

It therefore follows that

$$\sum_{i \neq j} \left(\sum_k P_{ik} \bar{\eta}_k P_{kj} \right)^2 = \sum_{i, j} \left(\sum_k P_{ik} \bar{\eta}_k P_{kj} \right)^2 - \sum_i \left(\sum_k P_{ik} \bar{\eta}_k P_{ki} \right)^2 \leq 2\bar{\mu}_\eta^2 K.$$

Also, by Lemma B1, $\sum_{i \neq j} P_{ij}^2 P_{jj}^2 \bar{\eta}_j^2 \leq \bar{\mu}_\eta^2 \sum_{i \neq j} P_{ij}^2 \leq \bar{\mu}_\eta^2 K$, so that

$$\sum_{i \neq j} \left(\sum_{k \notin \{i, j\}} P_{ik} \bar{\eta}_k P_{kj} \right)^2 \leq 3 \sum_{i \neq j} \left\{ \left(\sum_{k \notin \{i, j\}} P_{ik} \bar{\eta}_k P_{kj} \right)^2 + P_{ij}^2 P_{ii}^2 \bar{\eta}_i^2 + P_{ij}^2 P_{jj}^2 \bar{\eta}_j^2 \right\} \leq 6\bar{\mu}_\eta^2 K.$$

From the previous expression for $E[\hat{\psi}_3^2]$ we then have

$$E[\hat{\psi}_3^2] \leq C\bar{\sigma}_W^2 \bar{\sigma}_Y^2 \bar{\mu}_\eta^2 K \leq Cr_n^{-2} K \longrightarrow 0.$$

Then $\hat{\psi}_3 \xrightarrow{p} 0$ by M.

Next, consider $\hat{\psi}_4$. Note that for $i \neq j \neq k$ and $r \neq s \neq t$ we have $E[\tilde{W}_i P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j \tilde{W}_r P_{rs} \bar{\eta}_s P_{st} \bar{y}_t] = 0$ except when $i = r$. Thus,

$$E[\hat{\psi}_4^2] = \sum_i E[\tilde{W}_i^2] \left(\sum_{j \neq i} \sum_{k \notin \{i, j\}} P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j \right)^2 \leq \bar{\sigma}_W^2 \sum_i \left(\sum_{j \neq i} \sum_{k \notin \{i, j\}} P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j \right)^2.$$

Note that for $i \neq j$,

$$\sum_{k \notin \{i,j\}} P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j = \sum_k P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j - P_{ii} \bar{\eta}_i P_{ij} \bar{y}_j - P_{ij} \bar{\eta}_j P_{jj} \bar{y}_j.$$

Therefore, for fixed i ,

$$\begin{aligned} \sum_{j \neq i} \sum_{k \notin \{i,j\}} P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j &= \sum_{j \neq i} \left(\sum_k P_{ik} \bar{\eta}_k P_{kj} \bar{y}_j - P_{ii} \bar{\eta}_i P_{ij} \bar{y}_j - P_{ij} \bar{\eta}_j P_{jj} \bar{y}_j \right) \\ &= \sum_k P_{ik} \bar{\eta}_k \check{y}_k - P_{ii} \bar{\eta}_i \check{y}_i - \sum_j P_{ij} \bar{\eta}_j P_{jj} \bar{y}_j - \sum_k P_{ik}^2 \bar{\eta}_k \bar{y}_i - 2P_{ii}^2 \bar{\eta}_i \bar{y}_i. \end{aligned}$$

Note that by P idempotent we have

$$\sum_j \sum_k P_{jk} \bar{\eta}_j \check{y}_j \bar{\eta}_k \check{y}_k \leq \sum_j \bar{\eta}_j^2 \check{y}_j^2 \leq \bar{\mu}_\eta^2 \sum_j \check{y}_j^2 \leq \bar{\mu}_\eta^2 \sum_j \bar{y}_j^2 \leq n \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \leq C.$$

Then it follows that

$$\begin{aligned} \sum_i \left\{ \sum_k P_{ik} \bar{\eta}_k \check{y}_k \right\}^2 &= \sum_i \sum_j \sum_k P_{ij} \bar{\eta}_j \check{y}_j P_{ik} \bar{\eta}_k \check{y}_k = \sum_j \sum_k \bar{\eta}_j \check{y}_j \bar{\eta}_k \check{y}_k \sum_i P_{ij} P_{ik} \\ &= \sum_j \sum_k P_{jk} \bar{\eta}_j \check{y}_j \bar{\eta}_k \check{y}_k \leq C. \end{aligned}$$

Also, using similar reasoning,

$$\begin{aligned} \sum_i (P_{ii} \bar{\eta}_i \check{y}_i)^2 &\leq \sum_i \bar{\eta}_i^2 \check{y}_i^2 \leq n \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \leq C, \\ \sum_i \left(\sum_j P_{ij} \bar{\eta}_j P_{jj} \bar{y}_j \right)^2 &\leq \sum_i \bar{\eta}_i^2 P_{ii}^2 \bar{y}_i^2 \leq \sum_i \bar{\eta}_i^2 \check{y}_i^2 \leq C, \\ \sum_i \left(\bar{y}_i \sum_k P_{ik}^2 \bar{\eta}_k \right)^2 &\leq \bar{\mu}_Y^2 \sum_{i,k,\ell} P_{ik}^2 P_{i\ell}^2 \bar{\eta}_k \bar{\eta}_\ell \leq \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \sum_{i,k,\ell} P_{ik}^2 P_{i\ell}^2 \leq K \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \leq C, \\ \sum_i P_{ii}^4 \bar{\eta}_i^2 \bar{y}_i^2 &\leq n \bar{\mu}_\eta^2 \bar{\mu}_Y^2. \end{aligned}$$

Then using the fact that $(\sum_{r=1}^5 A_r)^2 \leq 5 \sum_{r=1}^5 A_r^2$ it follows that $E[\hat{\psi}_4^2] \leq \bar{\sigma}_W^2 C \leq C/r_n \rightarrow 0$, so that $\hat{\psi}_4 \xrightarrow{P} 0$ by M.

Next, consider $\hat{\psi}_6$. Note that for $i \neq k$,

$$\sum_{j \notin \{i,k\}} \bar{w}_i P_{ik} P_{kj} \bar{y}_j = \bar{w}_i P_{ik} \check{y}_k - \bar{w}_i P_{ik}^2 \bar{y}_i - \bar{w}_i P_{ik} P_{kk} \bar{y}_k.$$

Then for fixed k ,

$$\begin{aligned} \sum_{i \neq k} \sum_{j \notin \{i, k\}} \bar{w}_i P_{ik} P_{kj} \bar{y}_j &= \sum_i \left(\bar{w}_i P_{ik} \check{y}_k - \bar{w}_i P_{ik}^2 \bar{y}_i - \bar{w}_i P_{ik} P_{kk} \bar{y}_k \right) - \bar{w}_k P_{kk} \check{y}_k + 2\bar{w}_k P_{kk}^2 \bar{y}_k \\ &= \check{w}_k \check{y}_k - \sum_i \bar{w}_i P_{ik}^2 \bar{y}_i - \check{w}_i P_{kk} \bar{y}_k - \bar{w}_k P_{kk} \check{y}_k + 2\bar{w}_k P_{kk}^2 \bar{y}_k \end{aligned}$$

Then using the fact that $(\sum_{r=1}^5 A_r)^2 \leq 5 \sum_{r=1}^5 A_r^2$ we have

$$\begin{aligned} E[\hat{\psi}_6^2] &= \sum_k E[\hat{\eta}_k^2] \left(\sum_{i \neq k} \sum_{j \notin \{i, k\}} \bar{w}_i P_{ik} P_{kj} \bar{y}_j \right)^2 \\ &\leq 5\bar{\sigma}_\eta^2 \sum_k \left(\check{w}_k^2 \check{y}_k^2 + \sum_{i, j} P_{kj}^2 P_{ki}^2 \bar{w}_i \bar{y}_i \bar{w}_j \bar{y}_j + \check{w}_k^2 P_{kk}^2 \check{y}_k^2 + \bar{w}_k^2 P_{kk}^2 \check{y}_k^2 + \bar{w}_k^2 P_{kk}^4 \bar{y}_k^2 \right) \\ &\leq 5\bar{\sigma}_\eta^2 \left(\sum_k \check{w}_k^2 \check{y}_k^2 + \bar{\mu}_W^2 \bar{\mu}_Y^2 \sum_{i, j, k} P_{kj}^2 P_{ki}^2 + \bar{\mu}_Y^2 \sum_k \check{w}_k^2 + \bar{\mu}_W^2 \sum_k \check{y}_k^2 + n \bar{\mu}_W^2 \bar{\mu}_Y^2 \right) \\ &\leq 5\bar{\sigma}_\eta^2 \left(\sum_k \check{w}_k^2 \check{y}_k^2 + 4n \bar{\mu}_W^2 \bar{\mu}_Y^2 \right) \leq C \sum_k \check{w}_k^2 \check{y}_k^2 + Cn/n^2 \leq C \sum_k \check{w}_k^2 \check{y}_k^2 + o(1). \end{aligned}$$

Now let π_n be such that $\Delta_n = \max_i |a_i - Z_i' \pi_n| \rightarrow 0$, let $\alpha_n = \pi_n / \sqrt{n}$ and note that $\max_{i \leq n} |\bar{w}_i - Z_i' \alpha_n| = \Delta_n / \sqrt{n}$. Let $\bar{w} = (\bar{w}_1, \dots, \bar{w}_n)'$. Then

$$\begin{aligned} |\bar{w}_i - \check{w}_i| &= \left| \bar{w}_i - Z_i' (Z' Z)^{-1} Z' \bar{w} \right| = \left| \bar{w}_i - Z_i' \alpha_n - Z_i' (Z' Z)^{-1} Z' (\bar{w} - Z \alpha_n) \right| \\ &\leq \Delta_n / \sqrt{n} + \left(\sum_j P_{ij}^2 \right)^{1/2} \left(\sum_j [\bar{w}_j - Z_j' \alpha_n]^2 \right)^{1/2} \leq C \Delta_n. \end{aligned}$$

Then by the triangle inequality, $\max_{i \leq n} |\check{w}_i| \leq \max_{i \leq n} |\bar{w}_i| + \Delta_n \rightarrow 0$, so that

$$\sum_k \check{w}_k^2 \check{y}_k^2 \leq \left(\max_{i \leq n} |\check{w}_i| \right)^2 \sum_k \check{y}_k^2 = o(1) \sum_k \bar{y}_k^2 \rightarrow 0.$$

Then we have $E[\hat{\psi}_6^2] \rightarrow 0$, so that

7 References

- Ackerberg, D.A. and P. Deveraux (2003). “Improved JIVE Estimators for Overidentified Models with and without Heteroskedasticity,” Working Paper, University of California, Los Angeles.
- Andrews, D.W.K. and J.H. Stock (2005). “Testing with Many Weak Instruments,” Working Paper, Harvard University.

- Angrist, J.D., G.W. Imbens, and A. Krueger (1999), "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57-67.
- Angrist, J. D. and A. B. Krueger (1995). "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business and Economic Statistics*, 13, 225-235.
- Bekker, P.A. (1994). "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657-681.
- Bekker, P. A. and J. van der Ploeg (2005), "Instrumental Variable Estimation Based on Grouped Data," *Statistica Neerlandica* 59, 506-508.
- Blomquist, S. and M. Dahlberg (1999). "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics* 14, 69-88.
- Chao, J.C. and N.R. Swanson (2005). "Consistent Estimation with a Large Number of Weak Instruments," *Econometrica*, 73, 1673-1692.
- Chao, J.C., W.K. Newey and N.R. Swanson (2006). "Jackknife GMM with Many Weak Instruments", Working Paper, MIT.
- Chao, J.C. and N.R. Swanson (2006). "Asymptotic Normality of Single-Equation Estimators for the Case with a Large Number of Weak Instruments," in: *Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, pp. 82-124. Cambridge: Cambridge University Press.
- R. Davidson and J.G. MacKinnon (2006). "The Case Against JIVE," *Journal of Applied Econometrics* 21, 827-833.
- Donald, S.G. and W.K. Newey (2001). "Choosing the Number of Instruments," *Econometrica*, 69, 1161-1191.
- Gänsler, P. and W. Stute (1977). *Wahrscheinlichkeitstheorie* (Springer-Verlag, New York).
- Gao, C. and K. Lahiri (2000). "A Comparison of Some Recent Bayesian and Classical Procedures for Simultaneous Equation Models with Weak Instruments," Working Paper, SUNY Albany.

- Hahn, J. (2002). "Optimal Inference with Many Instruments," *Econometric Theory*, 18, 140-168.
- Hahn, J., J. Hausman, and G. Kuersteiner (2001). "Accuracy of Higher Order Bias and MSE Approximations," *Econometrics Journal*, 7, 272-306.
- Han, C. and P.C.B. Phillips (2006). "GMM With Many Moment Conditions," *Econometrica* 74, 147-192.
- Hansen, C., J.A. Hausman, and W.K. Newey (2006). "Estimation with Many Instrumental Variables", Working Paper, MIT.
- Hausman, J.A., W.K. Newey, T.M. Woutersen, J. Chao, and N.R. Swanson (2007). "IV Estimation with Heteroskedasticity and Many Instruments", Working Paper, MIT.
- Horn, R. A. and C. R. Johnson (1985). *Matrix Analysis*. New York: Cambridge University Press.
- Kelejian, H.H. and I.R. Prucha (2001). "On the Asymptotic Distribution of the Moran I Test Statistic with Applications," *Journal of Econometrics*, 104, 219-57.
- 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.
- Magnus, J.R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (John Wiley & Sons, New York).
- Moreira, M. (2003). "A Conditional Likelihood Ratio Test for Structural Model," *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. and R. J. Smith (2004). "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

- Newey, W.K. and F. Windmeijer (2006). "GMM with Many Weak Moment Conditions", Working Paper, MIT.
- Phillips, G.D.A. and C. Hale (1977). "The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems," *International Economic Review*, 18, 219-228.
- Phillips, P.C.B. (1983). "Small Sample Distribution Theory in Econometric Models of Simultaneous Equations," in Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics, Vol I*. Amsterdam: North Holland.
- Rothenberg, T.J. (1983). "Asymptotic Properties of Some Estimators in Structural Models," in S. Karlin, T. Amemiya, and L. Goodman, eds., *Studies in Econometrics, Time Series, and Multivariate Statistics*. San Diego: Academic Press.
- Staiger, D. and J.H. Stock (1997). "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65, 557-586.
- Smith, R. J. (1997). "Alternative Semi-Parametric Likelihood Approaches to Generalized Methods of Moments Estimation," *Economic Journal*, 107, 503-519.
- Stock, J.H. and M. Yogo (2005a). "Testing for Weak Instruments in Linear IV Regression," in D.W.K. Andrews and J.H. Stock eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, pp. 80-108. Cambridge: Cambridge University Press.
- Stock, J.H. and M. Yogo (2005b). "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," in D.W.K. Andrews and J.H. Stock eds. *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, pp. 109-120. Cambridge: Cambridge University Press.
- Wang, J. and E. Zivot (1998). "Inference on Structural Parameters in Instrumental Variables Regressions with Weak Instruments," *Econometrica*, 66, 1389-1404.
- White, H. (2001). *Asymptotic Theory for Econometricians*. San Diego: Academic Press.