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

John C. Chao[1], Norman R. Swanson[2], Jerry A. Hausman[3], Whitney K. Newey[3], and Tiemen Woutersen[5]

[1]University of Maryland, [2]Rutgers University, [3]MIT, and [4]Johns Hopkins University

August, 2007
Revised, November, 2007

Abstract

This paper derives the limiting distributions of alternative jackknife IV (JIV) estimators and gives formulae for accompanying consistent standard errors in the presence of 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 that standard errors are consistent provided that $\frac{\sqrt{K_n}}{r_n} \to 0$, as $n \to \infty$, 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 $\frac{K_n}{r_n} \to 0$. 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 orders 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. Jerry A. Hausman, Department of Economics, MIT, Cambridge MA 02142-1347, jhausman@mit.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. Tiemen Woutersen: Department of Economics, Johns Hopkins University, Baltimore, MD, USA 21218, woutersen@jhu.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), Ackerberg 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 limited information maximum likelihood (LIML) and bias corrected 2SLS (B2SLS) estimators are not (see e.g. Bekker and van der Ploeg (2005), Ackerberg and Deveraux (2003), Chao and Swanson (2006), 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 weak instrument sequence of Chao and Swanson (2005). To be precise we show that JIV estimators are consistent and asymptotically normal when $\frac{\sqrt{K_n}}{r_n} \to 0$, as $n \to \infty$, where K_n and r_n denote, the number of instruments and the rate of growth of the so-called concentration parameter, respectively. In contrast, consistency of LIML and B2SLS generally requires that $\frac{K_n}{r_n} \to 0$, as $n \to \infty$, meaning that 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 the asymptotic normality of JIV, this paper gives a central limit theorem for quadratic (and, more generally, bilinear) forms associated with an idempotent matrix. This theorem 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

Fuller (1977) estimators that are robust to heteroskedasticity, and that are as efficient as $LIML$ under homoskedasticity.

This paper is a substantially altered and revised version of Chao and Swanson (2004), in which we now allow for the many instrument sequence of Kunitomo (1980), Morimune (1983) and Bekker (1994); and in which we give a refined version of the central limit theorem from the earlier paper.

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

$$\begin{aligned} \underset{n\times 1}{y} &= \underset{n\times G}{X}\underset{G\times 1}{\delta_0} + \underset{n\times 1}{\varepsilon}, \\ X &= \Upsilon + U, \end{aligned}$$

where n is the number of observations, G is 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 an $n \times K$ matrix, Z, of instrumental variable observations with $rank(Z) = K$. Let $\mathcal{Z} = (\Upsilon, Z)$, and assume that $E[\varepsilon|\mathcal{Z}] = 0$ and $E[U|\mathcal{Z}] = 0$.

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 let $\Upsilon_i = f_0(w_i)$ be an vector of unknown functions of a vector w_i of underlying instruments and let $Z_i = (p_{1K}(w_i), ..., p_{KK}(w_i))'$, 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 $P = Z(Z'Z)^{-1}Z'$. Additionally, let $\bar{\Pi}_{-i} = (Z'Z - Z_iZ_i')^{-1}(Z'X -$

$Z_i X_i')$ 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

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

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 that

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

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

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

As we will show, these estimators are consistent and asymptotically normal under heteroskedasticity, when $\sqrt{K}/r_n \longrightarrow 0$, where r_n is proportional to the concentration parameter. In contrast, consistency of LIML and Fuller (1977) requires $K/r_n \longrightarrow 0$, when P_{ii} is asymptotically correlated with $E[X_i \varepsilon_i | \mathcal{Z}] / E[\varepsilon_i^2 | \mathcal{Z}]$, 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 $E[X_i \varepsilon_i | \mathcal{Z}]$, as discussed in Ackerberg and Deveraux (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, $\tilde{\xi}_i = (1 - P_{ii})^{-1}\left(y_i - X_i'\tilde{\delta}\right)$. Using the independence of the observations, $E[\varepsilon_i|\mathcal{Z}] = 0$, and the exclusion of the $i = j$ terms in the double sums, it follows that

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

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

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

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

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

This estimator is robust to heteroskedasticity, as it allows $Var(\xi_i|\mathcal{Z})$ to vary over i and $E[X_i \varepsilon_i|\mathcal{Z}]$ 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

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

The variance estimator for $\hat{\delta}$ is then 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 of our regularity conditions are conditional on $\mathcal{Z} = (\Upsilon, Z)$. 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: $K = K_n \longrightarrow \infty$, 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, $rank(Z) = K$ and $P_{ii} \leq C$, $(i = 1, ..., n)$.[1]

The restriction that $rank(Z) = K$ is a normalization that requires excluding redundant columns from Z. It can be verified in particular cases. For instance, when w_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 $K < n$.[2] The condition $P_{ii} \leq C < 1$ implies that $K/n \leq C$, because $K/n = \sum_{i=1}^n P_{ii}/n \leq C$.

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

Assumption 2: There is a $G \times G$ matrix $S_n = \tilde{S}_n \operatorname{diag}(\mu_{1n}, ..., \mu_{Gn})$ 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 $\mu_{jn}/\sqrt{n} \longrightarrow 0$, $\mu_n = \min_{1 \leq j \leq G} \mu_{jn} \longrightarrow \infty$, and $\sqrt{K}/\mu_n^2 \longrightarrow 0$. Also, there is $C > 0$ such that with probability one $\|\sum_{i=1}^n z_i z_i'/n\| \leq C$ and $\lambda_{\min}\left(\sum_{i=1}^n z_i z_i'/n\right) \geq C$, for n sufficiently large.

Here $r_n = \mu_n^2$ can be interpeted as being proportional to the concentration parameter. For instance, note that in the $G = 1$ case, we have that $\sum_{i=1}^n \Upsilon_i^2 = \mu_n^2 \sum_{i=1}^n z_i^2/n$, so that since $\sum_{i=1}^n z_i^2/n$ is 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 as in 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 implies modeling the case where we have 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

[1] Note that, in this paper, C is the generic notation for a constant. It is used to denote different constants in different parts of the paper. Hence, although in Assumption 1, C is taken to be less than 1, in other parts of the paper, it might not be.

[2] The observations $w_1, ..., w_T$ 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 i with $a'Z_i = a'p^K(w_i) \neq 0$, implying $a'Z'Za > 0$.

$\mu_{jn}/\sqrt{n} \longrightarrow 0$) 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.

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

Assumption 3: There is a constant, C, such that conditional on $\mathcal{Z} = (\Upsilon, Z)$, the observations $(\varepsilon_1, U_1), ..., (\varepsilon_n, U_n)$ are independent, with $E[\varepsilon_i|\mathcal{Z}] = 0$, $E[U_i|\mathcal{Z}] = 0$, $\sup_i E[\varepsilon_i^2|\mathcal{Z}] < C$, and $\sup_i E[\|U_i\|^2|\mathcal{Z}] \leq C$, almost surely.

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

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

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

We can easily interpret all of these conditions in the context of the 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 $p^K(w) \overset{def}{=} (p_{1K}(w), ..., p_{K-G_2,K}(w))'$. 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 $z_i = \pi_{Kn} Z_i$, for some π_{Kn} 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 where the number of instruments can grow as fast as the sample size. When $\mu_n^2/n \longrightarrow 0$, it is a modified version thereof, where the model is more weakly identified.

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

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

By construction we have that $\Upsilon_i = S_n z_i/\sqrt{n}$. Assumption 2 imposes the requirements that for n sufficiently large, and with probability one,

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

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

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

THEOREM 1: *Suppose that Assumptions 1-4 are satisfied. Then,* $\mu_n^{-1} S_n'(\tilde{\delta} - \delta_0), \tilde{\delta} \xrightarrow{p} 0, \xrightarrow{p} \delta_0$, $\mu_n^{-1} S_n'(\hat{\delta} - \delta_0) \xrightarrow{p} 0$, *and* $\hat{\delta} \xrightarrow{p} \delta_0$.

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

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

To give asymptotic normality results, we need to describe the asymptotic variances. We will outline results that do not depend on convergence of various moment matrices; and so we will write the asymptotic variances as a function of n (rather than as a limit). Let $\sigma_i^2 = E\left[\varepsilon_i^2|\mathcal{Z}\right]$, where, for notational simplicity, we have suppressed the possible dependence of σ_i^2 on $\mathcal{Z}$. Moreover, let

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

When K/r_n is bounded, the conditional asymptotic variance given $\mathcal{Z}$ of $S_n'(\tilde{\delta} - \delta_0)$ is

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

and the conditional asymptotic variance of $S_n'(\hat{\delta} - \delta_0)$ is

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

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

Theorem 2: *Suppose that 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 the 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 rate of divergence of the concentration parameter, is the usual one for estimation using 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) for LIML, Fuller's modified LIML, and B2SLS, when r_n grows at the same rate as K and slower than n, under homoskedasticity.

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

$K/r_n \longrightarrow 0$, 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 standard errors. If $K/r_n \longrightarrow \infty$, then these terms will dominate, and will slow down the convergence rate of the estimators. When $K/r_n \longrightarrow \infty$, the conditional asymptotic variance given $\mathcal{Z}$ of $\sqrt{r_n/K}S_n'(\tilde{\delta}-\delta_0)$ is

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

and the conditional asymptotic variance of $\sqrt{r_n/K}S_n'(\hat{\delta}-\delta_0)$ is

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

When $K/r_n \longrightarrow \infty$, the (conditional) 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, our 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 $K/r_n \longrightarrow \infty$. If L_n is bounded, and with probability one there is a $C > 0$ such that $\lambda_{\min}\left(L_n\bar{V}_n^*L_n'\right) \geq C$, then*

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

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

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

Here, the convergence rate is related to the size of $(\sqrt{r_n/K})S_n$. In the simple case where δ is a scalar, we can take $S_n = \sqrt{r_n}$, giving a convergence rate of $\sqrt{K}/r_n$. In this case, the theorem states that $\left(r_n/\sqrt{K}\right)(\tilde{\delta}-\delta_0)$ is asymptotically normal. It is interesting that $\sqrt{K}/r_n \longrightarrow 0$ 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 samples. Since $\sum_{i=1}^n P_{ii} = tr(P) = K$, 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

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

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

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

By the fact that $(1-P_{ii})^{-1} > 1$, we have that $\bar{V}_n^2 \geq V_n^2$ 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.

Next, we turn to results pertaining to the consistency of the asymptotic variance estimators, and to the use of these estimators in hypothesis testing. To proceed, first let z_{ig} denote the g^{th} element of z_i, and let $e_{i,n}$ be the i^{th} column of an $n \times n$ identity matrix. Further, define $z_{\cdot g} = (z_{1g}, ..., z_{ng})'$. Also, define $\check{z}_{ig} = Z_i(Z'Z)^{-1}Z'z_{\cdot g} = e_{i,n}'Pz_{\cdot g}$, so that $\check{z}_{ig}$ is the i^{th} element of the projection of $z_{\cdot g}$ on the column space of Z. Finally, define $b_{g,n} = \min\left\{\sqrt{n}/\mu_{gn}, \sqrt{n/K}\right\}$. We impose the following additional condition.

Assumption 6: *There exists a positive constant C, and a positive integer N, such that for $g = 1, ..., G$;*

$$\sup_{n \geq N} E_{\mathcal{Z}}\left[\left(\frac{1}{b_{g,n}} \max_{1 \leq i \leq n} |z_{ig}|\right)^6\right] \leq C, \quad \sup_{n \geq N} E_{\mathcal{Z}}\left[\left(\frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |\check{z}_{ig}|\right)^6\right] \leq C.$$

In addition, there exists a π_n *such that* $\max_{i\le n}\|z_i-\pi_n Z_i\| \longrightarrow 0$.

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

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

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

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

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

COROLLARY 6: *Suppose that Assumptions 1-6 are satisfied, and that* c *and* b_n *are such that* $b_n c' S_n^{-1\prime}$ *is bounded. If* K/r_n *is bounded and* $b_n^2 c' S_n^{-1\prime}\bar{V}_n S_n^{-1}c \ge C$, *or if* $K/r_n \longrightarrow \infty$ *and* $b_n^2 c' S_n^{-1\prime}\bar{V}_n^* S_n^{-1}c \ge C$, *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 $c' S_n^{-1\prime} V_n S_n^{-1} c/b_n^2 \geq C$, or if $K/r_n \longrightarrow \infty$ and $c' S_n^{-1\prime} V_n^* S_n^{-1} c/b_n^2 \geq C$, 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, under heteroskedasticity of unkown form. In the same setup, LIML, 2SLS, and B2SLS are inconsistent.

5 Appendix A - Proofs of Theorems

We will define a number of notation and abbreviations which will be used not only in this appendix but also in Appendix B. Let C denote a generic positive constant that may be different in different uses and let M, CS, and T denote the Markov inequality, the Cauchy-Schwartz inequality, and the Triangle inequality respectively. Also, for random variables W_i, Y_i, and η_i and for $\mathcal{Z} = (\Upsilon, Z)$, let $\bar{w}_i = E[W_i|\mathcal{Z}]$, $\tilde{W}_i = W_i - \bar{w}_i$, $\bar{y}_i = E[Y_i|\mathcal{Z}]$, $\tilde{Y}_i = Y_i - \bar{y}_i$, $\bar{\eta}_i = E[\eta_i|\mathcal{Z}]$, $\tilde{\eta}_i = \eta_i - \bar{\eta}_i$, $\bar{y} = (\bar{y}_1, ..., \bar{y}_n)'$, $\bar{w} = (\bar{w}_1, ..., \bar{w}_n)'$,

$$\begin{aligned} \bar{\mu}_W &= \max_{1 \le i \le n} |\bar{w}_i|, \ \bar{\mu}_Y = \max_{1 \le i \le n} |\bar{y}_i|, \ \bar{\mu}_\eta = \max_{1 \le i \le n} |\bar{\eta}_i|, \ \bar{\sigma}_W^2 = \max_{i \le n} Var[W_i|\mathcal{Z}], \\ \bar{\sigma}_Y^2 &= \max_{i \le n} Var[Y_i|\mathcal{Z}], \ \bar{\sigma}_\eta^2 = \max_{i \le n} Var[\eta_i|\mathcal{Z}]; \end{aligned}$$

where, in order to simplify notation, we have suppressed the dependence on $\mathcal{Z}$ of the various quantities ($\bar{w}_i$, $\tilde{W}_i$, $\bar{y}_i$, $\tilde{Y}_i$,$\bar{\eta}_i$, $\tilde{\eta}_i$, $\bar{\mu}_W$, $\bar{\mu}_Y$, $\bar{\mu}_\eta$, $\bar{\sigma}_W^2$, $\bar{\sigma}_Y^2$, and $\bar{\sigma}_\eta^2$) defined above. Furthermore, for random variable X, define $\|X\|_{L_2,\mathcal{Z}} = \sqrt{E[X^2|\mathcal{Z}]}$.

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, conditional on $\mathcal{Z} = (\Upsilon, Z)$, (W_i, Y_i), $(i = 1, ..., n)$ are independent with probability one and if W_i and Y_i are scalars, and P is symmetric, idempotent of rank K then for $\bar{w} = E[(W_1, ..., W_n)'|\mathcal{Z}]$, $\bar{y} = E[(Y_1, ..., Y_n)'|\mathcal{Z}]$, $\bar{\sigma}_{W_n} = \max_{i \le n} Var(W_i|\mathcal{Z})^{1/2}$, $\bar{\sigma}_{Y_n} = \max_{i \le n} Var(Y_i|\mathcal{Z})^{1/2}$, there exists a positive constant C such that

$$\left\| \sum_{i \ne j} P_{ij} W_i Y_j - \sum_{i \ne j} P_{ij} \bar{w}_i \bar{y}_j \right\|_{L_2,\mathcal{Z}}^2 \le C D_n \quad a.s.\ \mathbb{P}_{\mathcal{Z}}$$

where $D_n = K \bar{\sigma}_{W_n}^2 \bar{\sigma}_{Y_n}^2 + \bar{\sigma}_{W_n}^2 \bar{y}' \bar{y} + \bar{\sigma}_{Y_n}^2 \bar{w}' \bar{w}$.

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

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

Let $D_{1n} = \bar{\sigma}_{Wn}^2 \bar{\sigma}_{Yn}^2$. Note that for $i \ne j$ and $k \ne \ell$, $E[\tilde{w}_i \tilde{y}_j \tilde{w}_k \tilde{y}_\ell|\mathcal{Z}]$ is zero unless $i = k$ and $j = \ell$ or

$i = \ell$ and $j = k$. Then by CS and $\sum_j P_{ij}^2 = P_{ii}$,

$$\begin{aligned} E\left[\left(\sum\nolimits_{i\neq j} P_{ij}\tilde{w}_i\tilde{y}_j\right)^2 |\mathcal{Z}\right] &= \sum_{i\neq j}\sum_{k\neq \ell} P_{ij}P_{k\ell}E\left[\tilde{w}_i\tilde{y}_j\tilde{w}_k\tilde{y}_\ell|\mathcal{Z}\right] \\ &= \sum_{i\neq j} P_{ij}^2\left(E[\tilde{w}_i^2|\mathcal{Z}]E[\tilde{y}_j^2|\mathcal{Z}] + E[\tilde{w}_i\tilde{y}_i|\mathcal{Z}]E[\tilde{w}_j\tilde{y}_j|\mathcal{Z}]\right) \\ &\leq 2D_{1n}\sum_{i\neq j} P_{ij}^2 \leq 2D_{1n}\sum_i P_{ii} = 2D_{1n}K. \end{aligned}$$

Also, for $\tilde{w} = (\tilde{w}_1, ..., \tilde{w}_n)'$, 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 conditional on $\mathcal{Z}$, we have $E\left[\tilde{w}\tilde{w}'|\mathcal{Z}\right] \leq \bar{\sigma}_{W_n}^2 I_n$ $a.s.$ $P_{\mathcal{Z}}$, so that

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

Then by T we have

$$\left\|\sum\nolimits_{i\neq j} P_{ij}\tilde{w}_i\bar{y}_j\right\|_{L_2,\mathcal{Z}}^2 \leq \left\|\bar{y}'P\tilde{w}\right\|_{L_2,\mathcal{Z}}^2 + \left\|\sum\nolimits_i P_{ii}\bar{y}_i\tilde{w}_i\right\|_{L_2,\mathcal{Z}}^2 \leq C\bar{\sigma}_{W_n}^2\bar{y}'\bar{y} \;\; a.s.\ \mathbb{P}_{\mathcal{Z}}.$$

Interchanging the roles of Y_i and W_i gives $\left\|\sum\nolimits_{i\neq j} P_{ij}\bar{w}_i\tilde{y}_j\right\|_{L_2,\mathcal{Z}}^2 \leq C\bar{\sigma}_{Y_n}^2\bar{w}'\bar{w}$ $a.s.$ $P_{\mathcal{Z}}$. The conclusion then follows by T. Q.E.D.

Lemma A2: Suppose that conditional on $\mathcal{Z}$, the following conditions hold almost surely under the induced measure $P_{\mathcal{Z}}$: i) P is a symmetric, idempotent matrix with $rank(P) = K$, $P_{ii} \leq C < 1$; ii) $(W_{1n}, U_1, \varepsilon_1)$, ..., $(W_{nn}, U_n, \varepsilon_n)$ are independent and $D_n = \sum_{i=1}^n E\left[W_{in}W_{in}'|\mathcal{Z}\right]$ is bounded for n sufficiently large; iii) $E\left[W_{in}'|\mathcal{Z}\right] = 0$, $E[U_i|\mathcal{Z}] = 0$, $E[\varepsilon_i|\mathcal{Z}] = 0$ and there exists a constant C such that $E[\|U_i\|^4|\mathcal{Z}] \leq C$, $E[\varepsilon_i^4|\mathcal{Z}] \leq C$; iv) $\sum_{i=1}^n E\left[\|W_{in}\|^4|\mathcal{Z}\right] \longrightarrow 0$. Suppose further that $K \longrightarrow \infty$ as $n \to \infty$; then for $\bar{\Sigma}_n \stackrel{\text{def}}{=} \sum_{i\neq j} P_{ij}^2\left(E[U_iU_i'|\mathcal{Z}]E[\varepsilon_j^2|\mathcal{Z}] + E[U_i\varepsilon_i|\mathcal{Z}]E[\varepsilon_jU_j'|\mathcal{Z}]\right)/K$ and any bounded sequences c_{1n} and c_{2n} of conformable vectors with $\Xi_n = c_{1n}'D_nc_{1n} + c_{2n}'\bar{\Sigma}_nc_{2n} > C$, it follows that

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

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.

Lemma A3: If, conditional on $\mathcal{Z}$, $(W_i, Y_i), (i = 1, ..., n)$ are independent, and if W_i and Y_i are scalars; then there exists a positive constant C such that

$$\left\| \sum\nolimits_{i\neq j} P_{ij}^2 W_i Y_j - E\left[\sum\nolimits_{i\neq j} P_{ij}^2 W_i Y_j | \mathcal{Z}\right] \right\|_{L_2,\mathcal{Z}}^2 \leq C B_n \quad a.s.\ \mathbb{P}_{\mathcal{Z}}$$

where $B_n = K\left\{\bar{\sigma}_W^2 \bar{\sigma}_Y^2 + \bar{\sigma}_W^2 \bar{\mu}_Y^2 + \bar{\mu}_W^2 \bar{\sigma}_Y^2\right\}$.

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 $k \neq \ell$, $E\left[\tilde{w}_i \tilde{y}_j \tilde{w}_k \tilde{y}_\ell | \mathcal{Z}\right]$ 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 $\sum_j P_{ij} = P_{ii}$,

$$\begin{aligned}
E\left[\left(\sum\nolimits_{i\neq j} P_{ij}^2 \tilde{w}_i \tilde{y}_j\right)^2 | \mathcal{Z}\right] &= \sum_{i\neq j}\sum_{k\neq \ell} P_{ij}^2 P_{k\ell}^2 E\left[\tilde{w}_i \tilde{y}_j \tilde{w}_k \tilde{y}_\ell | \mathcal{Z}\right] \\
&= \sum_{i\neq j} P_{ij}^4 \left(E\left[\tilde{w}_i^2 | \mathcal{Z}\right] E\left[\tilde{y}_j^2 | \mathcal{Z}\right] + E\left[\tilde{w}_i \tilde{y}_i | \mathcal{Z}\right] E\left[\tilde{w}_j \tilde{y}_j | \mathcal{Z}\right]\right) \\
&\leq 2\bar{\sigma}_W^2 \bar{\sigma}_Y^2 \sum_{i\neq j} P_{ij}^4 \leq 2K \bar{\sigma}_W^2 \bar{\sigma}_Y^2 \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.
\end{aligned}$$

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 conditional on Z, we have $E[\tilde{w}\tilde{w}'|Z] \leq \bar{\sigma}_{Wn}^2 I_n$, so that

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

$$E\left[\left(\sum\nolimits_i P_{ii}^2 \bar{y}_i \tilde{w}_i\right)^2 | \mathcal{Z}\right] = \sum_i P_{ii}^4 E[\tilde{w}_i^2 | \mathcal{Z}] \bar{y}_i^2 \leq K \bar{\sigma}_W^2(\mathcal{Z}) \bar{\mu}_Y^2(\mathcal{Z}) \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.$$

Then by T we have $\left\|\sum_{i\neq j} P_{ij}^2 \tilde{w}_i \bar{y}_j\right\|_{L_2,\mathcal{Z}}^2 \leq \left\|\tilde{w}' \tilde{P} \bar{y}\right\|_{L_2,\mathcal{Z}}^2 + \left\|\sum_i P_{ii}^2 \bar{y}_i \tilde{w}_i\right\|_{L_2,\mathcal{Z}}^2 \leq CK \bar{\sigma}_W^2 \bar{\mu}_Y^2 \quad a.s.\ P_{\mathcal{Z}}$. Interchanging the roles of Y_i and W_i gives $\left\|\sum_{i\neq j} P_{ij}^2 \bar{w}_i \tilde{y}_j\right\|_{L_2,\mathcal{Z}}^2 = CK \bar{\mu}_W^2 \bar{\sigma}_Y^2 \quad a.s.\ P_{\mathcal{Z}}$. 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: Suppose that conditional on $\mathcal{Z}$; $(W_1, Y_1, \eta_1), ..., (W_n, Y_n, \eta_n)$ are independent, and suppose that $KE_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2 \bar{\mu}_W^2 \bar{\mu}_Y^2\right] \to 0$ as $n \to \infty$ and there exists a positive constant C such that, for n sufficiently large, $E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] \leq C$, $E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^4\right] \leq C$, $K^2 E_{\mathcal{Z}}\left[\bar{\mu}_W^4\right] \leq C$, $K^2 E_{\mathcal{Z}}\left[\bar{\mu}_Y^4\right] \leq C$. Suppose further that with probability one for n large enough, $\sum_i \bar{w}_i \leq C$, $\sum_i \bar{y}_i \leq C$,$Var(W_i|\mathcal{Z}) \leq C/r_n$, and $Var(Y_i|\mathcal{Z}) \leq C/r_n$ and that there exists π_n such that $\max_{i\leq n} |a_i - Z_i'\pi_n| \longrightarrow 0$ $a.s.$ $P_{\mathcal{Z}}$, and $\sqrt{K}/r_n \longrightarrow 0$; then

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

Proof: Given in Appendix B.

Lemma A5: If Assumptions 1-3 are satisfied; then,

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

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 $\left\|S_n^{-1}\right\| \leq C/\sqrt{r_n}$,

$$\begin{aligned} E[Y_i|\mathcal{Z}] &= z_{ik}/\sqrt{n}, Var(Y_i|\mathcal{Z}) \leq C/r_n \ a.s.\mathbb{P}_{\mathcal{Z}}, E[W_i|\mathcal{Z}] = z_{i\ell}/\sqrt{n}(1-P_{ii}), \\ Var(W_i|\mathcal{Z}) &\leq C/r_n \ a.s.\mathbb{P}_{\mathcal{Z}}. \end{aligned}$$

Note that with probabiliy one

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

Since $e_k' S_n^{-1}\tilde{H}S_n^{-1\prime}e_\ell = e_k' S_n^{-1}\sum_{i\neq j} X_i P_{ij} X_j' S_n^{-1\prime} e_\ell/(1-P_{jj}) = \sum_{i\neq j} W_i P_{ij} Y_j$ and $P_{ij}\bar{w}_i\bar{y}_j = P_{ij} z_{ik} z_{j\ell}/n(1-P_{jj})$, applying Lemma A1 and the conditional version of M, we deduce that for any $\upsilon > 0$

$$P\left(\left|e_k' S_n^{-1}\tilde{H}S_n^{-1\prime}e_\ell - \sum\nolimits_{i\neq j} e_k' z_i P_{ij}(1-P_{jj})^{-1} z_j' e_\ell/n\right| \geq \upsilon \mid \mathcal{Z}\right) \to 0 \ \ a.s. \ \mathbb{P}_{\mathcal{Z}}.$$

Moreover, note that, clearly, for some $\epsilon > 0$

$$\sup_n E\left[\left|P\left(\left|e_k' S_n^{-1}\tilde{H}S_n^{-1\prime}e_\ell - \sum\nolimits_{i\neq j} e_k' z_i P_{ij}(1-P_{jj})^{-1} z_j' e_\ell/n\right| \geq \upsilon \mid \mathcal{Z}\right)\right|^{1+\epsilon}\right] < \infty,$$

(take, for example, $\epsilon = 1$). Hence, by a version of the dominated convergence theorem, as given by Theorem 25.12 of Billingsley (1986), it follows that as $n \to \infty$,

$$
\begin{aligned}
& P\left(\left|e_k' S_n^{-1} \tilde{H} S_n^{-1\prime} e_\ell - \sum\nolimits_{i \neq j} e_k' z_i P_{ij}(1-P_{jj})^{-1} z_j' e_\ell / n\right| \geq \upsilon\right) \\
= \;& E\left[P\left(\left|e_k' S_n^{-1} \tilde{H} S_n^{-1\prime} e_\ell - \sum\nolimits_{i \neq j} e_k' z_i P_{ij}(1-P_{jj})^{-1} z_j' e_\ell / n\right| \geq \upsilon \mid \mathcal{Z}\right)\right] \\
\to \;& 0.
\end{aligned}
$$

The above argument establishes the first conclusion for the (k, ℓ) element. Now, doing this for every element completes the proof for the first conclusion.

For the second conclusion, apply Lemma A1 with $Y_i = e_k' S_n^{-1} X_i$ as before and $W_i = \varepsilon_i / \sqrt{r_n}(1 - P_{ii})$ and follow the same argument as that used to establish the first conclusion above.

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 with probability one

$$
\begin{aligned}
\sqrt{K} \bar{\sigma}_{W_n} \bar{\sigma}_{Y_n} & \leq C\sqrt{K}/r_n \longrightarrow 0, \bar{\sigma}_{W_n}\sqrt{\bar{y}'\bar{y}} \leq C r_n^{-1/2} \sqrt{\sum z_{ik}^2 / n} \longrightarrow 0, \\
\bar{\sigma}_{Y_n} \sqrt{\bar{w}'\bar{w}} & \longrightarrow 0.
\end{aligned}
$$

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

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

Lemma A6: If Assumptions 1-4 are satisfied; then,

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

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

$$
\begin{aligned}
\sum_{i=1}^n \|z_i - \bar{z}_i\|^2 / n &= \|(I-P)z\|^2 / n = tr(z'(I-P)z/n) = tr[(z - Z\pi_{Kn}')'(I-P)(z - Z\pi_{Kn}')/n] \\
&\leq tr[(z - Z\pi_{Kn}')'(z - Z\pi_{Kn}')/n] = \sum_{i=1}^n \|z_i - \pi_{Kn} Z_i\|^2 / n \longrightarrow 0 \;\; a.s.\ \mathbb{P}_{\mathcal{Z}}.
\end{aligned}
$$

It follows that with probability one

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

Then

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

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

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

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

<u>Proof of Theorem 1</u>: First, note that by $\lambda_{\min}\left(S_n S_n'/r_n\right) \geq \lambda_{\min}\left(\tilde{S}_n \tilde{S}_n'\right) \geq C$ 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 approaching one $\lambda_{\min}(S_n^{-1}\tilde{H}S_n^{-1\prime}) \geq C$ as the sample size grows. Hence $\left(S_n^{-1}\tilde{H}S_n^{-1\prime}\right)^{-1} = O_p(1)$. By eq. (1),

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

All of the previous statements are conditional on $Z = (\Upsilon, Z)$ 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 costant $v > 0$, 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,

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

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 $E\left[\xi_i^2 | \mathcal{Z}\right] \leq C \;\; a.s.\ P_{\mathcal{Z}}$, so that

$$\begin{aligned} E\left[\left\|\sum_{i=1}^n (z_i - \bar{z}_i)\, \xi_i / \sqrt{n}\right\|^2 | \mathcal{Z}\right] &= \sum_{i=1}^n \|z_i - \bar{z}_i\|^2 E\left[\xi_i^2 | \mathcal{Z}\right] / n \\ &\leq C \sum_{i=1}^n \|z_i - \bar{z}_i\|^2 / n \longrightarrow 0 \;\; a.s.\ \mathbb{P}_{\mathcal{Z}}. \end{aligned}$$

Note further that

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

so that by a conditional version of M, we deduce that for any $\upsilon > 0$

$$\begin{aligned} & P\left(\left\|S_n^{-1} \sum_{i \neq j} X_i P_{ij} \xi_j - \sum_i z_i (1 - P_{ii}) \xi_i / \sqrt{n} - S_n^{-1} \sum_{i \neq j} U_i P_{ij} \xi_j\right\| \geq \upsilon \mid \mathcal{Z}\right) \\ = \;& P\left(\left\|\sum_{i=1}^n (z_i - \bar{z}_i)\, \xi_i / \sqrt{n}\right\| \geq \upsilon \mid \mathcal{Z}\right) \to 0 \;\; a.s.\ \mathbb{P}_{\mathcal{Z}}. \end{aligned}$$

Moreover, note that there exists $\epsilon > 0$ such that

$$\sup_n E\left[\left|P\left(\left\|\sum_{i=1}^n (z_i - \bar{z}_i)\, \xi_i / \sqrt{n}\right\| \geq \upsilon \mid \mathcal{Z}\right)\right|^{1+\epsilon}\right] < \infty,$$

(take, for example, $\epsilon = 1$). Hence, by a version of the dominated convergence theorem, as given by Theorem 25.12 of Billingsley (1986), it follows that as $n \to \infty$,

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

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

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

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 the almost sure boundedness of $\sum_i z_i z_i'/n$, Assumption 3, and the boundedness of K/μ_n^2, it follows that Γ_n is bounded with probability one for n sufficiently large. Also, $E[\xi_i^2|Z] \geq C > 0$, 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 $\lambda_{\min}(\Gamma_n) \geq C > 0$ with probability one for all n large enough. It follows that Γ_n^{-1} exists and is bounded in n with probability one 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, $W_{in} = z_i(1-P_{ii})\xi_i/\sqrt{n}$, $c_{1n} = \Gamma_n^{-1/2}\alpha$, and $c_{2n} = \sqrt{K} S_n^{-1}\Gamma_n^{-1/2}\alpha$. Note that condition i) of Lemma A2 is satisfied. Also, by the boundedness of $\sum_i z_i z_i'/n$ and $E[\xi_i^2|Z]$ with probability one, condition ii) of Lemma A2 is satsifed and condition iii) is satisfied by Assumptions 3 and 5. Also, by $(1-P_{ii})^{-1} \leq C$ and Assumption 5,

$$\sum_{i=1}^n E\left[\|W_{in}\|^4 |\mathcal{Z}\right] \leq C \sum_{i=1}^n \|z_i\|^4 /n^2 \longrightarrow 0 \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.$$

so condition iv) is satisfied. Finally, condition v) is satsfied by hypothesis. Note also that with probability one $c_{1n} = \Gamma_n^{-1/2}\alpha$ and $c_{2n} = \left(\sqrt{K}/\mu_n\right)\mu_n S_n^{-1}\Gamma_n^{-1/2}\alpha$ are bounded by the boundedness of $\sqrt{K}/\mu_n$, $\mu_n S_n^{-1}$, and Γ_n^{-1}. Moreover, $\Xi_n = \alpha'\alpha$ by construction and, thus, Ξ_n is also bounded and $\Xi_n \geq C > 0$. Then by the conclusion of Lemma A2,

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

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

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

Note that $B_n = \bar{V}_n^{-1/2}\bar{H}_n^{-1}\Gamma_n^{1/2}$ is an orthogonal matrix, since $B_n B_n' = \bar{V}_n^{-1/2}\bar{V}_n\bar{V}_n^{-1/2\prime} = I$. Also, $\bar{V}_n^{-1/2}$ is bounded by $\lambda_{\min}(\bar{V}_n) \geq C > 0$, and $\Gamma_n^{1/2}$ is also bounded by Γ_n bounded. By Lemma A6,

$(S_n^{-1}\tilde{H}S_n^{-1\prime})^{-1} = \bar{H}_n^{-1} + o_p(1)$. Therefore, we have

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

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 Slutzky 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

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

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, $r_n/K \longrightarrow 0$. 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

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

Note that with probability one, Γ_n is bounded given that $\mu_n S_n^{-1}$ is bounded, $E[\|U_i\|^2|Z] \leq C$, and $E[\xi_j^2|Z] \leq C$ $a.s.\ P_{\mathcal{Z}}$. 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

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

Now let α be a nonzero vector and apply Lemma A2 with $W_{in} = 0$, $\varepsilon_i = \xi_i$, $c_{1n} = 0$, and $c_{2n} = \alpha'\left(\bar{L}_n\Gamma_n\bar{L}_n'\right)^{-1/2}\bar{L}_n\mu_n S_n^{-1}$. We have $c_{2n}'Var\left(\sum_{i\neq j} U_i P_{ij}\xi_j/\sqrt{K}\right)c_{2n} = \alpha'\alpha > 0$ 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

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

Let L_n be as specified in the statement of the result such that $\lambda_{\min}\left(L_n \bar{V}_n^* L_n'\right) \geq C > 0$ with probability one 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, $\left(\bar{L}_n \Gamma_n \bar{L}_n'\right)^{-1/2}$ and $\Gamma_n^{1/2}$ are bounded with probability one. By Lemma A6, $(S_n^{-1} \tilde{H} S_n^{-1\prime})^{-1} = \bar{H}_n^{-1} + o_p(1)$. Therefore, we have

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

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

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

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, we need to verify the conditions of Lemma A4. To proceed, note first 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

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

Let η_i be any element $-2\xi_i X_i^{P\prime}$ or of $X_i^P X_i^{P\prime}$. Also, let $\widetilde{s}_{kg}$ be the $(k,g)^{th}$ element of $\tilde{S}_n$, and let $\Upsilon_{i,k} = \sum_{g=1}^G \widetilde{s}_{kg} \mu_{gn} z_{ig}/\sqrt{n}$ and $u_{i,k}$ denote the k^{th} element of Υ_i and U_i, respectively. Note that if

$\eta_i = -2\xi_i X_i^{P\prime} e_k$, then

$$\begin{aligned} E[\eta_i|\mathcal{Z}] &= \frac{E[\varepsilon_i u_{i,k}|\mathcal{Z}]}{1\{JIV1\}(1-P_{ii})^2 + (1 - 1\{JIV1\})}, \\ Var[\eta_i|\mathcal{Z}] &= \frac{\Upsilon_{i,k}^2 E[\varepsilon_i^2|\mathcal{Z}] + 2\Upsilon_{i,k} E[\varepsilon_i^2 u_{i,k}|\mathcal{Z}] + Var(\varepsilon_i u_{i,k}|\mathcal{Z})}{1\{JIV1\}(1-P_{ii})^4 + (1 - 1\{JIV1\})}, \end{aligned}$$

where $1\{JIV1\}$ is an indicator function which takes on the value of 1 for the case of $JIV1$ and 0 for the case of $JIV2$. On the other hand, if $\eta_i = e_k' X_i^P X_i^{P\prime} e_\ell$, then

$$\begin{aligned} E[\eta_i|\mathcal{Z}] &= \frac{\Upsilon_{i,k}\Upsilon_{i,\ell} + E[u_{i,k}u_{i,\ell}|\mathcal{Z}]}{1\{JIV1\}(1-P_{ii})^2 + (1 - 1\{JIV1\})}, \\ Var[\eta_i|\mathcal{Z}] &= \left[1\{JIV1\}(1-P_{ii})^4 + (1 - 1\{JIV1\})\right]^{-1} \\ &\quad \times \left\{\Upsilon_{i,k}^2 E[u_{i,\ell}^2|\mathcal{Z}] + 2\Upsilon_{i,k}\Upsilon_{i,\ell} E[u_{i,k}u_{i,\ell}|\mathcal{Z}] + \Upsilon_{i,\ell}^2 E[u_{i,k}^2|\mathcal{Z}]\right. \\ &\quad \left. + 2\Upsilon_{i,k} E[u_{i,k}u_{i,\ell}^2|\mathcal{Z}] + 2\Upsilon_{i,\ell} E[u_{i,k}^2 u_{i,\ell}|\mathcal{Z}] + Var(u_{i,k}u_{i,\ell}|\mathcal{Z})\right\}. \end{aligned}$$

Now, since $\tilde{S}_n$ is bounded by Assumption 2, we have that

$$\max_{1\le i\le n} |\Upsilon_{i,k}| = \max_{1\le i\le n} \left|\sum_{g=1}^{G} \widetilde{s}_{kg}\mu_{gn} z_{ig}/\sqrt{n}\right| \le C \sum_{g=1}^{G} \frac{\mu_{gn}}{\sqrt{n}} \max_{1\le i\le n} |z_{ig}|$$

Moreover, with probability one $\sup_{i,k} E\left[u_{i,k}^4|\mathcal{Z}\right] \le C < \infty$, $\sup_i E\left[\varepsilon_i^4|\mathcal{Z}\right] \le C < \infty$, and $P_{ii} \le C < 1$, so that whether $\eta_i = -2\xi_i X_i^{P\prime} e_k$ or $e_k' X_i^P X_i^{P\prime} e_\ell$, we can obtain overall upper bounds for the conditional mean and variance of the form:

$$\begin{aligned} \bar{\mu}_\eta(\mathcal{Z}) &= \max_{1\le i\le n} |E[\eta_i|\mathcal{Z}]| \le C \sum_{g=1}^{G} \left(\frac{\mu_{gn}}{\sqrt{n}} \max_{1\le i\le n} |z_{ig}|\right) + C \\ \bar{\sigma}_\eta^2(\mathcal{Z}) &= \max_{1\le i\le n} Var[\eta_i|\mathcal{Z}] \le C \sum_{g=1}^{G} \left(\frac{\mu_{gn}}{\sqrt{n}} \max_{1\le i\le n} |z_{ig}|\right)^2 + C \sum_{g=1}^{G} \left(\frac{\mu_{gn}}{\sqrt{n}} \max_{1\le i\le n} |z_{ig}|\right) + C \end{aligned}$$

It follows in this case that

$$\begin{aligned} E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2(\mathcal{Z})\right] &\le C \sum_{g=1}^{G} E_{\mathcal{Z}} \left(\frac{\mu_{gn}}{\sqrt{n}} \max_{1\le i\le n} |z_{ig}|\right)^2 + C \le C, \\ E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^4(\mathcal{Z})\right] &\le C \sum_{g=1}^{G} E_{\mathcal{Z}} \left(\frac{\mu_{gn}}{\sqrt{n}} \max_{1\le i\le n} |z_{ig}|\right)^4 + C \sum_{g=1}^{G} E_{\mathcal{Z}} \left(\frac{\mu_{gn}}{\sqrt{n}} \max_{1\le i\le n} |z_{ig}|\right)^2 + C \le C. \end{aligned}$$

Next, take $W_i = e_k' \dot{X}_i$ and $Y_j = e_\ell' \dot{X}_j$ and let $\widetilde{u}_{i,k} = e_k' S_n^{-1} U_i$, and note that in this case

$$\begin{aligned} \bar{w}_i &= \frac{1}{\sqrt{n}} |z_{ik}|, \ \bar{y}_i = \frac{1}{\sqrt{n}} |z_{i\ell}| \\ \bar{\mu}_W(\mathcal{Z}) &= \max_{1\le i\le n} |\bar{w}_i| = \frac{1}{\sqrt{n}} \max_{1\le i\le n} |z_{ik}|, \ \bar{\mu}_Y(\mathcal{Z}) = \max_{1\le i\le n} |\bar{y}_i| = \frac{1}{\sqrt{n}} \max_{1\le i\le n} |z_{i\ell}|, \end{aligned}$$

so that, by Assumption 2, we have that for n sufficiently large, we have that with probability one

$$\sum_{i=1}^{n} \bar{w}_i = \frac{1}{n}\sum_{i=1}^{n} z_{ik}^2 \leq C, \ \ \sum_{i=1}^{n} \bar{y}_i = \frac{1}{n}\sum_{i=1}^{n} z_{i\ell}^2 \leq C.$$

Moreover, for n sufficiently large,

$$\begin{aligned} K^2 E_{\mathcal{Z}}\left[\bar{\mu}_W^4\left(\mathcal{Z}\right)\right] &\leq E_{\mathcal{Z}}\left(\sqrt{\frac{K}{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^4 \leq C, \ K^2 E_{\mathcal{Z}}\left[\bar{\mu}_Y^4\left(\mathcal{Z}\right)\right] \leq E_{\mathcal{Z}}\left(\sqrt{\frac{K}{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^4 \leq C, \\ Var\left(W_i|\mathcal{Z}\right) &= E\left[\widetilde{u}_{i,k}^2|\mathcal{Z}\right] \leq C/r_n \ \ a.s.\ \mathbb{P}_{\mathcal{Z}}, \ Var\left(Y_j|\mathcal{Z}\right) = E\left[\widetilde{u}_{i,k}^2|\mathcal{Z}\right] \leq C/r_n \ \ a.s.\ \mathbb{P}_{\mathcal{Z}}, \end{aligned}$$

and

$$\begin{aligned} & K E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\left(\mathcal{Z}\right)\bar{\mu}_W^2\left(\mathcal{Z}\right)\bar{\mu}_Y^2\left(\mathcal{Z}\right)\right] \\ \leq\ & C\sum_{g=1}^{G} E_{\mathcal{Z}}\left[\frac{1}{K^{1/3}}\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ig}|\right)^2\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^2\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^2\right] \\ & +C\sum_{g=1}^{G} E_{\mathcal{Z}}\left[\frac{1}{K^{1/3}}\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ig}|\right)\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^2\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^2\right] \\ & +C E_{\mathcal{Z}}\left[\left(\frac{K^{1/4}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^2\left(\frac{K^{1/4}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^2\right] \\ \leq\ & C\sum_{g=1}^{G} E_{\mathcal{Z}}\left[\frac{1}{3K^{1/3}}\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ig}|\right)^2+\frac{1}{3}\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^2+\frac{1}{3}\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^2\right]^3 \\ & +C\sum_{g=1}^{G} E_{\mathcal{Z}}\left[\frac{1}{3K^{1/3}}\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ig}|\right)+\frac{1}{3}\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^2+\frac{1}{3}\left(\frac{K^{1/3}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^2\right]^3 \\ & +C E_{\mathcal{Z}}\left(\frac{K^{1/4}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^4 + C E_{\mathcal{Z}}\left(\frac{K^{1/4}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^4 \\ \leq\ & C\sum_{g=1}^{G}\frac{1}{K^{1/3}}E_{\mathcal{Z}}\left[\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ig}|\right)^6\right]+C\frac{1}{K}E_{\mathcal{Z}}\left(\sqrt{\frac{K}{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^6+C\frac{1}{K}E_{\mathcal{Z}}\left(\sqrt{\frac{K}{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^6 \\ & +\sum_{g=1}^{G}\frac{1}{K^{1/3}}E_{\mathcal{Z}}\left[\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ig}|\right)^3\right]+C\frac{1}{K}E_{\mathcal{Z}}\left(\sqrt{\frac{K}{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^4+C\frac{1}{K}E_{\mathcal{Z}}\left(\sqrt{\frac{K}{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^4 \\ =\ & O\left(K^{-1/3}\right) = o\left(1\right), \end{aligned}$$

where the second inequality above follows from the arithmetic-geometric mean inequality and the third inequality follows from Loève's c_r inequality. Applying Lemma A4, we deduce that $\sum_{i\neq j\neq k} e_k'\dot{X}_i P_{ik}\eta_k P_{kj}\dot{X}_j' e_\ell =$

$O_p(1)$. Doing this for every element, we then get $\sum_{i\neq j\neq k}\dot{X}_iP_{ik}\eta_kP_{kj}\dot{X}_j' = O_p(1)$. Next, let $\hat{\Delta}_n$ denote a sequence of random variables converging to zero in probability, and it follows that

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

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 $\hat{\Delta}\sum_{i\neq j\neq k}\dot{X}_iP_{ik}\eta_kP_{kj}\dot{X}_j'$, so by the triangle inequality $\hat{\Sigma}_1 - \dot{\Sigma}_1 \xrightarrow{p} 0$.

Let $d_i = \sum_{g=1}^{G}\mu_{gn}\max_{1\leq i\leq n}|z_{i,g}|/\sqrt{n} + |\varepsilon_i| + \|U_i\|$, $\hat{A} = (1+\left\|\hat{\delta}\right\|)$ for JIV1, $\hat{A} = (1+\left\|\tilde{\delta}\right\|)$ for JIV2, $\hat{B} = \left\|\hat{\delta}-\delta_0\right\|$ for JIV1, and $\hat{B} = \left\|\tilde{\delta}-\delta_0\right\|$ 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,

$$\begin{aligned}
\|X_i\| &\leq \|\Upsilon_i\| + \|U_i\| \leq C\sqrt{\sum_{g=1}^{G}\left(\frac{\mu_{gn}^2}{n}\right)z_{i,g}^2} + \|U_i\| \leq C\sum_{g=1}^{G}\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i,g}|\right) + \|U_i\| \leq Cd_i \\
,\left\|\dot{X}_i\right\| &\leq Cr_n^{-1/2}d_i,\ \left|\hat{\xi}_i - \xi_i\right| \leq C\left|X_i'(\hat{\delta}-\delta_0)\right| \leq Cd_i\hat{B}, \\
\left|\hat{\xi}_i\right| &\leq C\left|X_i'(\delta_0-\hat{\delta})\right| + |\xi_i| \leq Cd_i\hat{A}, \\
\left|\hat{\xi}_i^2 - \xi_i^2\right| &\leq \left(|\xi_i| + \left|\hat{\xi}_i\right|\right)\left|\hat{\xi}_i - \xi_i\right| \leq Cd_i(1+\hat{A})d_i\hat{B} \leq Cd_i^2\hat{A}\hat{B}, \\
\left\|\dot{X}_i\left(\hat{\xi}_i - \xi_i\right)\right\| &\leq Cr_n^{-1/2}d_i^2\hat{B},\ \left\|\dot{X}_i\hat{\xi}_i\right\| \leq Cr_n^{-1/2}d_i^2\hat{A},\ \left\|\dot{X}_i\xi_i\right\| \leq Cr_n^{-1/2}d_i^2.
\end{aligned}$$

Also note that

$$\begin{aligned}
\max_{1\leq i\leq n}E\left[d_i^4|\mathcal{Z}\right] &= \max_{1\leq i\leq n}E\left[\left(\sum_{g=1}^{G}\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i,g}|\right) + |\varepsilon_i| + \|U_i\|\right)^4 \mid \mathcal{Z}\right] \\
&\leq C\left[\sum_{g=1}^{G}\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i,g}|\right)^4 + \max_{1\leq i\leq n}E\left[|\varepsilon_i|^4|\mathcal{Z}\right] + \max_{1\leq i\leq n}E\left[\|U_i\|^4|\mathcal{Z}\right]\right],
\end{aligned}$$

and, thus, we have

$$
\begin{aligned}
& E\left[\sum_{i\neq j} P_{ij}^2 d_i^2 d_j^2 r_n^{-1}\right] \\
\leq\ & C E_{\mathcal{Z}}\left[\sum_{i\neq j} P_{ij}^2 E\left[d_i^4|\mathcal{Z}\right] r_n^{-1} + \sum_{i\neq j} P_{ij}^2 E\left[d_j^4|\mathcal{Z}\right] r_n^{-1}\right] \\
\leq\ & C E_{\mathcal{Z}}\left[\max_{1\leq i\leq n} E\left[d_i^4|\mathcal{Z}\right] \sum_{i,j} P_{ij}^2 r_n^{-1}\right] \\
=\ & C E_{\mathcal{Z}}\left[\max_{1\leq i\leq n} E\left[d_i^4|\mathcal{Z}\right] \sum_{i} P_{ii} r_n^{-1}\right] \\
=\ & C\left(\frac{K}{r_n}\right) E_{\mathcal{Z}}\left[\max_{1\leq i\leq n} E\left[d_i^4|\mathcal{Z}\right]\right] \\
\leq\ & C\left(\frac{K}{r_n}\right)\left[\sum_{g=1}^{G} E_{\mathcal{Z}}\left\{\left(\frac{\mu_{gn}}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i,g}|\right)^4\right\} + \ E_{\mathcal{Z}}\left\{\max_{1\leq i\leq n} E\left[|\varepsilon_i|^4|\mathcal{Z}\right]\right\} + E_{\mathcal{Z}}\left\{\max_{1\leq i\leq n} E\left[\|U_i\|^4|\mathcal{Z}\right]\right\}\right] \\
\leq\ & CK/r_n,
\end{aligned}
$$

where the third inequality above follows from Assumptions . Hence, we deduce by the Markov inequality that $\sum_{i\neq j} P_{ij}^2 d_i^2 d_j^2 r_n^{-1} = O_p(K/r_n)$. It then follows that

$$
\left\|\sum_{i\neq j} P_{ij}^2\left(\dot{X}_i\dot{X}_i'\left(\hat{\xi}_j^2-\xi_j^2\right)\right)\right\| \leq \sum_{i\neq j} P_{ij}^2\left\|\dot{X}_i\right\|^2\left|\hat{\xi}_j^2-\xi_j^2\right| \leq C r_n^{-1}\sum_{i\neq j} P_{ij}^2 d_i^2 d_j^2 \hat{A}\hat{B} = o_p\left(K/r_n\right).
$$

We also have

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

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

Lemma A8: If Assumptions 1-6 are satisfied then

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

where $\epsilon_n = \max\left\{K/r_n^2, 1/K\right\}$.

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 $Var(\xi_i^2|\mathcal{Z}) \leq C\ a.s.\ P_{\mathcal{Z}}$ and $r_n \leq Cn$, so that for $\widetilde{u}_{i,k} = e_k' S_n^{-1} U_i$,

$$\begin{aligned}
E[(\dot{X}_{ik}\dot{X}_{i\ell})^2|\mathcal{Z}] &\leq CE[\dot{X}_{ik}^4 + \dot{X}_{i\ell}^4|\mathcal{Z}] \leq C\left\{z_{ik}^4/n^2 + E[\widetilde{u}_{i,k}^4|\mathcal{Z}] + z_{i\ell}^4/n^2 + E[\widetilde{u}_{i,\ell}^4|\mathcal{Z}]\right\} \\
&\leq C\left[\left(\frac{1}{\sqrt{n}}\max_{1\leq i\leq n}|z_{ik}|\right)^4 + \left(\frac{1}{\sqrt{n}}\max_{1\leq i\leq n}|z_{i\ell}|\right)^4 + r_n^{-2}\right] \quad a.s.\ \mathbb{P}_{\mathcal{Z}}, \\
E[(\dot{X}_{ik}\xi_i)^2|\mathcal{Z}] &\leq CE[(z_{ik}^2\xi_i^2/n + \widetilde{u}_{i,k}^2\xi_i^2)|\mathcal{Z}] \\
&\leq C\left[\sqrt{\left(n^{-1/2}\max_{1\leq i\leq n}|z_{ik}|\right)^4}\sqrt{E[\xi_i^4|\mathcal{Z}]} + \sqrt{E[\widetilde{u}_{i,k}^4|\mathcal{Z}]}\sqrt{E[\xi_i^4|\mathcal{Z}]}\right] \\
&\leq C\left[\sqrt{\left(n^{-1/2}\max_{1\leq i\leq n}|z_{ik}|\right)^4} + r_n^{-1}\right] \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.
\end{aligned}$$

Also, we have, for $\Omega_i(\mathcal{Z}) = E[U_iU_i'|\mathcal{Z}]$,

$$E\left[\dot{X}_i\dot{X}_i'|\mathcal{Z}\right] = z_iz_i'/n + S_n^{-1}\Omega_i(\mathcal{Z})S_n^{-1\prime}, E\left[\dot{X}_i\xi_i|\mathcal{Z}\right] = S_n^{-1}E\left[U_i\xi_i|\mathcal{Z}\right].$$

Next let W_i be $e_k'\dot{X}_i\dot{X}_i'e_\ell$ for some k and ℓ, and note that

$$E\left[W_i|\mathcal{Z}\right] = e_k'S_n^{-1}E\left[U_iU_i'|\mathcal{Z}\right]S_n^{-1\prime}e_\ell + z_{ik}z_{i\ell}/n,$$

so that

$$\max_{1\leq i\leq n}|E[W_i|\mathcal{Z}]| \leq C\left[\left(n^{-1/2}\max_{1\leq i\leq n}|z_{ik}|\right)^2 + \left(n^{-1/2}\max_{1\leq i\leq n}|z_{i\ell}|\right)^2 + r_n^{-1}\right] \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.$$

Moreover,

$$\begin{aligned}
\max_{1\leq i\leq n}Var\left(W_i|\mathcal{Z}\right) &= \max_{1\leq i\leq n}\left[Var\left\{\left(\widetilde{u}_{i,k} + z_{ik}/\sqrt{n}|\mathcal{Z}\right)\left(\widetilde{u}_{i,\ell} + z_{i\ell}/\sqrt{n}\right)|\mathcal{Z}\right\}\right] \\
&= \max_{1\leq i\leq n}\left\{n^{-1}z_{ik}^2E\left[\widetilde{u}_{i,\ell}^2|\mathcal{Z}\right] + 2n^{-1}z_{ik}z_{i\ell}E\left[\widetilde{u}_{i,k}\widetilde{u}_{i,\ell}|\mathcal{Z}\right] + n^{-1}z_{i\ell}^2E\left[\widetilde{u}_{i,k}^2|\mathcal{Z}\right]\right| \\
&\quad +2n^{-1/2}z_{ik}E\left[\widetilde{u}_{i,k}\widetilde{u}_{i,\ell}^2|\mathcal{Z}\right] + 2n^{-1/2}z_{i\ell}E\left[\widetilde{u}_{i,\ell}\widetilde{u}_{i,k}^2|\mathcal{Z}\right] + Var\left(\widetilde{u}_{i,k}\widetilde{u}_{i,\ell}|\mathcal{Z}\right)\Big\} \\
&\leq C\left[r_n^{-1}\left(n^{-1/2}\max_{1\leq i\leq n}|z_{ik}|\right)^2 + r_n^{-1}\left(n^{-1/2}\max_{1\leq i\leq n}|z_{i\ell}|\right)^2 + r_n^{-3/2}\left(n^{-1/2}\max_{1\leq i\leq n}|z_{ik}|\right)\right. \\
&\quad \left. +r_n^{-3/2}\left(n^{-1/2}\max_{1\leq i\leq n}|z_{i\ell}|\right) + r_n^{-2}\right] \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.
\end{aligned}$$

Also let $Y_i = \xi_i^2$. Then, applying Lemma A3 for this W_i and Y_i gives that with probability one in $P_{\mathcal{Z}}$

$$
\begin{aligned}
& \left\| \sum\nolimits_{i \neq j} P_{ij}^2 e_k' \dot{X}_i \dot{X}_i' e_\ell \xi_j^2 - \sum_{i \neq j} P_{ij}^2 \left(z_{ik} z_{i\ell}/n + S_n^{-1} \Omega_i(\mathcal{Z}) S_n^{-1\prime} \right) E\left[\xi_j^2 | \mathcal{Z} \right] \right\|_{L_2, \mathcal{Z}}^2 \\
\leq \;& CK(\bar{\sigma}_{W_n}^2 \bar{\sigma}_{Y_n}^2 + \bar{\sigma}_{W_n}^2 \bar{\mu}_{Y_n}^2 + \bar{\mu}_{W_n}^2 \bar{\sigma}_{Y_n}^2) \\
\leq \;& C \left\{ \frac{1}{K} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{ik}| \right)^4 + \frac{1}{K} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{i\ell}| \right)^4 + r_n^{-1} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{ik}| \right)^2 \right. \\
& \left. + r_n^{-1} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{i\ell}| \right)^2 + \frac{\sqrt{K}}{r_n^{3/2}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{ik}| \right) + \frac{\sqrt{K}}{r_n^{3/2}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{i\ell}| \right) + \frac{K}{r_n^2} \right\}.
\end{aligned}
$$

Taking expectation with respect to the distribution of $\mathcal{Z}$ and theorem 16.1 of Billingsley (1986), we get

$$
\begin{aligned}
& E \left(\sum\nolimits_{i \neq j} P_{ij}^2 e_k' \dot{X}_i \dot{X}_i' e_\ell \xi_j^2 - \sum_{i \neq j} P_{ij}^2 \left(z_{ik} z_{i\ell}/n + e_k' S_n^{-1} \Omega_i(\mathcal{Z}) S_n^{-1\prime} e_\ell \right) E\left[\xi_j^2 | \mathcal{Z} \right] \right)^2 \\
= \;& E_{\mathcal{Z}} \left\| \sum\nolimits_{i \neq j} P_{ij}^2 e_k' \dot{X}_i \dot{X}_i' e_\ell \xi_j^2 - \sum_{i \neq j} P_{ij}^2 \left(z_{ik} z_{i\ell}/n + S_n^{-1} \Omega_i(\mathcal{Z}) S_n^{-1\prime} \right) E\left[\xi_j^2 | \mathcal{Z} \right] \right\|_{L_2, \mathcal{Z}}^2 \\
\leq \;& C \left\{ \frac{1}{K} E_{\mathcal{Z}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{ik}| \right)^4 + \frac{1}{K} E_{\mathcal{Z}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{i\ell}| \right)^4 + r_n^{-1} E_{\mathcal{Z}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{ik}| \right)^2 \right. \\
& \left. + r_n^{-1} E_{\mathcal{Z}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{i\ell}| \right)^2 + \frac{\sqrt{K}}{r_n^{3/2}} E_{\mathcal{Z}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{ik}| \right) + \frac{\sqrt{K}}{r_n^{3/2}} E_{\mathcal{Z}} \left(\sqrt{\frac{K}{n}} \max_{1 \leq i \leq n} |z_{i\ell}| \right) + \frac{K}{r_n^2} \right\} \\
= \;& O(\epsilon_n) = o(1),
\end{aligned}
$$

where $\epsilon_n = \max\left\{ K/r_n^2, 1/K \right\}$. It follows by M that

$$
\sum_{i \neq j} P_{ij}^2 e_k' \dot{X}_i \dot{X}_i' e_\ell \xi_j^2 = \sum_{i \neq j} P_{ij}^2 \left(z_{ik} z_{i\ell}/n + e_k' S_n^{-1} \Omega_i(\mathcal{Z}) S_n^{-1\prime} e_\ell \right) E\left[\xi_j^2 | \mathcal{Z} \right] + O_p(\sqrt{\epsilon_n}).
$$

Do this for every element of $\dot{X}_i \dot{X}_i'$ then yields the representation

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

Now, by a similar argument using Lemma A3 and taking W_i and Y_i to be elements of $\dot{X}_i \xi_i$, we can show that

$$
\sum_{i \neq j} P_{ij}^2 \dot{X}_i \xi_i \xi_j \dot{X}_j' = S_n^{-1} \sum_{i \neq j} P_{ij}^2 E\left[U_i \xi_i | \mathcal{Z} \right] E\left[\xi_j U_j' | \mathcal{Z} \right] S_n^{-1\prime} + O_p(\sqrt{\epsilon_n}).
$$

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

Proof of Theorem 4: Note that $\bar{X}_i = \sum_{j=1}^n P_{ij} X_j$, so that

$$
\begin{aligned}
& \sum_{i=1}^n (\bar{X}_i \bar{X}_i' - \hat{X}_i P_{ii} \bar{X}_i' - \bar{X}_i P_{ii} X_i') \hat{\xi}_i^2 \\
= & \sum_{i,j,k=1}^n P_{ik} P_{kj} X_i X_j' \hat{\xi}_k^2 - \sum_{i,j=1}^n P_{ii} P_{ij} X_i X_j' \hat{\xi}_i^2 - \sum_{i,j=1}^n P_{ij} P_{jj} X_i X_j' \hat{\xi}_j^2 \\
= & \sum_{i,j,k=1}^n P_{ik} P_{kj} X_i X_j' \hat{\xi}_k^2 - \sum_{i \neq j} P_{ii} P_{ij} X_i X_j' \hat{\xi}_i^2 - \sum_{i \neq j} P_{ij} P_{jj} X_i X_j' \hat{\xi}_j^2 - 2 \sum_{i=1}^n P_{ii}^2 X_i X_i' \hat{\xi}_i^2 \\
= & \sum_{i,j,k \notin \{i,j\}} P_{ik} P_{kj} X_i X_j' \hat{\xi}_k^2 - \sum_{i=1}^n P_{ii}^2 X_i X_i' \hat{\xi}_i^2 \\
= & \sum_{i \neq j \neq k} P_{ik} P_{kj} X_i X_j' \hat{\xi}_k^2 + \sum_{i \neq j}^n P_{ij}^2 X_i X_i' \hat{\xi}_j^2 - \sum_{i=1}^n P_{ii}^2 X_i X_i' \hat{\xi}_i^2.
\end{aligned}
$$

Also, for Z_i' and $\tilde{Z}_i'$ equal to the ith row of Z and $\tilde{Z} = Z(Z'Z)^{-1}$ we have

$$
\begin{aligned}
& \sum_{k=1}^K \sum_{\ell=1}^K \left(\sum_{i=1}^n \tilde{Z}_{ik} \tilde{Z}_{i\ell} X_i \hat{\xi}_i \right) \left(\sum_{j=1}^n Z_{jk} Z_{j\ell} X_j \hat{\xi}_j \right)' \\
= & \sum_{i,j=1}^n \left(\sum_{k=1}^K \sum_{\ell=1}^K \tilde{Z}_{ik} Z_{jk} \tilde{Z}_{i\ell} Z_{j\ell} \right) X_i \hat{\xi}_i \hat{\xi}_j X_j' = \sum_{i,j=1}^n (\sum_{k=1}^K \tilde{Z}_{ik} Z_{jk})^2 X_i \hat{\xi}_i \hat{\xi}_j X_j' \\
= & \sum_{i,j=1}^n (\tilde{Z}_i' Z_j)^2 X_i \hat{\xi}_i \hat{\xi}_j X_j' = \sum_{i,j=1}^n P_{ij}^2 X_i \hat{\xi}_i \hat{\xi}_j X_j'
\end{aligned}
$$

Adding this equation to the previous one then gives

$$
\begin{aligned}
\hat{\Sigma} &= \sum_{i \neq j \neq k} P_{ik} P_{kj} X_i X_j' \hat{\xi}_k^2 + \sum_{i \neq j} P_{ij}^2 X_i X_i' \hat{\xi}_j^2 - \sum_{i=1}^n P_{ii}^2 X_i X_i' \hat{\xi}_i^2 + \sum_{i,j=1}^n P_{ij}^2 X_i \hat{\xi}_i \hat{\xi}_j X_j' \\
&= \sum_{i \neq j \neq k} P_{ik} P_{kj} X_i X_j' \hat{\xi}_k^2 + \sum_{i \neq j} P_{ij}^2 (X_i X_i' \hat{\xi}_j^2 + X_i \hat{\xi}_i \hat{\xi}_j X_j'),
\end{aligned}
$$

giving the equality in Section 2.

Let $\dot{\sigma}_i^2 = E\left[\xi_i^2 | \mathcal{Z}\right]$ and $\bar{z}_i = \sum_j P_{ij} z_j = e_i' P z$. Then

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

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

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

It follows that

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

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

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

since $\epsilon_n \to 0$. 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, since $\bar{H}_n^{-1}$, $\bar{\Omega}_n + \bar{\Psi}_n$, H_n^{-1}, and $\Omega_n + \Psi_n$ are all bounded $a.s.\ P_{\mathcal{Z}}$ for n sufficiently large, we have, making use of Lemma A6,

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

giving the first conclusion.

Next, consider the case where $K/r_n \longrightarrow \infty$. Then for JIV1, where $\xi_i = \varepsilon_i/(1-P_{ii})$ and $\dot{\sigma}_i^2 = \sigma_i^2/(1-P_{ii})^2$, the almost sure boundedness of $\bar{\Omega}_n$ for n sufficiently large implies that 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 the fact that $\bar{H}_n^{-1}$, $(r/K_n)\bar{\Psi}_n$, H_n^{-1}, and $(r/K_n)\Psi_n$ are all bouneded with probability one for n sufficiently large and by Lemma A6,

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

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), $B_n\hat{A}S_n^{-1\prime} = B_nAS_n^{-1\prime} + o_p(1)$. Then multiplying by B_n and using the conclusion of Theorem 4 we have

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

for $F_n = B_nAS_n^{-1}\bar{V}_n^{1/2}$ and $\bar{Y}_n = \bar{V}_n^{-1/2}S_n'(\delta - \delta_0)$, and note that the third equality above follows from the Slutzky Theorem given the continuity of the square root matrix. By Theorem 2, $\bar{Y}_n \xrightarrow{d} N(0, I_G)$. Then since $L_n = (F_nF_n')^{-1/2}F_n$ satisfies $L_nL_n' = I$, it follows from the Slutzky 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 $B_nAS_n^{-1\prime} = b_nc'S_n^{-1\prime}$ is bounded by hypothesis so condition ii) of Theorem 5 is satisfied. Also, $B_n(\bar{A} - A)S_n^{-1\prime} = 0$ so condition iii) of Theorem 5 is satisfied. If K/r_n is bounded then by hypothesis, $\lambda_{\min}(B_nAS_n^{-1\prime}\bar{V}_nS_n^{-1}A'B_n') = b_n^2c'S_n^{-1\prime}\bar{V}_nS_n^{-1}c \geq C$ or if $K/r_n \longrightarrow \infty$ then $\lambda_{\min}(B_nAS_n^{-1\prime}\bar{V}_n^*S_n^{-1}A'B_n') = b_n^2c'S_n^{-1\prime}\bar{V}_n^*S_n^{-1}c \geq C$, 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 I_2 of the set $\left\{(i,j)_{i,j=1}^n\right\}$ and any subset I_3 of $\left\{(i,j,k)_{i,j,k=1}^n\right\}$, (a) $\sum_{\mathcal{I}_2} P_{ij}^4 \leq K$; (b) $\sum_{\mathcal{I}_3} P_{ij}^2P_{jk}^2 \leq K$; and (c) $\sum_{\mathcal{I}_3}\left|P_{ij}^2P_{ik}P_{jk}\right| \leq K$, with probability one for n sufficiently large.

Proof: By Assumption 2, $Z'Z$ is nonsingular with probability one for n sufficiently large. Also, by P idempotent, $rank(P) = tr(P) = K$, $0 \leq P_{ii} \leq 1$, and $\sum_{j=1}^{n} P_{ij}^2 = P_{ii}$. Therefore, with probability one for n large enough,

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

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 with probability one for n sufficiently large, a) $tr\left[(P-D)^4\right] \leq CK$; b) $\left| \sum_{i<j<k<l} P_{ik} P_{jk} P_{il} P_{jl} \right| \leq CK$, and c) $|S_n| \leq CK$, where $D = diag(P_{11}, ..., P_{nn})$.

Proof: To show part (a), note that

$$\begin{aligned}
(P-D)^4 &= (P - PD - DP + D^2)^2 = P - PD - PDP + PD^2 - PDP + PDPD + PD^2P - PD^3 \\
&\quad -DP + DPD + DPDP - DPD^2 + D^2P - D^2PD - D^3P + D^4.
\end{aligned}$$

Note that $tr(A') = tr(A)$ and $tr(AB) = tr(BA)$ for any square matrices A and B. Then,

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

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

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

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 $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ and $\mathrm{tr}(A') = \mathrm{tr}(A)$,

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

Next, compute each of the terms. Note that

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

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

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

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

Summing up gives the result

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

Then by the triangle inequality and Lemma B1 we have

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

with probability one for n sufficiently large, thus, giving part c). That is, $S_n = O_{a.s.}(K)$.

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

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

Note that by Lemma A1,

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

Also, note that

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

and

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

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

$$
E\left[\Delta_1^2|\mathcal{Z}\right] = E\left[\Delta_2^2|\mathcal{Z}\right] + E\left[\Delta_3^2|\mathcal{Z}\right] + 2E\left[\Delta_2\Delta_3|\mathcal{Z}\right] = O_{a.s.}(K) + 2S_n = O_{a.s.}(K).
$$

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

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

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

Proof: From the proof of Lemma B2 and by Lemma B1 and Lemma B2 b) we have for n sufficiently large

$$
\begin{aligned}
\|LL'\|^2 &= \mathrm{tr}(LL'LL') = \sum_{i<j} P_{ij}^4 + 2\sum_{i<j<k}\left(P_{ij}^2P_{ik}^2 + P_{ik}^2P_{jk}^2\right) + 4\sum_{i<j<k<\ell} P_{ik}P_{kj}P_{j\ell}P_{\ell i} \\
&\leq C(K + \left|\sum_{i<j<k<\ell} P_{ik}P_{kj}P_{j\ell}P_{\ell i}\right|) \leq CK \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.
\end{aligned}
$$

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

Lemma B4: Let A and B be $n \times n$ matrices, and let $\underline{K}_{nn}$ $\left(n^2 \times n^2\right)$ denote a commutation matrix so that $\underline{K}_{nn} vec\,(A) = vec\,(A')$ Then,

$$
tr\left\{(A \otimes B)\,\underline{K}_{nn}\right\} = tr\left\{AB\right\}.
$$

Proof: To proceed, first note that, $\underline{K}_{nn}$ has the explicit form

$$\underline{K}_{nn} = \left(I_n \otimes e_{1,n} \vdots I_n \otimes e_{2,n} \vdots \cdots \vdots I_n \otimes e_{n,n} \right)'$$

where $e_{i,n}$ is the i^{th} column of an $n \times n$ identity matrix. Now, by direct calculation, we have that

$$\begin{aligned}
& tr\left\{ (A \otimes B) \underline{K}_{nn} \right\} \\
= \; & tr\left\{ \underline{K}_{nn} (A \otimes B) \right\} \\
= \; & tr\left\{ \begin{pmatrix} A \otimes e'_{1,n}B \\ A \otimes e'_{2,n}B \\ \vdots \\ A \otimes e'_{n,n}B \end{pmatrix} \right\} \\
= \; & \sum_{i=1}^{n} \left(e'_{i,n} A e_{1,n} \right) \left(e'_{1,n} B e_{i,n} \right) + \sum_{i=1}^{n} \left(e'_{i,n} A e_{2,n} \right) \left(e'_{2,n} B e_{i,n} \right) + \cdots + \sum_{i=1}^{n} \left(e'_{i,n} A e_{n,n} \right) \left(e'_{n,n} B e_{i,n} \right) \\
= \; & e'_{1,n} B \left(\sum_{i=1}^{n} e_{i,n} e'_{i,n} \right) A e_{n,1} + e'_{2,n} B \left(\sum_{i=1}^{n} e_{i,n} e'_{i,n} \right) A e_{2,n} + \cdots + e'_{n,n} B \left(\sum_{i=1}^{n} e_{i,n} e'_{i,n} \right) A e_{n,n} \\
= \; & e'_{1,n} B A e_{1,n} + e'_{2,n} B A e_{2,n} + \cdots \cdot + e'_{n,n} B A e_{n,n} \\
= \; & tr\left\{ BA \right\} = tr\left\{ AB \right\}. \text{ Q.E.D.}
\end{aligned}$$

For Lemma A5 below, let $\phi_i(\mathcal{Z})$ $(i = 1, ..., n)$ denote some sequence of measurable functions. In application of this lemma, we will take $\phi_i(\mathcal{Z})$ to be either conditional variances or conditional covariances given Z.

Lemma B5: Suppose that $i)$ P is a symmetric, idempotent matrix with rank$(P) = K$ and $P_{ii} \leq C < 1$; $ii)$ $(u_1, \varepsilon_1), ..., (u_n, \varepsilon_n)$ are independent conditional on Z; $iii)$ there exists a constant C such that, with probability one, $\sup_i E\left(u_i^4 | \mathcal{Z}\right) \leq C$, $\sup_i E\left(\varepsilon_i^4 | \mathcal{Z}\right) \leq C$, and $\sup_i |\phi_i| = \sup_i |\phi_i(\mathcal{Z})| \leq C$. Then, with probability one,

a) $E\left[\left(\frac{1}{K} \sum_{1 \leq\, i \,<\, k \,\leq\, n} P_{ki}^2 \phi_k (u_i \varepsilon_i - \gamma_i) \right)^2 | \mathcal{Z} \right] \to 0$;

b) $E\left[\left(\frac{1}{K} \sum_{1 \leq\, i \,<\, j \,<\, k \,\leq\, n} P_{ki} P_{kj} \phi_k (u_i \varepsilon_j + u_j \varepsilon_i) \right)^2 | \; \mathcal{Z} \right] \to 0$;

c) $E\left[\left(\frac{1}{K} \sum_{1 \leq\, i \,<\, j \,<\, k \,\leq\, n} P_{ki} P_{kj} \phi_k \varepsilon_i \varepsilon_j \right)^2 | \mathcal{Z} \right] \to 0$;

d) $E\left[\left(\frac{1}{K} \sum_{1 \leq\, i \,<\, j \,<\, k \,\leq\, n} P_{ki} P_{kj} \phi_k u_i u_j \right)^2 | \mathcal{Z} \right] \to 0$.

Proof: To set some notation, let $\sigma_i^2 = E[\varepsilon_i^2 | \mathcal{Z}]$, $\omega_i^2 = \omega_{in}^2(\mathcal{Z}) = E[u_i^2 | \mathcal{Z}]$, and $\gamma_i = \gamma_{in}(\mathcal{Z}) = E[u_i \varepsilon_i | \mathcal{Z}]$, where in order to simplify notation, we suppress the dependence of σ_i^2 on $\mathcal{Z}$ and of ω_i^2 and γ_i on $\mathcal{Z}$ and n.

Now, to show part (a), note that

$$
\begin{aligned}
& E\left[\left(\frac{1}{K}\sum\nolimits_{1\leq\ i\ <\ k\ \leq\ n}P_{ki}^{2}\phi_{k}\left(u_{i}\varepsilon_{i}-\gamma_{i}\right)\right)^{2}\ \mid\mathcal{Z}\right] \\
= & \frac{1}{K^{2}}\sum\nolimits_{1\leq\ i\ <\ k\ \leq\ n}P_{ki}^{4}\phi_{k}^{2}\left\{E\left(u_{i}^{2}\varepsilon_{i}^{2}|\mathcal{Z}\right)-\gamma_{i}^{2}\right\} \\
& +\frac{2}{K^{2}}\sum\nolimits_{1\leq\ i\ <\ k\ <\ l\ \leq\ n}P_{ki}^{2}P_{li}^{2}\phi_{k}\phi_{l}\left\{E\left(u_{i}^{2}\varepsilon_{i}^{2}|\mathcal{Z}\right)-\gamma_{i}^{2}\right\} \\
\leq & \frac{1}{K^{2}}\sum\nolimits_{1\leq\ i\ <\ k\ \leq\ n}P_{ki}^{4}\phi_{k}^{2}\left\{\sqrt{E\left(u_{i}^{4}|\mathcal{Z}\right)E\left(\varepsilon_{i}^{4}|\mathcal{Z}\right)}+E\left(u_{i}^{2}|\mathcal{Z}\right)E\left(\varepsilon_{i}^{2}|\mathcal{Z}\right)\right\} \\
& +\frac{2}{K^{2}}\sum\nolimits_{1\leq\ i\ <\ k\ <\ l\ \leq\ n}P_{ki}^{2}P_{li}^{2}\left|\phi_{k}\right|\left|\phi_{l}\right|\left\{\sqrt{E\left(u_{i}^{4}|\mathcal{Z}\right)E\left(\varepsilon_{i}^{4}|\mathcal{Z}\right)}+E\left(u_{i}^{2}|\mathcal{Z}\right)E\left(\varepsilon_{i}^{2}|\mathcal{Z}\right)\right\} \\
\leq & C\left\{\frac{1}{K^{2}}\sum\nolimits_{1\leq\ i\ <\ k\ \leq\ n}P_{ki}^{4}+\frac{2}{K^{2}}\sum\nolimits_{1\leq\ i\ <\ k\ <\ l\ \leq\ n}P_{ki}^{2}P_{li}^{2}\right\}\rightarrow 0\ \ a.s.\ \ \mathbb{P}_{\mathcal{Z}},
\end{aligned}
$$

where the first inequality is the result of applying T and a conditional version of CS, the second inequality follows by hypothesis, and the convergence to zero almost surely follows from applying Lemma B1 parts (a) and (b).

To show part (b), first let L be a lower triangular matrix with $(i,j)^{th}$ element $L_{ij}=P_{ij}1\left(i>j\right)$ as in Lemma B3 above, and define $D_{\gamma}=diag\left(\gamma_{1},...,\gamma_{n}\right)$, $D_{\phi}=diag\left(\phi_{1},...,\phi_{n}\right)$, $u=\left(u_{1},...,u_{n}\right)'$, and $\varepsilon=\left(\varepsilon_{1},...,\varepsilon_{n}\right)'$. It then follows by direct multiplication that

$$
\varepsilon'L'D_{\phi}Lu-tr\left\{L'D_{\phi}LD_{\gamma}\right\}=\sum_{1\leq\ i\ <\ k\ \leq\ n}P_{ki}^{2}\phi_{k}\left(u_{i}\varepsilon_{i}-\gamma_{i}\right)+\sum_{1\leq\ i\ <\ j\ <\ k\ \leq\ n}P_{ki}P_{kj}\phi_{k}\left(u_{i}\varepsilon_{j}+u_{j}\varepsilon_{i}\right),
$$

so that by making use of Loève's c_r inequality, we have that

$$
\begin{aligned}
& \frac{1}{K^{2}}E\left[\left(\sum\nolimits_{1\leq\ i\ <\ j\ <\ k\ \leq\ n}P_{ki}P_{kj}\phi_{k}\left(u_{i}\varepsilon_{j}+u_{j}\varepsilon_{i}\right)\right)^{2}\mid\mathcal{Z}\right] \\
\leq\ & 2\frac{1}{K^{2}}E\left[\left(u'L'D_{\phi}L\varepsilon-tr\left\{L'D_{\phi}LD_{\gamma}\right\}\right)^{2}|\mathcal{Z}\right] \\
& +2\frac{1}{K^{2}}E\left[\left(\sum\nolimits_{1\leq\ i\ <\ k\ \leq\ n}P_{ki}^{2}\phi_{k}\left(u_{i}\varepsilon_{i}-\gamma_{i}\right)\right)^{2}|\mathcal{Z}\right] \qquad (2)
\end{aligned}
$$

It has already been shown in the proof of part (a) that $\left(1/K^{2}\right)E\left[\left(\sum\nolimits_{1\leq\ i\ <\ k\ \leq\ n}P_{ki}^{2}\phi_{k}\left(u_{i}\varepsilon_{i}-\gamma_{i}\right)\right)^{2}|\mathcal{Z}\right]\rightarrow 0\ \ a.s.\ P_{\mathcal{Z}}$, so what remains to be shown is that $\left(1/K^{2}\right)E\left[\left(u'L'D_{\phi}L\varepsilon-tr\left\{L'D_{\phi}LD_{\gamma}\right\}\right)^{2}|\mathcal{Z}\right]\rightarrow 0\ \ a.s.$

$P_{\mathcal{Z}}$. To show the latter, first note that

$$\begin{aligned}
& \frac{1}{K^2} E\left[\left(u'L'D_\phi L\varepsilon - tr\left\{L'D_\phi LD_\gamma\right\}\right)^2 \mid \mathcal{Z}\right] \\
= & \frac{1}{K^2} E\left[\left(u'L'D_\phi L\varepsilon\right)^2 \mid \mathcal{Z}\right] - \frac{1}{K^2}\left[tr\left\{L'D_\phi LD_\gamma\right\}\right]^2 \\
= & \frac{1}{K^2} E\left[u'L'D_\phi L\varepsilon \otimes u'L'D_\phi L\varepsilon \mid \mathcal{Z}\right] - \frac{1}{K^2}\left[tr\left\{L'D_\phi LD_\gamma\right\}\right]^2 \\
= & \frac{1}{K^2} E\left[tr\left\{\left(u' \otimes u'\right)\left(L'D_\phi L \otimes L'D_\phi L\right)\left(\varepsilon \otimes \varepsilon\right)\right\} \mid \mathcal{Z}\right] - \frac{1}{K^2}\left[tr\left\{L'D_\phi LD_\gamma\right\}\right]^2 \\
= & \frac{1}{K^2} tr\left\{\left(L'D_\phi L \otimes L'D_\phi L\right) E\left[\varepsilon u' \otimes \varepsilon u' | \mathcal{Z}\right]\right\} - \frac{1}{K^2}\left[tr\left\{L'D_\phi LD_\gamma\right\}\right]^2. \qquad (3)
\end{aligned}$$

Next, note that, by straightforward calculation, we have

$$\begin{aligned}
& E\left[\varepsilon u' \otimes \varepsilon u' | \mathcal{Z}\right] \\
= & \begin{pmatrix} \sigma_1^2\omega_1^2 e_1e_1' & \sigma_1^2\omega_2^2 e_1e_2' & \cdots & \sigma_1^2\omega_n^2 e_1e_n' \\ \sigma_2^2\omega_1^2 e_2e_1' & \sigma_2^2\omega_2^2 e_2e_2' & \cdots & \sigma_2^2\omega_n^2 e_2e_n' \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n^2\omega_1^2 e_ne_1' & \sigma_n^2\omega_2^2 e_ne_2' & \cdots & \sigma_n^2\omega_n^2 e_ne_n' \end{pmatrix} + \begin{pmatrix} \gamma_1^2 e_1e_1' & \gamma_1\gamma_2 e_2e_1' & \cdots & \gamma_1\gamma_n e_ne_1' \\ \gamma_2\gamma_1 e_1e_2' & \gamma_2^2 e_2e_2' & \cdots & \gamma_2\gamma_n e_ne_2' \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_n\gamma_1 e_1e_n' & \gamma_n\gamma_2 e_2e_n' & \cdots & \gamma_n^2 e_ne_n' \end{pmatrix} \\
& + \begin{pmatrix} \vartheta_1 e_1e_1' & \underset{n\times n}{0} & \cdots & \underset{n\times n}{0} \\ \underset{n\times n}{0} & \vartheta_2 e_2e_2' & \cdots & \underset{n\times n}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{n\times n}{0} & \underset{n\times n}{0} & \cdots & \vartheta_n e_ne_n' \end{pmatrix} + \begin{pmatrix} \gamma_1 \otimes D_\gamma & \underset{n\times n}{0} & \cdots & \underset{n\times n}{0} \\ \underset{n\times n}{0} & \gamma_2 \otimes D_\gamma & \cdots & \underset{n\times n}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{n\times n}{0} & \underset{n\times n}{0} & \cdots & \gamma_n \otimes D_\gamma \end{pmatrix} \\
= & \left(D_\sigma \otimes I_n\right) vec\left(I_n\right) vec\left(I_n\right)\left(D_\omega \otimes I_n\right) + \left(D_\gamma \otimes I_n\right) \underline{K}_{nn} \left(D_\gamma \otimes I_n\right) + \underline{E}' D_\vartheta \underline{E} + \left(D_\gamma \otimes D_\gamma\right), \qquad (4)
\end{aligned}$$

where $\underline{K}_{nn}$ is an $n^2 \times n^2$ commutation matrix such that for any $n \times n$ matrix A, $\underline{K}_{nn} vec\left(A\right) = vec\left(A'\right)$. (See Magnus and Neudecker, 1988, pages 46-48 for more on commutation matrices.) Also, here, $D_\gamma = diag\left(\gamma_1, ..., \gamma_n\right)$, $D_\sigma = diag\left(\sigma_1^2, ..., \sigma_n^2\right)$, $D_\omega = diag\left(\omega_1^2, ..., \omega_n^2\right)$, $D_\vartheta = diag\left(\vartheta_1, ..., \vartheta_n\right)$ with $\vartheta_i = E\left[\varepsilon_i^2 u_i^2 | \mathcal{Z}\right] - \sigma_i^2\omega_i^2 - 2\gamma_i^2$ for $i = 1, ..., n$, and $\underline{E} = \left(e_1 \otimes e_1 \vdots e_2 \otimes e_2 \vdots \cdots \vdots e_n \otimes e_n\right)'$; and e_i is the i^{th}

column of an $n \times n$ identity matrix . It follows from (3) and (4) that

$$
\begin{aligned}
& \frac{1}{K^2} E\left[\left(u' L' D_\phi L \varepsilon - tr\left\{L' D_\phi L D_\gamma\right\}\right)^2 \mid \mathcal{Z}\right] \\
= & \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right) E\left[\varepsilon u' \otimes \varepsilon u' | \mathcal{Z}\right]\right\} - \frac{1}{K^2}\left[tr\left\{L' D_\phi L D_\gamma\right\}\right]^2 \\
= & \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right)\left(D_\sigma \otimes I_n\right) vec\left(I_n\right) vec\left(I_n\right)'\left(D_\omega \otimes I_n\right)\right\} \\
& + \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right)\left(D_\gamma \otimes I_n\right) \underline{K}_{nn}\left(D_\gamma \otimes I_n\right)\right\} \\
& + \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right) \underline{E}' D_\vartheta \underline{E}\right\} + \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right)\left(D_\gamma \otimes D_\gamma\right)\right\} \\
& - \frac{1}{K^2}\left[tr\left\{L' D_\phi L D_\gamma\right\}\right]^2 \\
= & \frac{1}{K^2} vec\left(I_n\right)'\left(D_\omega L' D_\phi L D_\sigma \otimes L' D_\phi L\right) vec\left(I_n\right) + \frac{1}{K^2} tr\left\{\left(D_\gamma L' D_\phi L D_\gamma \otimes L' D_\phi L\right) \underline{K}_{nn}\right\} \\
& + \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right) \underline{E}' D_\vartheta \underline{E}\right\} + \frac{1}{K^2} tr\left\{\left(L' D_\phi L D_\gamma \otimes L' D_\phi L D_\gamma\right)\right\} \\
& - \frac{1}{K^2}\left[tr\left\{L' D_\phi L D_\gamma\right\}\right]^2 \\
= & \frac{1}{K^2} tr\left\{L' D_\phi L D_\omega L' D_\phi L D_\sigma\right\} + \frac{1}{K^2} tr\left\{\left(D_\gamma L' D_\phi L D_\gamma \otimes L' D_\phi L\right) \underline{K}_{nn}\right\} \\
& + \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right) \underline{E}' D_\vartheta \underline{E}\right\} + \frac{1}{K^2}\left[tr\left\{L' D_\phi L D_\gamma\right\}\right]^2 - \frac{1}{K^2}\left[tr\left\{L' D_\phi L D_\gamma\right\}\right]^2 \\
= & \frac{1}{K^2} tr\left\{L' D_\phi L D_\omega L' D_\phi L D_\sigma\right\} + \frac{1}{K^2} tr\left\{\left(D_\gamma L' D_\phi L D_\gamma \otimes L' D_\phi L\right) \underline{K}_{nn}\right\} \\
& + \frac{1}{K^2} tr\left\{\left(L' D_\phi L \otimes L' D_\phi L\right) \underline{E}' D_\vartheta \underline{E}\right\} \qquad (5)
\end{aligned}
$$

Focusing first on the first term of (5), we get

$$
\begin{aligned}
& \frac{1}{K^2} tr\left\{L' D_\phi L D_\omega L' D_\phi L D_\sigma\right\} \\
\leq\ & \sqrt{\frac{1}{K^2} tr\left\{L' D_\phi L D_\omega^2 L' D_\phi L\right\}} \sqrt{\frac{1}{K^2} tr\left\{L' D_\phi L D_\sigma^2 L' D_\phi L\right\}} \\
\leq\ & \sqrt{\max_{1\leq i\leq n} \omega_i^4} \sqrt{\max_{1\leq i\leq n} \sigma_i^4} \frac{1}{K^2} tr\left\{L' D_\phi L L' D_\phi L\right\} \\
\leq\ & \sqrt{\max_{1\leq i\leq n} \omega_i^4} \sqrt{\max_{1\leq i\leq n} \sigma_i^4} \sqrt{\frac{1}{K^2} tr\left\{D_\phi L L' D_\phi^2 L L' D_\phi\right\}} \sqrt{\frac{1}{K^2} tr\left\{L L' L L'\right\}} \\
\leq\ & \sqrt{\max_{1\leq i\leq n} \omega_i^4} \sqrt{\max_{1\leq i\leq n} \sigma_i^4} \sqrt{\max_{1\leq\ i\ \leq\ n} \phi_i^2} \sqrt{\frac{1}{K^2} tr\left\{L L' D_\phi^2 L L'\right\}} \sqrt{\frac{1}{K^2} tr\left\{L L' L L'\right\}} \\
\leq\ & \sqrt{\max_{1\leq i\leq n} \omega_i^4} \sqrt{\max_{1\leq i\leq n} \sigma_i^4} \left(\max_{1\leq\ i\ \leq\ n} \phi_i^2\right) \frac{1}{K^2} tr\left\{L' L L' L\right\} \\
\leq\ & \sqrt{\max_{1\leq i\leq n} E\left(u_i^4 | \mathcal{Z}\right)} \sqrt{\max_{1\leq i\leq n} E\left(\varepsilon_i^4 | \mathcal{Z}\right)} \left(\max_{1\leq\ i\ \leq\ n} \phi_i^2\right) \frac{1}{K^2} tr\left\{L' L L' L\right\} \\
\leq\ & C \frac{1}{K^2} tr\left\{L' L L' L\right\} = \frac{C}{K^2} \left\|L L'\right\|^2 \quad a.s.\ \mathbb{P}_{\mathcal{Z}},
\end{aligned} \tag{6}
$$

where the first and third inequalities follow from CS; the second and fourth inequalities follows from the fact that let A be an $n \times n$ matrix and $\Lambda = diag\left(\lambda_1, ..., \lambda_n\right)$ such that $\lambda_i \geq 0$ for all i, then,

$$
tr\left\{A' \Lambda A\right\} \leq \left(\max_{1\leq\ i\ \leq\ n} \lambda_i\right) tr\left(A' A\right); \tag{7}
$$

the sixth inequality follows from the conditional version of the Jensen's inequality and the last inequality follows in light of the assumptions of this lemma.

Turning our attention now to the second term of (5), we see that by Lemma B4

$$
\begin{aligned}
& \frac{1}{K^2} tr\left\{\left(D_\gamma L' D_\phi L D_\gamma \otimes L' D_\phi L\right) \underline{K}_{nn}\right\} \\
=\ & \frac{1}{K^2} tr\left\{D_\gamma L' D_\phi L D_\gamma L' D_\phi L\right\} = \frac{1}{K^2} tr\left\{L' D_\phi L D_\gamma L' D_\phi L D_\gamma\right\}.
\end{aligned}
$$

It follows by calculations similar to that used to obtain (6), we obtain

$$
\begin{aligned}
& \frac{1}{K^2} tr\left\{\left(D_\gamma L' D_\phi L D_\gamma \otimes L' D_\phi L\right) \underline{K}_{nn}\right\} \\
\leq\ & \left(\max_{1\leq i\leq n} \gamma_i^2\right) \left(\max_{1\leq\ i\ \leq\ n} \phi_i^2\right) \frac{1}{K^2} tr\left\{L' L L' L\right\} \leq C \frac{1}{K^2} tr\left\{L' L L' L\right\} = \frac{C}{K^2} \left\|L L'\right\|^2 \quad a.s.\ \mathbb{P}_{\mathcal{Z}}.
\end{aligned} \tag{8}
$$

Finally, to analyze the third term of (5), we note that

$$
\begin{aligned}
& \frac{1}{K^2}\left|tr\left\{\left(L'D_\phi L\otimes L'D_\phi L\right)\underline{E}'D_\vartheta\underline{E}\right\}\right| \\
\leq\ & \frac{1}{K^2}\sum_{i=1}^{n}|\vartheta_i|\,tr\left\{e_i'L'D_\phi Le_iL'D_\phi Le_ie_i'\right\}=\frac{1}{K^2}\sum_{i=1}^{n}|\vartheta_i|\left(e_i'L'D_\phi Le_i\right)^2 \\
\leq\ & \frac{1}{K^2}\sum_{i=1}^{n}|\vartheta_i|\left(e_i'L'D_\phi^2Le_i\right)\left(e_i'L'Le_i\right)\leq\left(\max_{1\leq\ i\ \leq\ n}\phi_i^2\right)\frac{1}{K^2}\sum_{i=1}^{n}|\vartheta_i|\left(e_i'L'Le_i\right)^2 \\
\leq\ & \left\{\sqrt{E\left[\varepsilon_i^4|\mathcal{Z}\right]}\sqrt{E\left[u_i^4|\mathcal{Z}\right]}+\left(\max_{1\leq i\leq n}\sigma_i^2\right)\left(\max_{1\leq i\leq n}\omega_i^2\right)+2\left(\max_{1\leq i\leq n}\gamma_i^2\right)\right\} \\
& \times\left(\max_{1\leq\ i\ \leq\ n}\phi_i^2\right)\frac{1}{K^2}\sum_{i=1}^{n}\left(e_i'L'Le_i\right)^2 \\
\leq\ & C\frac{1}{K^2}\sum_{i=1}^{n}\left(e_i'L'Le_i\right)^2\leq C\frac{1}{K^2}\sum_{i=1}^{n}\left(e_i'P'Pe_i\right)^2=C\frac{1}{K^2}\sum_{i=1}^{n}\left(e_i'Pe_i\right)^2=C\frac{1}{K^2}\sum_{i=1}^{n}P_{ii}^2 \\
\leq\ & C\frac{1}{K^2}\sum_{i=1}^{n}P_{ii}=C\left(1/K\right)\quad a.s.\ \mathbb{P}_{\mathcal{Z}}, \qquad (9)
\end{aligned}
$$

where the first inequality above follows from T, the second inequality follows from CS, the third inequality makes use of (7), the fourth inequality uses CS and T, the fifth inequality follows as a consequence of the assumptions of this lemma, and the last inequality holds since $P_{ii}<1$.

In light of (5), it follows from (6), (8), (9), and Lemma B3 that

$$\frac{1}{K^2}E\left[\left(u'L'D_\phi L\varepsilon-tr\left\{L'D_\phi LD_\gamma\right\}\right)^2\ \mid\mathcal{Z}\right]\leq 2C\left(1/K^2\right)\left\|LL'\right\|^2+C\left(1/K\right)\leq C/K\ \ a.s.\ \mathbb{P}_{\mathcal{Z}};$$

which shows part (b).

It is easily seen that parts (c) and (d) can be proved in essentially the same way as part (b) (by taking $u_i=\varepsilon_i$); hence, to avoid redundancy, we do not give detailed arguments for these parts. 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 $w_{in}=b_{1n}'W_{in}$ and $u_i=b_{2n}'U_i$, where we suppress the n subscript on u_i for notational convenience. Then

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

Also, $E\left[\left\|w_{1n}\right\|^4|\mathcal{Z}\right]\leq\sum_i E\left[\left\|w_{in}\right\|^4|\mathcal{Z}\right]\leq C\sum_i E\left[\left\|W_{in}\right\|^4|\mathcal{Z}\right]\longrightarrow 0\ \ a.s.\ P_{\mathcal{Z}}$; so that, by a conditional version of M, we deduce that for any $\upsilon>0$

$$P\left(|w_{1n}|\geq\upsilon\mid\mathcal{Z}\right)\to 0\ \ a.s.\ \mathbb{P}_{\mathcal{Z}}.$$

Moreover, note that $\sup_n E\left[|P(|w_{1n}| \geq \upsilon \mid \mathcal{Z})|^2\right] < \infty$. It follows that, by Theorem 25.12 of Billingsley (1986)

$$P(|w_{1n}| \geq \upsilon\ ) = E\left[P(|w_{1n}| \geq \upsilon \mid \mathcal{Z})\right] \to 0 \text{ as } n \to \infty;$$

that is, $w_{1n} \xrightarrow{p} 0$ unconditionally. Hence,

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

Now, we will show that $Y_n \xrightarrow{d} N(0,1)$ by first showing that, conditional on $\mathcal{Z}$,

$$\sum\nolimits_{i=2}^{n} y_{in} \xrightarrow{d} N(0,1),$$

with probability one. To proceed, let $X_i = (W_{in}', U_i', \varepsilon_i)'$ for $i = 1, ..., n$. Define the σ-fields $F_{i,n} = \sigma(\mathcal{X}_1, ..., \mathcal{X}_i)$ for $i = 1, ..., n$. Note that, by construction, $F_{i-1,n} \subseteq F_{i,n}$. Moreover, it is straightforward to verify that, conditional on Z, $\{y_{in}, \mathcal{F}_{i,n}, 1 \leq i \leq n, n \geq 2\}$ is a martingale difference array, and we can apply the martingale central limit theorem. Moreover, as before, let $\sigma_i^2 = E[\varepsilon_i^2|\mathcal{Z}]$, $\omega_i^2 = \omega_{in}^2(\mathcal{Z}) = E[u_i^2|\mathcal{Z}]$, and $\gamma_i = \gamma_{in}(\mathcal{Z}) = E[u_i\varepsilon_i|\mathcal{Z}]$, where in order to simplify notation, we suppress the dependence of σ_i^2 on Z and of ω_i^2 and γ_i on Z and n. Now, note that $E[w_{in}\bar{y}_{jn}|Z] = 0$, for all i and j and that

$$\begin{aligned}
E\left[(\bar{y}_{in})^2|\mathcal{Z}\right] &= \sum_{j<i}\sum_{k<i} E\left[(u_j P_{ij}\varepsilon_i + u_i P_{ij}\varepsilon_j)(u_k P_{ik}\varepsilon_i + u_i P_{ik}\varepsilon_k)|\mathcal{Z}\right]/K \\
&= \sum_{j<i} P_{ij}^2\left[\omega_j^2\sigma_i^2 + \omega_i^2\sigma_j^2 + 2\gamma_i\gamma_j\right]/K.
\end{aligned}$$

Thus,

$$\begin{aligned}
s_n^2(\mathcal{Z}) &= E\left[\left(\sum\nolimits_{i=2}^{n} y_{in}\right)^2|\mathcal{Z}\right] = \sum_{i=2}^{n}(E\left[w_{in}^2|\mathcal{Z}\right] + E\left[\bar{y}_{in}^2|\mathcal{Z}\right]) \\
&= b_{1n}' D_n(\mathcal{Z})\, b_{1n} - E\left[w_{1n}^2|\mathcal{Z}\right] + \sum_{i\neq j} P_{ij}^2\left[\omega_j^2\sigma_i^2 + \omega_i^2\sigma_j^2 + 2\gamma_i\gamma_j\right]/K \\
&= b_{1n}' D_n(\mathcal{Z})\, b_{1n} + b_{2n}'\bar{\Sigma}_n(\mathcal{Z})\, b_{2n} + o_{a.s.}(1) = 1 + o_{a.s.}(1) \longrightarrow 1 \quad a.s.
\end{aligned}$$

Thus, $s_n^2(\mathcal{Z})$ is bounded and bounded away from zero with probability one. Also,

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

By condition iv), $\sum_{i=2}^{n} E\left[\|W_{in}\|^4 \,|\mathcal{Z}\right] \longrightarrow 0$. 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 $\sum_j P_{ij}^2 = P_{ii}$, we have that with probability one

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

Similarly, $\sum_{i=2}^{n} E\left[(\bar{y}_{in}^{u})^4 \,|\mathcal{Z}\right] \longrightarrow 0$ with probability one, so that

$$
\sum_{i=2}^{n} E\left[\bar{y}_{in}^4 | \mathcal{Z}\right] \leq C \sum_{i=2}^{n} \left\{ E\left[(\bar{y}_{in}^{\varepsilon})^4 \,|\mathcal{Z}\right] + E\left[(\bar{y}_{in}^{u})^4 \,|\mathcal{Z}\right] \right\} \longrightarrow 0 \quad a.s. \;\; \mathbb{P}_{\mathcal{Z}}
$$

Then by T we have

$$
\sum_{i=2}^{n} E\left[y_{in}^4 | \mathcal{Z}\right] \longrightarrow 0 \; a.s.
$$

Conditional on Z, to apply the martingale central limit theorem, it suffices to show that for any $\epsilon > 0$

$$
P\left(\left| \sum\nolimits_{i=2}^{n} E\left[y_{in}^2 | \mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}\right] - s_n^2\left(\mathcal{Z}\right) \right| \geq \epsilon \mid \mathcal{Z} \right) \to 0 \quad a.s. \; \mathbb{P}_{\mathcal{Z}}. \tag{10}
$$

Now, note that $E\left[w_{in} \bar{y}_{in} | \mathcal{Z}\right] = 0$ $a.s.$and, thus, we can write

$$
\begin{aligned}
& \sum_{i=2}^{n} E\left[y_{in}^2 | \mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}\right] - s_n^2\left(\mathcal{Z}\right) \\
= \quad & \sum_{i=2}^{n} \left(E[w_{in}^2 | \mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}] - E[w_{in}^2 | \mathcal{Z}] \right) + \sum_{i=2}^{n} E\left[w_{in} \bar{y}_{in} | \mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}\right] \\
& + \sum_{i=2}^{n} \left(E[\bar{y}_{in}^2 | \mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}] - E[\bar{y}_{in}^2 | \mathcal{Z}] \right)
\end{aligned} \tag{11}
$$

We will show that each term on the right-hand side of (11) converges to zero with probability one. To proceed, note first that by independence of $W_{1n}, ..., W_{nn}$ conditional on Z,

$$
E[w_{in}^2 | \mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}] = E[w_{in}^2 | \mathcal{Z}] \quad a.s. \; \mathbb{P}_{\mathcal{Z}}.
$$

Next, note that

$$
E\left[w_{in} \bar{y}_{in} | \mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}\right] = E[w_{in} u_i | \mathcal{Z}] \sum_{j<i} P_{ij} \varepsilon_j / \sqrt{K} + E[w_{in} \varepsilon_i | \mathcal{Z}] \sum_{j<i} P_{ij} u_j / \sqrt{K}.
$$

Let $\delta_i = E[w_{in}u_i|Z]$ 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 $\delta = (\delta_1, ..., \delta_n)$. Then, $\sum_{i=2}^{n} \sum_{j<i} \delta_i P_{ij}\varepsilon_j/\sqrt{K} = \delta'\bar{P}'\varepsilon/\sqrt{K}$. By CS $\delta'\delta = \sum_{i=1}^{n} \left(E\left[w_{in}u_i|\mathcal{Z}\right]\right)^2 \leq \sum_{i=1}^{n} E[w_{in}^2|Z]E[u_i^2|Z_i] \leq C$ with probability one. By Lemma B3, $\left\|\bar{P}'\bar{P}\right\| \leq \sqrt{K}$ with probability one, which in turn implies that $\lambda_{\max}\left(\bar{P}'\bar{P}\right) \leq \sqrt{K}$ with probability one. It then follows given $E\left[u_j^2|\mathcal{Z}_j\right] \leq C$ $a.s.$ that

$$E[(\delta'\bar{P}'\varepsilon/\sqrt{K})^2|\mathcal{Z}] \leq C\delta'\bar{P}'\bar{P}\delta/K \leq C\left\|\delta\right\|^2/\sqrt{K} \leq C/\sqrt{K} \longrightarrow 0 \;\; a.s.\; \mathbb{P}_{\mathcal{Z}},$$

so that by M we have that, for any $\epsilon > 0$,

$$P\left(\left|\delta\left(\mathcal{Z}\right)'\bar{P}'\varepsilon/\sqrt{K}\right| \geq \epsilon|\mathcal{Z}\right) \to 0 \;\; a.s.\; \mathbb{P}_{\mathcal{Z}}.$$

Similarly, we have $\sum_{i=2}^{n} E\left[w_{in}u_i|\mathcal{Z}\right]\sum_{j<i} P_{ij}\varepsilon_j/\sqrt{K} \longrightarrow 0 \;\; a.s.\; P_{\mathcal{Z}}$. Therefore, it follows by T that, for any $\epsilon > 0$,

$$P\left(\left|\sum_{i=2}^{n} E\left[w_{in}\bar{y}_{in}|\mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}\right]\right| \geq \epsilon|\mathcal{Z}\right) \to 0 \;\; a.s.\; \mathbb{P}_{\mathcal{Z}}.$$

To finish showing that eq. (10) is satisfied it only remains to show that, for any $\epsilon > 0$,

$$P\left(\left|\sum_{i=2}^{n}\left(E\left[\bar{y}_{in}^2|\mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}\right] - E[\bar{y}_{in}^2||\mathcal{Z})\right|\geq \epsilon|\mathcal{Z}\right) \to 0 \;\; a.s.\; \mathbb{P}_{\mathcal{Z}}. \tag{12}$$

Now, write

$$\begin{aligned}
&\sum_{i=2}^{n}\left(E\left[\bar{y}_{in}^2|\mathcal{X}_1, ..., \mathcal{X}_{i-1}, \mathcal{Z}\right] - E[\bar{y}_{in}^2|\mathcal{Z}]\right) \\
= \;& \sum_{j<i} \omega_i^2 P_{ij}^2\left(\varepsilon_j^2 - \sigma_j^2\right)/K + 2\sum_{j<k<i} \omega_i^2 P_{ij}P_{ik}\varepsilon_j\varepsilon_k/K \\
&+ \sum_{j<i} \sigma_i^2 P_{ij}^2\left(u_j^2 - \omega_j^2\right)/K + 2\sum_{j<k<i} \sigma_i^2 P_{ij}P_{ik}u_ju_k/K \\
&+ \sum_{j<i} \gamma_i P_{ij}^2\left(u_j\varepsilon_j - \gamma_j\right)/K + 2\sum_{j<k<i} \gamma_i P_{ij}P_{ik}(u_j\varepsilon_k + u_k\varepsilon_j)/K.
\end{aligned}$$

By applying part (a) of Lemma B5 with $\phi_i = \gamma_i$, ω_i^2, and then σ_i^2 in succession; we obtain, with probability one, that

$$\begin{aligned}
E\left[\left(\sum_{j<i} \gamma_i P_{ij}^2\left[u_j\varepsilon_j - \gamma_j\right]/K\right)^2 \mid \mathcal{Z}\right] &\longrightarrow 0, \\
E\left[\left(\sum_{j<i} \omega_i^2 P_{ij}^2\left[\varepsilon_j^2 - \sigma_j^2\right]/K\right)^2 \mid \mathcal{Z}\right] &\longrightarrow 0, \\
E\left[\left(\sum_{j<i} \sigma_i^2 P_{ij}^2\left[u_j^2 - \omega_j^2\right]/K\right)^2 \mid \mathcal{Z}\right] &\longrightarrow 0.
\end{aligned}$$

Moreover, applying part (b) of Lemma B5 with $\phi_i = \gamma_i$, we obtain

$$E\left[\left(\sum_{j<k<i} \gamma_i P_{ij} P_{ik} \left[u_j \varepsilon_k + u_k \varepsilon_i\right]/K\right)^2 \mid \mathcal{Z}\right] \longrightarrow 0 \ \ a.s.\ \mathbb{P}_{\mathcal{Z}}.$$

Similarly, conditional on Z, all of the remaining terms in eq. (13) converge in mean square to zero with probability one by parts (c) and (d) of Lemma B5.

The above argument shows that as $n \to \infty$

$$P\left(Y_n \leq y \mid \mathcal{Z}\right) \to \Phi\left(y\right) \quad a.s.\ \mathbb{P}_{\mathcal{Z}},$$

for every real number y, where $\Phi\left(y\right)$ denotes the cdf of a standard normal distribution. Moreover, it is clear that, for some $\epsilon > 0$,

$$\sup_n E\left[\left|P\left(Y_n \leq y \mid \mathcal{Z}\right)\right|^{1+\epsilon}\right] < \infty,$$

(take, for example, $\epsilon = 1$). Hence, by a version of the dominated convergence theorem, as given by Theorem 25.12 of Billingsley (1986), we deduce that

$$P\left(Y_n \leq y\right) = E\left[P\left(Y_n \leq y \mid \mathcal{Z}\right)\right] \to E\left[\Phi\left(y\right)\right] = \Phi\left(y\right),$$

which gives the desired conclusion. Q.E.D.

Proof of Lemma A4: Let $\bar{w}_i$, $\tilde{W}_i$, $\bar{y}_i$, $\tilde{Y}_i$,$\bar{\eta}_i$, $\tilde{\eta}_i$, $\bar{\mu}_W$, $\bar{\mu}_Y$, $\bar{\mu}_\eta$, $\bar{\sigma}^2_W$, $\bar{\sigma}^2_Y$, and $\bar{\sigma}^2_\eta$ are as defined at the beginning of Appendix A; and note that, in order to simplify notation, we have suppressed the dependence of these quantities on Z. Also, let

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

be predicted values from projecting $\bar{y} = (\bar{y}_1, ..., \bar{y}_n)'$ and $\bar{w} = (\bar{w}_1, ..., \bar{w}_n)'$ on the column space of Z. Now, define $e_{k,n}$ to be the k^{th} column of an $n \times n$ identity matrix, and we can write $\breve{w}_k = e'_{k,n} P\bar{w}$ and $\breve{y}_k = e'_{k,n} P\bar{y}$. Note that

$$\sum_i \breve{y}_i^2 = \bar{y}'P' \sum_i e_{k,n}\, e'_{k,n}\, P\bar{y} = \bar{y}'P'P\bar{y} = \bar{y}'P\bar{y} \leq \bar{y}'\bar{y} = \sum_i \bar{y}_i^2, \tag{13}$$

$$\sum_i \breve{w}_i^2 = \bar{w}'P' \sum_i e_{k,n}\, e'_{k,n}\, P\bar{w} = \bar{w}'P'P\bar{w} = \bar{w}'P\bar{w} \leq \bar{w}'\bar{w} = \sum_i \bar{w}_i^2. \tag{14}$$

By independent observations conditional on Z, we have

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

To show that $A_n = O_p(1)$, it suffices to show that there exists a positive constant C such that $E\left[|A_n|\right] \leq C < \infty$ for n sufficiently large. To proceed, note first that

$$
\begin{aligned}
& E_{\mathcal{Z}}\left[\left|\sum_k \breve{w}_k \bar{\eta}_k \breve{y}_k\right|\right] \leq E_{\mathcal{Z}}\left[\left|\sum_k \bar{\eta}_k \overline{w}' P e_{k,n}\, e_{k,n}'\, P \bar{y}\right|\right] \\
\leq\ & E_{\mathcal{Z}}\left[\sqrt{\sum\nolimits_k \bar{w}' P e_{k,n}\, e_{k,n}'\, P \bar{w}}\sqrt{\sum\nolimits_k \bar{\eta}_k^2 \bar{y}' P e_{k,n}\, e_{k,n}'\, P \bar{y}}\right] \leq E_{\mathcal{Z}}\left[\max_{1 \leq k \leq n} |\bar{\eta}_k| \sqrt{\bar{w}' P \bar{w}} \sqrt{\bar{y}' P \bar{y}}\right] \\
\leq\ & E_{\mathcal{Z}}\left[\bar{\mu}_\eta \sqrt{\sum_i \bar{w}_i^2} \sqrt{\sum_j \bar{y}_j^2}\right] \leq \frac{1}{2} E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] + \frac{1}{2} E_{\mathcal{Z}}\left[\sum_i \bar{w}_i^2 \sum_j \bar{y}_j^2\right] \leq C,
\end{aligned}
$$

where the first inequality follows from CS, the third inequality follows from (13) and (14), the fourth inequality follows from inequality $2E|XY| \leq E\left(X^2\right) + E\left(Y^2\right)$, and the last inequality follows by hypothesis. In a similar manner, we have that

$$
\begin{aligned}
& E_{\mathcal{Z}}\left[\left|\sum_i \bar{w}_i P_{ii} \bar{\eta}_i \breve{y}_i\right|\right] \leq E_{\mathcal{Z}}\left[\sqrt{\sum\nolimits_i \bar{w}_i^2 P_{ii}^2} \sqrt{\sum\nolimits_k \bar{\eta}_k^2 \bar{y}' P e_{k,n}\, e_{k,n}'\, P \bar{y}}\right] \\
\leq\ & E_{\mathcal{Z}}\left[\max_{1 \leq k \leq n} |\bar{\eta}_k| \sqrt{\sum_i \bar{w}_i^2} \sqrt{\sum_k \bar{y}_k^2}\right] \leq \frac{1}{2} E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] + \frac{1}{2} E_{\mathcal{Z}}\left[\sum_i \bar{w}_i^2 \sum_j \bar{y}_j^2\right] \leq C,
\end{aligned}
$$

and $E_{\mathcal{Z}}\left[\left|\sum_i \breve{w}_i P_{ii} \bar{\eta}_i \bar{y}_i\right|\right]$ is bounded in the same way as well. Moreover, using Lemma B1 and the inequality $2E|XY| \leq E\left(X^2\right) + E\left(Y^2\right)$, we obtain

$$
\begin{aligned}
& E_{\mathcal{Z}}\left[\left|\sum_{i,k} \bar{w}_i \bar{y}_i P_{ik}^2 \bar{\eta}_k\right|\right] \leq E_{\mathcal{Z}}\left[\sum_{i,k} \left|\bar{w}_i \bar{y}_i P_{ik}^2 \bar{\eta}_k\right|\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_W \bar{\mu}_Y \bar{\mu}_\eta \sum_{i,k} P_{ik}^2\right] \\
=\ & K E_{\mathcal{Z}}\left[\bar{\mu}_W \bar{\mu}_Y \bar{\mu}_\eta\right] = E_{\mathcal{Z}}\left[\bar{\mu}_\eta \sqrt{K} \bar{\mu}_W \sqrt{K} \bar{\mu}_Y\right] \\
\leq\ & \frac{1}{2} E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] + \frac{1}{2} E_{\mathcal{Z}}\left[K \bar{\mu}_W^2 K \bar{\mu}_Y^2\right] \leq \frac{1}{2} E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] + \frac{1}{4} K^2 E_{\mathcal{Z}}\left[\bar{\mu}_W^4\right] + \frac{1}{4} K^2 E_{\mathcal{Z}}\left[\bar{\mu}_Y^4\right] \leq C.
\end{aligned}
$$

Also, by similar argument, $E_{\mathcal{Z}}\left[\left|\sum_i \bar{w}_i \bar{y}_i P_{ii}^2 \bar{\eta}_i\right|\right] \leq C$. Thus, $E_{\mathcal{Z}}\left[|A_n|\right] \leq C$ holds by T, from which it then follows by M that $A_n = O_p(1)$.

Next, note that

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

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 $\hat{\psi}_r \xrightarrow{p} 0$, $(r = 1, ..., 7)$. 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,

$$KE_{\mathcal{Z}}\left[\bar{\sigma}_W^2\bar{\sigma}_\eta^2\bar{\sigma}_Y^2\right] \leq C\left(K/r_n^2\right)E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\right] \leq C\left(K/r_n^2\right) \longrightarrow 0.$$

By Lemma B1, we have

$$\begin{aligned}|\hat{\tau}_1| &= \sum_{i\neq j\neq k} E[(\tilde{W}_i P_{ik}\tilde{\eta}_k P_{kj}\tilde{Y}_j)^2] = E_{\mathcal{Z}}\left(\sum_{i\neq j\neq k} E[\tilde{W}_i^2|\mathcal{Z}]P_{ik}^2 E[\tilde{\eta}_k^2|\mathcal{Z}]P_{kj}^2 E[\tilde{Y}_j^2|\mathcal{Z}]\right)\\ &\leq KE_{\mathcal{Z}}\left[\bar{\sigma}_W^2\bar{\sigma}_\eta^2\bar{\sigma}_Y^2\right] \leq C\left(K/r_n^2\right) \longrightarrow 0.\end{aligned}$$

Similarly, by CS,

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

Next, by Lemma B1 and CS

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

Similarly,

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

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

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

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

$$
\begin{aligned}
& E_{\mathcal{Z}}\left\{\left[\sum\nolimits_{i\neq k} P_{ik}^2\left(\sum\nolimits_{j\notin\{i,k\}} P_{kj}\bar{y}_j\right)^2\right]^2\right\} \\
\leq\ & 9E_{\mathcal{Z}}\left\{\left[\sum\nolimits_{i\neq k} P_{ik}^2\left(\breve{y}_k^2 + P_{ki}^2\bar{y}_i^2 + P_{kk}^2\bar{y}_k^2\right)\right]^2\right\} \\
\leq\ & 9E_{\mathcal{Z}}\left(\sum\nolimits_k P_{kk}\left(\breve{y}_k^2 + 2\bar{y}_k^2\right)\right)^2 \leq 9E_{\mathcal{Z}}\left(\sum\nolimits_k \breve{y}_k^2 + 2\sum\nolimits_k \bar{y}_k^2\right)^2 \leq 81E_{\mathcal{Z}}\left[\left(\sum\nolimits_k \bar{y}_k^2\right)^2\right] \leq C.
\end{aligned}
$$

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

$$
\begin{aligned}
E[\hat{\psi}_2^2] &= E_{\mathcal{Z}}\left\{\sum\nolimits_{i\neq k} E[\tilde{W}_i^2|\mathcal{Z}]P_{ik}^2E[\tilde{\eta}_k^2|\mathcal{Z}]\left(\sum\nolimits_{j\notin\{i,k\}} P_{kj}\bar{y}_j\right)^2\right. \\
&\qquad \left. + \sum_{i\neq k} E[\tilde{W}_i\tilde{\eta}_i|\mathcal{Z}]P_{ik}^2E[\tilde{W}_k\tilde{\eta}_k|\mathcal{Z}]\left(\sum\nolimits_{j\notin\{i,k\}} P_{kj}\bar{y}_j\right)\left(\sum\nolimits_{j\notin\{i,k\}} P_{ij}\bar{y}_j\right)\right\} \\
&\leq 2E_{\mathcal{Z}}\left[\bar{\sigma}_W^2\bar{\sigma}_\eta^2\sum\nolimits_{i\neq k} P_{ik}^2\left(\sum\nolimits_{j\notin\{i,k\}} P_{kj}\bar{y}_j\right)^2\right] \\
&\leq \frac{C}{r_n}\left\{\sqrt{E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^4\right]}\sqrt{E_{\mathcal{Z}}\left[\sum\nolimits_{i\neq k} P_{ik}^2\left(\sum\nolimits_{j\notin\{i,k\}} P_{kj}\bar{y}_j\right)^2\right]^2}\right\} \leq \frac{C}{r_n} \longrightarrow 0.
\end{aligned}
$$

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 $E[\tilde{W}_iP_{ik}\bar{\eta}_kP_{kj}\tilde{Y}_j\tilde{W}_rP_{rs}\bar{\eta}_sP_{st}\tilde{Y}_t] = 0$ except when $i = r$ and $j = t$ or $i = t$ and $j = r$. Thus,

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

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

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

It therefore follows that

$$
\begin{aligned}
& E_{\mathcal{Z}}\left[\sum_{i\neq j}\left(\sum_k P_{ik}\bar{\eta}_k P_{kj}\right)^2\right] \\
= \; & E_{\mathcal{Z}}\left[\sum_{i,j}\left(\sum_k P_{ik}\bar{\eta}_k P_{kj}\right)^2\right] - E_{\mathcal{Z}}\left[\sum_i\left(\sum_k P_{ik}\bar{\eta}_k P_{ki}\right)^2\right] \leq 2K E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right].
\end{aligned}
$$

Also, by Lemma B1, $E_{\mathcal{Z}}\left[\sum_{i\neq j} P_{ij}^2 P_{jj}^2\bar{\eta}_j^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_{i\neq j} P_{ij}^2\right] \leq K E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2(\mathcal{Z})\right]$, so that

$$
\begin{aligned}
& E_{\mathcal{Z}}\left[\sum_{i\neq j}\left(\sum_{k\notin\{i,j\}} P_{ik}\bar{\eta}_k P_{kj}\right)^2\right] \\
\leq \; & 3E_{\mathcal{Z}}\left[\sum_{i\neq j}\left\{\left(\sum_{k=1}^n P_{ik}\bar{\eta}_k P_{kj}\right)^2 + P_{ij}^2 P_{ii}^2\bar{\eta}_i^2 + P_{ij}^2 P_{jj}^2\bar{\eta}_j^2\right\}\right] \leq 12K E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right].
\end{aligned}
$$

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

$$
E[\hat{\psi}_3^2] \leq \frac{C}{r_n^2} E_{\mathcal{Z}}\left[\sum_{i\neq j}\left(\sum_{k\notin\{i,j\}} P_{ik}\bar{\eta}_k P_{kj}\right)^2\right] \leq C\left(\frac{K}{r_n^2}\right) E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] \leq C\left(\frac{K}{r_n^2}\right) \longrightarrow 0.
$$

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,

$$\begin{aligned}
E[\hat{\psi}_4^2] &= E_{\mathcal{Z}}\left[\sum_i E[\tilde{W}_i^2|\mathcal{Z}]\left(\sum_{j\neq i}\sum_{k\notin\{i,j\}} P_{ik}\bar{\eta}_k P_{kj}\bar{y}_j\right)^2\right] \\
&\leq E_{\mathcal{Z}}\left[\bar{\sigma}_W^2(\mathcal{Z})\sum_i\left(\sum_{j\neq i}\sum_{k\notin\{i,j\}} P_{ik}\bar{\eta}_k P_{kj}\bar{y}_j\right)^2\right] \\
&\leq \frac{C}{r_n}E_{\mathcal{Z}}\left[\sum_i\left(\sum_{j\neq i}\sum_{k\notin\{i,j\}} P_{ik}\bar{\eta}_k P_{kj}\bar{y}_j\right)^2\right].
\end{aligned}$$

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,

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

Note that by P idempotent we have

$$\begin{aligned}
&E_{\mathcal{Z}}\left[\sum_j\sum_k P_{jk}\bar{\eta}_j\breve{y}_j\bar{\eta}_k\breve{y}_k\right] \\
\leq\ & E_{\mathcal{Z}}\left[\sum_j \bar{\eta}_j^2\breve{y}_j^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_j\breve{y}_j^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_j\bar{y}_j^2\right] \leq CE_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] \leq C,
\end{aligned}$$

where the second to last inequality holds for n sufficiently large since by hypothesis $\sum_j \bar{y}_j^2 \leq C \ \ a.s.\ P_{\mathcal{Z}}$ for large enough n. Then, it follows that

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

Also, using similar reasoning,

$$\begin{aligned}
E_{\mathcal{Z}}\left[\sum_i (P_{ii}\bar{\eta}_i\breve{y}_i)^2\right] &\leq E_{\mathcal{Z}}\left[\sum_i \bar{\eta}_i^2\breve{y}_i^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i \breve{y}_i^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i \bar{y}_i^2\right] \leq CE_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] \leq C, \\
E_{\mathcal{Z}}\left[\sum_i\left(\sum_j P_{ij}\bar{\eta}_j P_{jj}\bar{y}_j\right)^2\right] &\leq E_{\mathcal{Z}}\left[\sum_i \bar{\eta}_i^2 P_{ii}^2\bar{y}_i^2\right] \leq E_{\mathcal{Z}}\left[\sum_i \bar{\eta}_i^2\bar{y}_i^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i \bar{y}_i^2\right] \leq C, \\
E_{\mathcal{Z}}\left[\sum_i\left(\bar{y}_i\sum_k P_{ik}^2\bar{\eta}_k\right)^2\right] &\leq E_{\mathcal{Z}}\left[\sum_i \bar{y}_i^2\sum_{k,\ell} P_{ik}^2P_{i\ell}^2\left|\bar{\eta}_k\right|\left|\bar{\eta}_\ell\right|\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i \bar{y}_i^2\sum_{k,\ell}P_{ik}^2P_{i\ell}^2\right] \\
&= E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i \bar{y}_i^2P_{ii}^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i \bar{y}_i^2\right] \leq CE_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] \leq C, \\
E_{\mathcal{Z}}\left[\sum_i P_{ii}^4\bar{\eta}_i^2\bar{y}_i^2\right] &\leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i P_{ii}^4\bar{y}_i^2\right] \leq E_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\sum_i \bar{y}_i^2\right] \leq CE_{\mathcal{Z}}\left[\bar{\mu}_\eta^2\right] \leq C.
\end{aligned}$$

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

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

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

Then for fixed k,

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

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] \\
= \;& E_{\mathcal{Z}}\left[\sum_k E\left[\tilde{\eta}_k^2|\mathcal{Z}\right]\left(\sum_{i\neq k}\sum_{j\notin\{i,k\}}\bar{w}_i P_{ik}P_{kj}\bar{y}_j\right)^2\right] \\
\leq \;& 5E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\sum_k\left\{\breve{w}_k^2\breve{y}_k^2+\sum_{i,j}P_{kj}^2P_{ki}^2\bar{w}_i\bar{y}_i\bar{w}_j\bar{y}_j\right\}\right] \\
& +5E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\sum_k\left\{\breve{w}_k^2P_{kk}^2\bar{y}_k^2+\bar{w}_k^2P_{kk}^2\breve{y}_k^2+\bar{w}_k^2P_{kk}^4\bar{y}_k^2\right\}\right] \\
\leq \;& 5E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\left(\sum_k\breve{w}_k^2\breve{y}_k^2+\bar{\mu}_W^2\bar{\mu}_Y^2\sum_{i,j,k}P_{kj}^2P_{ki}^2\right)\right] \\
& +5E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\left(\bar{\mu}_Y^2\sum_k\bar{w}_k+\bar{\mu}_W^2\sum_k\bar{y}_k^2+\bar{\mu}_W^2\sum_k P_{kk}^4\bar{y}_k^2\right)\right] \\
\leq \;& 5E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\sum_k\breve{w}_k^2\breve{y}_k^2\right]+5KE_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\bar{\mu}_W^2\bar{\mu}_Y^2\right]+15E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\bar{\mu}_Y^2\sum_k\bar{w}_k^2\right] \\
\leq \;& 5E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\left(\max_{1\leq k\leq n}|\breve{w}_k|\right)^2\sum_k\breve{y}_k^2\right]+5KE_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\bar{\mu}_W^2\bar{\mu}_Y^2\right]+15CE_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\bar{\mu}_Y^2\right] \\
\leq \;& 5E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\left\{\frac{1}{2}\left(\max_{1\leq k\leq n}|\breve{w}_k-\bar{w}_k|\right)^2+\frac{1}{2}\left(\max_{1\leq k\leq n}|\bar{w}_k|\right)^2\right\}\sum_k\bar{y}_k^2\right]++5KE_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\bar{\mu}_W^2\bar{\mu}_Y^2\right] \\
& +15C\sqrt{E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^4\right]}\sqrt{E_{\mathcal{Z}}\left[\bar{\mu}_Y^4\right]} \\
\leq \;& CE_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\left(\max_{1\leq k\leq n}|\breve{w}_k-\bar{w}_k|\right)^2\right]+CE_{\mathcal{Z}}\left[\bar{\sigma}_\eta^2\left(\max_{1\leq k\leq n}|\bar{w}_k|\right)^2\right]+o\left(1\right) \\
\leq \;& C\sqrt{E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^4\right]}\sqrt{E_{\mathcal{Z}}\left[\left(\max_{1\leq k\leq n}|\breve{w}_k-\bar{w}_k|\right)^4\right]}+C\sqrt{E_{\mathcal{Z}}\left[\bar{\sigma}_\eta^4\right]}\sqrt{E_{\mathcal{Z}}\left[\left(\frac{1}{\sqrt{n}}\max_{1\leq i\leq n}|a_i\left(\mathcal{Z}\right)|\right)^4\right]}+o(1) \\
\leq \;& C\sqrt{E_{\mathcal{Z}}\left[\left(\max_{1\leq k\leq n}|\breve{w}_k-\bar{w}_k|\right)^4\right]}+o(1)
\end{aligned}
$$

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

$$
\begin{aligned}
\max_i|\bar{w}_i-\breve{w}_i| &= \max_i\left|\bar{w}_i-Z_i'(Z'Z)^{-1}Z'\bar{w}\right| = \max_i\left|\bar{w}_i-Z_i'\alpha_n-Z_i'(Z'Z)^{-1}Z'(\bar{w}-Z\alpha_n)\right| \\
&\leq \Delta_n/\sqrt{n}+(\max_i\sum_j P_{ij}^2)^{1/2}(\sum_j\max_i\left[\bar{w}_i-Z_i'\alpha_n\right]^2)^{1/2}\leq C\Delta_n\to 0 \quad a.s.\mathbb{P}_{\mathcal{Z}}.
\end{aligned}
$$

Moreover, note that by Assumption 6, there exists a positive integer N and a real number $\epsilon > 0$ such that

$$\begin{aligned}\sup_{n \geq N} E_{\mathcal{Z}} \left(\max_{1 \leq k \leq n} |\breve{w}_k - \bar{w}_k| \right)^{4+\epsilon} &\leq C \left[\sup_{n \geq N} E_{\mathcal{Z}} \left(\max_{1 \leq k \leq n} |\breve{w}_k| \right)^{4+\epsilon} + \sup_{n \geq N} E_{\mathcal{Z}} \left(\max_{1 \leq k \leq n} |\bar{w}_k| \right)^{4+\epsilon} \right] \\ &\leq C\end{aligned}$$

It, thus, follows by dominated convergence that

$$E_{\mathcal{Z}} \left[\left(\max_{1 \leq k \leq n} |\breve{w}_k - \bar{w}_k| \right)^4 \right] \leq C E_{\mathcal{Z}} \left[\Delta_n^4 \right] \rightarrow 0,$$

so that we have $E[\hat{\psi}_6^2] \longrightarrow 0$; and the desired result follows by T. Q.E.D.

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.

Billingsley, P (1986). *Probability and Measure*, 2nd Edition. (John Wiley & Sons, New York).

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 (2004). "Estimation and Testing Using Jackknife IV in Heteroskedastic Regressions With Many Weak Instruments", Working Paper, Rutgers University.

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

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," *Journal of Business and Economic Statistics*, forthcoming.

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.