ASYMPTOTIC DISTRIBUTION OF JIVE IN A HETEROSKEDASTIC IV REGRESSION WITH MANY INSTRUMENTS

Econometric Theory, **28**, 2012, 42–86.
doi:10.1017/S0266466611000120

ASYMPTOTIC DISTRIBUTION OF JIVE IN A HETEROSKEDASTIC IV REGRESSION WITH MANY INSTRUMENTS

JOHN C. CHAO
University of Maryland

NORMAN R. SWANSON
Rutgers University

JERRY A. HAUSMAN AND WHITNEY K. NEWEY
MIT

TIEMEN WOUTERSEN
Johns Hopkins University

This paper derives the limiting distributions of alternative jackknife instrumental variables (JIV) estimators and gives formulas for accompanying consistent standard errors in the presence of heteroskedasticity and many instruments. The asymptotic framework includes the many instrument sequence of Bekker (1994, *Econometrica* 62, 657–681) and the many weak instrument sequence of Chao and Swanson (2005, *Econometrica* 73, 1673–1691). We show that JIV estimators are asymptotically normal and that standard errors are consistent provided that $\sqrt{K_n}/r_n \to 0$ as $n \to \infty$, where K_n and r_n denote, respectively, the number of instruments and the concentration parameter. This is in contrast to the asymptotic behavior of such classical instrumental variables estimators as limited information maximum likelihood, bias-corrected two-stage least squares, and two-stage least squares, all of which are inconsistent in the presence of heteroskedasticity, unless $K_n/r_n \to 0$. 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 the strength of the instruments as measured by the relative orders of magnitude of r_n and K_n.

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, N. Lott, M. Moriera, G.D.A. Phillips, P.C.B. Phillips, J. Stock, J. Wright, two anonymous referees, and a co-editor for helpful comments and suggestions. Address correspondence to: Whitney K. Newey, Department of Economics, MIT, E52-262D, Cambridge, MA 02142-1347, USA; e-mail: wnewey@mit.edu.

1. INTRODUCTION

It has long been known that the two-stage least squares (2SLS) estimator is biased with many instruments (see, e.g., Sawa, 1968; Phillips, 1983; and the references cited therein). In large part because of this problem, various approaches have been proposed in the literature to reduce the bias of the 2SLS estimator. In recent years, there has been interest in developing procedures that use "delete-one" fitted values in lieu of the usual first-stage ordinary least squares fitted values as the instruments employed in the second stage of the 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 Devereux (2009), 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, whereas 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 Devereux, 2009; Chao and Swanson, 2006; Hausman 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), Morimune (1983), 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 $\sqrt{K_n}/r_n \to 0$ as $n \to \infty$, where K_n and r_n denote the number of instruments and 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 the identification strength. We show that both the rate of convergence of the JIV estimator and the form of its asymptotic covariance matrix depend 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 the standard errors under heteroskedasticity and many instruments.

Hausman et al. (2007) also consider a jackknife form of LIML that is slightly more difficult to compute but is asymptotically efficient relative to JIV under many weak instruments and homoskedasticity. With heteroskedasticity, any of the estimators may outperform the others, as shown by Monte Carlo examples in Hausman et al. Hausman et al. also propose a jackknife version of the Fuller (1977) estimator that has fewer outliers.

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). 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 the jackknife versions of the LIML and Fuller (1977) estimators and in Chao, Hausman, Newey, Swanson, and Woutersen (2010) to derive a moment-based test.

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 the Appendixes.

2. THE MODEL AND ESTIMATORS

The model we consider is given by

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

where n is the number of observations, G is the number of right-hand-side variables, Υ is the reduced form matrix, and U is the disturbance matrix. For the asymptotic approximations, the elements of Υ will implicitly be allowed to depend on n, although we suppress the 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 $\text{rank}(Z) = K$. 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 ith row (observation) for X, Υ, and Z, respectively. We could let $\Upsilon_i = f_0(w_i)$ be a vector of unknown functions of a vector w_i of underlying instruments and let $Z_i = (p_{1K}(w_i), \ldots, p_{KK}(w_i))'$ 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., Newey, 1990).

To describe the estimators, let $P = Z(Z'Z)^{-1}Z'$ and P_{ij} denote the (i, j)th element of P. Additionally, let $\bar{\Pi}_{-i} = (Z'Z - Z_iZ_i')^{-1}(Z'X - Z_iX_i')$ be the reduced form coefficients obtained by regressing X on Z using all observations except the ith. The JIV1 estimator of Phillips and Hale (1977) is obtained as

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

Using standard results on recursive residuals, it follows that

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

Then, we have that

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

where $\Sigma_{i \neq j}$ denotes the double sum $\sum_i \sum_{j \neq i}$. The JIV2 estimator proposed by Angrist et al. (1999), JIVE2, has a similar form, except that $\Pi_{-i} = (Z'Z)^{-1} (Z'X - Z_i X_i')$ is used in place of $\bar{\Pi}_{-i}$. It is given by

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

To explain why JIV2 is a consistent estimator, it is helpful to consider JIV2 as a minimizer of an objective function. As usual, the limit of the minimizer will be the minimizer of the limit under appropriate regularity conditions. We focus on $\hat{\delta}$ to simplify the discussion. The estimator $\hat{\delta}$ satisfies $\hat{\delta} = \arg\min_\delta \hat{Q}(\delta)$, where

$$\hat{Q}(\delta) = \sum_{i \neq j} (y_i - X_i'\delta) P_{ij} (y_j - X_j'\delta).$$

Note that the difference between the 2SLS objective function $(y - X'\delta)P(y - X'\delta)$ and $\hat{Q}(\delta)$ is $\sum_{i=1}^n P_{ii}(y_i - X_i'\delta)^2$. This is a weighted least squares object that is a source of bias in 2SLS because its expectation is not minimized at δ_0 when X_i and ε_i are correlated. This object does not vanish asymptotically relative to $E[\hat{Q}(\delta)]$ under many (or many weak) instruments, leading to inconsistency of 2SLS. When observations are mutually independent, the inconsistency is caused by this term, so removing it to form $\hat{Q}(\delta)$ makes $\hat{\delta}$ consistent.

To explain further, consider the JIV2 objective function $\hat{Q}(\delta)$. Note that for $\tilde{U}_i(\delta) = \varepsilon_i - U_i'(\delta - \delta_0)$

$$\hat{Q}(\delta) = \hat{Q}_1(\delta) + \hat{Q}_2(\delta) + \hat{Q}_3(\delta), \qquad \hat{Q}_1(\delta) = \sum_{i \neq j} (\delta - \delta_0)' \Upsilon_i P_{ij} \Upsilon_j' (\delta - \delta_0),$$

$$\hat{Q}_2(\delta) = -2 \sum_{i \neq j} \tilde{U}_i(\delta) P_{ij} \Upsilon_j' (\delta - \delta_0), \qquad \hat{Q}_3(\delta) = \sum_{i \neq j} \tilde{U}_i(\delta) P_{ij} \tilde{U}_j(\delta).$$

Then by the assumptions $E[\tilde{U}_i(\delta)] = 0$ and independence of observations, we have $E[\hat{Q}(\delta)|\mathcal{Z}] = Q_1(\delta)$. Under the regularity conditions in Section 3, $\sum_{i \neq j} \Upsilon_i P_{ij} \Upsilon_j'$ is positive definite asymptotically, so $Q_1(\delta)$ is minimized at δ_0. Thus, the expectation $Q_1(\delta)$ of $\hat{Q}(\delta)$ is minimized at the true parameter δ_0; in the terminology of Han and Phillips (2006), the many instrument "noise" term in the expected objective function is identically zero.

For consistency of $\hat{\delta}$, it is also necessary that the stochastic components of $\hat{Q}(\delta)$ do not dominate asymptotically. The size of $\hat{Q}_1(\delta)$ (for $\delta \neq \delta_0$) is proportional to the concentration parameter that we denote by r_n. It turns out that $\hat{Q}_2(\delta)$ has size smaller than $\hat{Q}_1(\delta)$ asymptotically but $\hat{Q}_3(\delta)$ is $O_p(\sqrt{K_n})$ (Lemma A1 shows that the variance of $\hat{Q}_3(\delta)$ is proportional to K_n). Thus, to ensure that the expectation of $\hat{Q}(\delta)$ dominates the stochastic part of $\hat{Q}(\delta)$, it suffices to impose the restriction $\sqrt{K_n}/r_n \to 0$, which we do throughout the asymptotic theory. This condition was formulated in Chao and Swanson (2005).

The estimators $\tilde{\delta}$ and $\hat{\delta}$ are consistent and asymptotically normal with heteroskedasticity under the regularity conditions we impose, including $\sqrt{K_n}/r_n \to 0$. In contrast, consistency of LIML and Fuller (1977) require $K_n/r_n \to 0$ when P_{ii} is asymptotically correlated with $\mathrm{E}[X_i\varepsilon_i|\mathcal{Z}]/\mathrm{E}[\varepsilon_i^2|\mathcal{Z}]$, as discussed in Chao and Swanson (2004) and Hausman et al. (2007). This condition is also required for consistency of the bias-corrected 2SLS estimator of Donald and Newey (2001) when P_{ii} is asymptotically correlated with $\mathrm{E}[X_i\varepsilon_i|\mathcal{Z}]$, as discussed in Ackerberg and Devereux (2009). Thus, JIV estimators are robust to heteroskedasticity and many instruments (when K_n grows as fast as r_n), whereas LIML, Fuller (1977), or B2SLS estimators are not.

Hausman et al. (2007) also consider a JIV form of LIML, which is obtained by minimizing $\hat{Q}(\delta)/[(y - X\delta)'(y - X\delta)]$. The sum of squared residuals in the denominator makes computation somewhat more complicated; however, like LIML, it has an explicit form in terms of the smallest eigenvalue of a matrix. This JIV form of LIML is asymptotically efficient relative to $\hat{\delta}$ and $\tilde{\delta}$ under many weak instruments and homoskedasticity. With heteroskedasticity, $\hat{\delta}$ and $\tilde{\delta}$ may perform better than this estimator, as shown by Monte Carlo examples in Hausman et al.; they also propose a jackknife version of the Fuller (1977) estimator that has fewer outliers than the JIV form of LIML.

To motivate the form of the variance estimator for $\hat{\delta}$ and $\tilde{\delta}$, note 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 $\sum_{i \neq j} X_i P_{ij} \xi_j$,which leads 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 with an estimate, $\tilde{\xi}_i = (1 - P_{ii})^{-1}\left(y_i - X_i'\tilde{\delta}\right)$. Using the independence of the observations, $\mathrm{E}[\varepsilon_i|\mathcal{Z}] = 0$, and the exclusion of the $i = j$ terms in the double sums, it follows that

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

Removing the expectation and replacing ξ_i with $\tilde{\xi}_i$ gives

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

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 $\text{Var}(\xi_i|\mathcal{Z})$ and $\text{E}[X_i\xi_i|\mathcal{Z}]$ to vary over i.

A vectorized form of $\tilde{V}$ is easier to compute. Note that for $\tilde{X}_i = X_i/(1-P_{ii})$, we have $\tilde{H} = X'P\tilde{X} - \sum_i X_i P_{ii}\tilde{X}_i'$. Also, let $\bar{X} = PX$, $\tilde{Z} = Z(Z'Z)^{-1}$, and Z_i' and $\tilde{Z}_i'$ equal the ith row of Z and $\tilde{Z}$, respectively. Then, as shown in the proof of Theorem 4, we have

$$\tilde{\Sigma} = \sum_{i=1}^{n}(\bar{X}_i\bar{X}_i' - X_i P_{ii}\bar{X}_i' - \bar{X}_i P_{ii} X_i')\hat{\tilde{\xi}}_i^2 + \sum_{k=1}^{K}\sum_{\ell=1}^{K}\left(\sum_{i=1}^{n}\tilde{Z}_{ik}\tilde{Z}_{i\ell}X_i\hat{\tilde{\xi}}_i\right)\left(\sum_{j=1}^{n}Z_{jk}Z_{j\ell}X_j\hat{\tilde{\xi}}_j\right)'.$$

This formula can be computed quickly by software with fast vector operations, even when n is large.

An asymptotic variance estimator for $\hat{\delta}$ can be formed in an analogous way. Note that $\hat{H} = X'PX - \sum_i X_i P_{ii} X_i'$. Also for $\hat{\varepsilon}_i = y_i - X_i'\hat{\delta}$, we can estimate the middle matrix of the sandwich by

$$\hat{\Sigma} = \sum_{i=1}^{n}(\bar{X}_i\bar{X}_i' - X_i P_{ii}\bar{X}_i' - \bar{X}_i P_{ii} X_i')\hat{\varepsilon}_i^2 + \sum_{k=1}^{K}\sum_{\ell=1}^{K}\left(\sum_{i=1}^{n}\tilde{Z}_{ik}\tilde{Z}_{i\ell}X_i\hat{\varepsilon}_i\right)\left(\sum_{j=1}^{n}Z_{jk}Z_{j\ell}X_j\hat{\varepsilon}_j\right)'.$$

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 a transpose is not needed for the third matrix in $\hat{V}$.

3. MANY INSTRUMENT ASYMPTOTICS

Our asymptotic theory combines the many instrument asymptotics of Kunitomo (1980), Morimune (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 ith row of Z, ε, U, and Υ, respectively. Also let a.s. denote almost surely (i.e., with probability one) and a.s.n denote a.s. for n large enough (i.e., with probability one for all n large enough).

Assumption 1. $K = K_n \to \infty$, Z includes among its columns a vector of ones, for some $C < 1$, $\text{rank}(Z) = K$, and $P_{ii} \le C$, $(i = 1, \ldots, n)$ a.s.n.

In this paper, C is a generic notation for a positive constant that may be bigger or less than 1. Hence, although in Assumption 1 C is taken to be less than 1, in other parts of the paper it might not be. The restriction that $\text{rank}(Z) = K$ is a normalization that requires excluding redundant columns from Z. It can be verified in particular cases. For instance, when w_i is a continuously distributed scalar, $Z_i = p^K(w_i)$, and $p_{kK}(w) = w^{k-1}$, it can be shown that $Z'Z$ is nonsingular with probability one for $K < n$.[1] The condition $P_{ii} \le C < 1$ implies that $K/n \le C$ because $K/n = \sum_{i=1}^n P_{ii}/n \le C$.

Now, let $\lambda_{\min}(A)$ denote the smallest eigenvalue of a symmetric matrix A and for any matrix B, let $\|B\| = \sqrt{\text{tr}(B'B)}$.

Assumption 2. $\Upsilon_i = S_n z_i/\sqrt{n}$ where $S_n = \tilde{S}_n \,\text{diag}(\mu_{1n}, \ldots, \mu_{Gn})$, $\tilde{S}_n$ is $G \times G$ and bounded, and the smallest eigenvalue of $\tilde{S}_n \tilde{S}_n'$ is bounded away from zero. Also, for each j, either $\mu_{jn} = \sqrt{n}$ or $\mu_{jn}/\sqrt{n} \to 0$, $r_n = \left(\min_{1 \le j \le G} \mu_{jn}\right)^2 \to \infty$, and $\sqrt{K}/r_n \to 0$. Also, there is $C > 0$ such that $\left\|\sum_{i=1}^n z_i z_i'/n\right\| \le C$ and $\lambda_{\min}\left(\sum_{i=1}^n z_i z_i'/n\right) \ge 1/C$ a.s.n.

This condition is similar to Assumption 2 of Hansen, Hausman, and Newey (2008). It accommodates linear models where included instruments (e.g., a constant) have fixed reduced form coefficients and excluded instruments have coefficients that can shrink as the sample size grows. A leading example of such a model is a linear structural equation with one endogenous variable of the form

$$y_i = Z_{i1}'\delta_{01} + \delta_{0G} X_{iG} + \varepsilon_i, \tag{2}$$

where Z_{i1} is a $G_1 \times 1$ vector of included instruments (e.g., including a constant) and X_{iG} is an endogenous variable. Here the number of right-hand-side variables is $G_1 + 1 = G$. Let the reduced form be partitioned conformably with δ, as $\Upsilon_i = (Z_{i1}', \Upsilon_{iG})'$ and $U_i = (0, U_{iG})'$. Here the disturbances for the reduced form for Z_{i1} are zero because Z_{i1} is taken to be exogenous. Suppose that the reduced form for X_{iG} depends linearly on the included instrumental variables Z_{i1} and on an excluded instrument z_{iG} as in

$$X_{iG} = \Upsilon_{iG} + U_{iG}, \qquad \Upsilon_{iG} = \pi_1 Z_{i1} + \left(\sqrt{r_n/n}\right) z_{iG}.$$

Here we normalize z_{iG} so that r_n determines how strongly δ_G is identified, and we absorb into z_{iG} any other terms, such as unknown coefficients. For Assumption 2, we let $z_i = (Z_{i1}', z_{iG})'$ and require that the second moment matrix of z_i is bounded and bounded away from zero. This normalization allows r_n to determine the strength of identification of δ_G. For example, if $r_n = n$, then the coefficient on z_{iG} does not shrink, which corresponds to strong identification of δ_G. If r_n grows more slowly than n, then δ_G will be more weakly identified. Indeed, $1/\sqrt{r_n}$ will be the convergence rate for estimators of δ_G. We require $r_n \to \infty$ to avoid the weak instrument setting of Staiger and Stock (1997), where δ_G is not asymptotically identified.

For this model, the reduced form is

$$\Upsilon_i = \begin{bmatrix} Z_{i1} \\ \pi_1 Z_{i1} + \sqrt{r_n/n} z_{iG} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \pi_1 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \sqrt{r_n/n} \end{bmatrix} \begin{pmatrix} Z_{i1} \\ z_{iG} \end{pmatrix}.$$

This reduced form is as specified in Assumption 2 with

$$\tilde{S}_n = \begin{bmatrix} I & 0 \\ \pi_1 & 1 \end{bmatrix}, \qquad \mu_{jn} = \sqrt{n}, \qquad 1 \le j \le G_1, \qquad \mu_{Gn} = \sqrt{r_n}.$$

Note how this somewhat complicated specification is needed to accommodate fixed reduced form coefficients for included instrumental variables and excluded instruments with identifying power that depend on n. We have been unable to simplify Assumption 2 while maintaining the generality needed for such important cases.

We will not require that z_{iG} be known, only that it be approximated by a linear combination of the instrumental variables $Z_i = (Z_{i1}', Z_{i2}')'$. Implicitly, Z_{i1} and z_{iG} are allowed to depend on n. One important case is where the excluded instrument z_{iG} is an unknown linear combination of the instrumental variables $Z_i = (Z_{i1}', Z_{i2}')'$. For example, the many weak instrument setting of Chao and Swanson (2005) is one where the reduced form is given by

$$\Upsilon_{iG} = \pi_1 Z_{i1} + (\pi_2/\sqrt{n})' Z_{i2}$$

for a $K - G_1$ dimensional vector Z_{i2} of excluded instrumental variables. This model can be folded into our framework by specifying that

$$z_{iG} = \pi_2' Z_{i2} \Big/ \sqrt{K - G_1}, \qquad r_n = K - G_1.$$

Assumption 2 will then require that

$$\sum_i z_{iG}^2/n = (K - G_1)^{-1} \sum_i (\pi_2' Z_{i2})^2 \Big/ n$$

is bounded and bounded away from zero. Thus, the second moment $\sum_i (\pi_2' Z_{i2})^2/n$ of the term in the reduced form that identifies δ_{0G} must grow linearly in K, just as in Chao and Swanson (2005), leading to a convergence rate of $1/\sqrt{K - G_1} = 1/\sqrt{r_n}$.

In another important case, the excluded instrument z_{iG} could be an unknown function that can be approximated by a linear combination of Z_i. For instance, suppose that $z_{iG} = f_0(w_i)$ for an unknown function $f_0(w_i)$ of variables w_i. In this case, the instrumental variables could include a vector $p^K(w_i) \stackrel{\text{def}}{=} (p_{1K}(w_i), \ldots, p_{K-G_1,K}(w_i))'$ of approximating functions, such as polynomials or splines. Here the vector of instrumental variables would be $Z_i = (Z_{i1}', p^K(w_i)')'$. For $r_n = n$, this example is like Newey (1990) where Z_i includes approximating functions for the reduced form but the number of instruments can grow as fast as the sample size. Alternatively, if $r_n/n \to 0$, it is a modified version where δ_G is more weakly identified.

Assumption 2 also allows for multiple endogenous variables with a different strength of identification for each one, i.e., for different convergence rates. In the preceding example, we maintained the scalar endogenous variable for simplicity.

The r_n can be thought of as a version of the concentration parameter; it determines the convergence rate of estimators of δ_{0G} just as the concentration parameter does in other settings. For $r_n = n$, the convergence rate will be $\sqrt{n}$ where Assumptions 1 and 2 permit K to grow as fast as the sample size. This corresponds to a many instrument asymptotic approximation like Kunitomo (1980), Morimune (1983), and Bekker (1994). For r_n growing more slowly than n, the convergence rate will be slower than $1/\sqrt{n}$, which leads to an asymptotic approximation like that of Chao and Swanson (2005).

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

In other words, Assumption 3 requires the second conditional moments of the disturbances to be bounded.

Assumption 4. There is a π_K such that $\sum_{i=1}^n \|z_i - \pi_K Z_i\|^2/n \to 0$ a.s.

This condition allows an unknown reduced form that is approximated by a linear combination of the instrumental variables. These four assumptions give the consistency result presented in Theorem 1.

THEOREM 1. *Suppose that Assumptions 1–4 are satisfied. Then,* $r_n^{-1/2} S_n'(\tilde{\delta} - \delta_0) \xrightarrow{p} 0$, $\tilde{\delta} \xrightarrow{p} \delta_0$, $r_n^{-1/2} S_n'(\hat{\delta} - \delta_0) \xrightarrow{p} 0$, *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 $\sum_{i=1}^n \|z_i\|^4/n^2 \to 0$, $\sup_i \mathrm{E}[\varepsilon_i^4|\mathcal{Z}] < C$, and $\sup_i \mathrm{E}[\|U_i\|^4|\mathcal{Z}] \leq C$ a.s.

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

$$\bar{H}_n = \sum_{i=1}^n z_i z_i'/n, \qquad \bar{\Omega}_n = \sum_{i=1}^n z_i z_i' \sigma_i^2/n,$$

$$\bar{\Psi}_n = S_n^{-1} \sum_{i \neq j} P_{ij}^2 \Big(\mathrm{E}[U_i U_i' | \mathcal{Z}] \sigma_j^2 (1 - P_{jj})^{-2} + \mathrm{E}[U_i \varepsilon_i | \mathcal{Z}] (1 - P_{ii})^{-1} \mathrm{E}[\varepsilon_j U_j' | \mathcal{Z}] (1 - P_{jj})^{-1} \Big) S_n^{-1\prime},$$

$$H_n = \sum_{i=1}^{n} (1 - P_{ii}) z_i z_i' / n, \qquad \Omega_n = \sum_{i=1}^{n} (1 - P_{ii})^2 z_i z_i' \sigma_i^2 / n,$$

$$\Psi_n = S_n^{-1} \sum_{i \neq j} P_{ij}^2 \Big(\mathrm{E}[U_i U_i' | \mathcal{Z}] \sigma_j^2 + \mathrm{E}[U_i \varepsilon_i | \mathcal{Z}] \mathrm{E}[\varepsilon_j U_j' | \mathcal{Z}] \Big) S_n^{-1\prime}.$$

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 semidefinite matrix A, satisfying $A^{1/2} A^{1/2\prime} = A$. Also, for nonsingular A, let $A^{-1/2} = (A^{1/2})^{-1}$.

THEOREM 2. *Suppose that Assumptions 1–5 are satisfied, $\sigma_i^2 \geq C > 0$ a.s., and K/r_n is bounded. Then $\bar{V}_n$ and V_n are nonsingular a.s.n, and*

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

The entire S_n matrix in Assumption 2 determines the convergence rate of the estimators, where

$$S_n'(\hat{\delta} - \delta_0) = \operatorname{diag}(\mu_{1n}, \ldots, \mu_{Gn}) \tilde{S}_n'(\hat{\delta} - \delta_0)$$

is asymptotically normal. The convergence rate of the linear combination $e_j' \tilde{S}_n'(\hat{\delta} - \delta_0)$ will be $1/\mu_{jn}$, where e_j is the jth unit vector. Note that

$$y_i = X_i' \delta_0 + u_i = z_i' \operatorname{diag}(\mu_{1n}, \ldots, \mu_{Gn}) \tilde{S}_n' \delta_0 + U_i' \delta_0 + \varepsilon_i.$$

The expression following the second equality is the reduced form for y_i. Thus, the linear combination of structural parameters $e_j' \tilde{S}_n' \delta_0$ is the jth reduced form coefficient for y_i that corresponds to the variable $\left(\mu_{jn}/\sqrt{n}\right) z_{ij}$. This reduced form coefficient is estimated at the rate $1/\mu_{jn}$ by the linear combination $e_j' \tilde{S}_n' \hat{\delta}$ of the instrumental variables (IV) estimator $\hat{\delta}$. The minimum rate is $1/\sqrt{r_n}$, which is the inverse square root of the rate of growth of the concentration parameter. These rates will change 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 more slowly 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 if $K/r_n \to 0$, dropping these terms does not affect the asymptotic variance. When K/r_n is bounded but does not go to zero, these terms have the same order as the other terms, and it is important to account for their presence in the standard errors. If $K/r_n \to \infty$, then these terms dominate and slow down the convergence rate of the estimators. In this case, 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 $K/r_n \to \infty$, 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. 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 $K/r_n \to \infty$. If L_n is bounded and there is a $C > 0$ such that $\lambda_{\min}\left(L_n\bar{V}_n^*L_n'\right) \geq C$ a.s.n then*

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

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

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

Here the convergence rate is related to the size of $\left(\sqrt{r_n/K}\right)S_n$. In the simple case where δ is a scalar, we can take $S_n = \sqrt{r_n}$, which gives a convergence rate of $\sqrt{K}/r_n$. Then the theorem states that $\left(r_n/\sqrt{K}\right)(\tilde{\delta}-\delta_0)$ is asymptotically normal. It is interesting that $\sqrt{K}/r_n \to 0$ is a condition for consistency in this setting and also in the context of Theorem 1.

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 relative to their number, 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. In this case, the available instruments are weaker than 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. Because $\sum_{i=1}^n P_{ii} = \text{tr}(P) = K$, P_{ii} goes to zero when K grows more slowly than n, though precise conditions for this convergence 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. In this case, the asymptotic variance of $\tilde{\delta}$ is

$$\bar{V}_n = \bar{V}_n^1 + \bar{V}_n^2, \qquad \bar{V}_n^1 = \sigma^2 \bar{H}_n^{-1},$$

$$\bar{V}_n^2 = S_n^{-1}\sigma^2 \text{E}[U_i U_i'] \sum_{i \neq j} P_{ij}^2/(1-P_{jj})^2 S_n^{-1}$$
$$+ S_n^{-1}\text{E}[U_i\varepsilon_i]\text{E}[U_i'\varepsilon_i]S_n^{-1\prime} \sum_{i \neq j} P_{ij}^2(1-P_{ii})^{-1}(1-P_{jj})^{-1}.$$

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

$$V_n = V_n^1 + V_n^2, \qquad V_n^1 = \sigma^2 H_n^{-1}\left[\sum_{i=1}^n (1-P_{ii})^2 z_i z_i'/n\right] H_n^{-1},$$

$$V_n^2 = S_n^{-1}\left(\sigma^2 \text{E}[U_i U_i'] + \text{E}[U_i\varepsilon_i]\text{E}[U_i'\varepsilon_i]\right) S_n^{-1\prime} \sum_{i \neq j} P_{ij}^2.$$

By the fact that $(1-P_{ii})^{-1} > 1$, we have that $\bar{V}_n^2 \geq V_n^2$ in the positive semidefinite sense. Also, note that V_n^1 is the variance of an IV estimator with instruments $z_i(1-P_{ii})$ whereas $\bar{V}_n^1$ is the variance of the corresponding least squares estimator, so $\bar{V}_n^1 \leq V_n^1$. 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. We impose the following additional conditions.

Assumption 6. *There exist π_n and $C > 0$ such that a.s.* $\max_{i \leq n} \|z_i - \pi_n Z_i\| \to 0$ *and* $\sup_i \|z_i\| \leq C$.

The next result shows that our estimators of the asymptotic variance 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 $K/r_n \to \infty$, 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 conducting approximate inference concerning coefficients. To that end, we introduce Theorem 5.

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 $\bar{\delta}_k \xrightarrow{p} \delta_0$*,* $(k = 1, \ldots, \ell)$ *and* $\bar{A} = [\partial a_1(\bar{\delta})/\partial\delta, \ldots, \partial a_\ell(\bar{\delta})/\partial\delta]'$*, we have* $B_n(\bar{A} - A)S_n^{-1\prime} \xrightarrow{p} 0$.

Also suppose that there is $C > 0$ *such that* $\lambda_{\min}(B_n A S_n^{-1\prime}\bar{V}_n S_n^{-1} A' B_n') \geq C$ *if* K/r_n *is bounded or* $\lambda_{\min}(B_n A S_n^{-1\prime}\bar{V}_n^* S_n^{-1} A' B_n') \geq C$ *if* $K/r_n \to \infty$ *a.s.n. 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 there is $C \geq 0$ *such that* $\lambda_{\min}(B_n A S_n^{-1\prime}\bar{V}_n S_n^{-1} A' B_n') \geq C$ *if* K/r_n *is bounded or* $\lambda_{\min}(B_n A S_n^{-1\prime}\bar{V}_n^* S_n^{-1} A' B_n') \geq C$ *if* $K/r_n \to \infty$ *a.s.n, 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 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 1. *Suppose that Assumptions 1–6 are satisfied and* c *and* b_n *are such that* $b_n c' S_n^{-1\prime}$ *is bounded. If there is a* $C > 0$ *such that* $b_n^2 c' S_n^{-1\prime}\bar{V}_n S_n^{-1} c \geq C$ *if* K/r_n *is bounded or* $b_n^2 c' S_n^{-1\prime}\bar{V}_n^* S_n^{-1} c \geq C$ *if* $K/r_n \to \infty$ *a.s.n, then*

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

Also if there is a $C \geq 0$ *such that* $b_n^2 c' S_n^{-1\prime} V_n S_n^{-1} c \geq C$ *if* K/r_n *is bounded or* $b_n^2 c' S_n^{-1\prime} V_n^* S_n^{-1} c \geq C$ *if* $K/r_n \to \infty$ *a.s.n, then*

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

To show how the conditions of this result can be checked, we return to the previous example with one right-hand-side endogenous variable. The following result gives primitive conditions in that example for the conclusion of Corollary 1, i.e., for the asymptotic normality of a t-ratio.

COROLLARY 2. *If equation (2) holds, Assumptions 1–6 are satisfied for* $z_i = (Z_{i1}', z_{iG})$*,* $c \neq 0$ *is a constant vector, either* *(i)* $r_n = n$ *or*

(ii) K/r_n *is bounded and* $(-\pi_1, 1)c \neq 0$ *or*

(iii) $K/r_n \to \infty$, $(-\pi_1, 1)c \neq 0$, $\mathrm{E}[U_{iG}^2|\mathcal{Z}]$ *is bounded away from zero, and the sign of* $\mathrm{E}[\varepsilon_i U_{iG}|\mathcal{Z}]$ *is constant a.s., then*

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

The proof of this result shows how the hypotheses concerning b_n in Corollary 1 can be checked. The conditions of Corollary 2 are quite primitive. We have previously described how Assumption 2 is satisfied in the model of equation (2). Assumptions 1 and 3–6 are also quite primitive.

This result can be applied to show that t-ratios are asymptotically correct when the many instrument robust variance estimators are used. For the coefficient δ_G of the endogenous variable, note that $c = e_G$, so $(-\pi_1, 1)c = 1 \neq 0$. Therefore, if $\mathrm{E}[U_{iG}^2|\mathcal{Z}]$ is bounded away from zero and the sign of $\mathrm{E}[\varepsilon_i U_{iG}|\mathcal{Z}]$ is constant, it follows from Corollary 2 that

$$\frac{\hat{\delta}_G - \delta_{0G}}{\sqrt{\hat{V}_{GG}}} \xrightarrow{d} N(0,1).$$

Thus, the t-ratio for the coefficient of the endogenous variable is asymptotically correct across a wide range of different growth rates for r_n and K. The analogous result holds for each coefficient δ_j, $j \leq G_1$, of an included instrument as long as $\pi_{1j} \neq 0$ is not zero. If $\pi_{1j} = 0$, then the asymptotics are more complicated. For brevity, we will not discuss this unusual case here. The analogous results also hold for $\tilde{\delta}_G$.

4. CONCLUDING REMARKS

In this paper, we derived limiting distribution results for two alternative JIV estimators. These estimators are both consistent and asymptotically normal in the presence of many instruments under heteroskedasticity of unknown form. In the same setup, LIML, 2SLS, and B2SLS are inconsistent. 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 central limit theorem has already been used in Hausman et al. (2007) to derive the asymptotic properties of the jackknife versions of the LIML and Fuller (1977) estimators and in Chao et al. (2010) to derive a moment-based test that allows for heteroskedasticity and many instruments. Moreover, this new central limit theorem is potentially useful for other analyses involving many instruments.

NOTE

1. The observations $w_1, \ldots, w_n$ are distinct with probability one and therefore, by $K < n$, cannot all be roots of a Kth 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 $a' Z' Z a > 0$.

REFERENCES

Abadir, K.M. & J.R. Magnus (2005) *Matrix Algebra*. Cambridge University Press.
Ackerberg, D.A. & P. Devereux (2009) Improved JIVE estimators for overidentified models with and without heteroskedasticity. *Review of Economics and Statistics* 91, 351–362.
Angrist, J.D., G.W. Imbens, & A. Krueger (1999) Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14, 57–67.
Bekker, P.A (1994) Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681.
Bekker, P.A. & J. van der Ploeg (2005) Instrumental variable estimation based on grouped data. *Statistica Neerlandica* 59, 506–508.
Billingsley, P. (1986) *Probability and Measure*, 2nd ed. Wiley.
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., J.A. Hausman, W.K. Newey, N.R. Swanson, & T. Woutersen (2010) Testing Overidentifying Restrictions with Many Instruments and Heteroskedasticity. Working paper, MIT.
Chao, J.C. & N.R. Swanson (2004) Estimation and Testing Using Jackknife IV in Heteroskedastic Regressions with Many Weak Instruments. Working paper, Rutgers University.
Chao, J.C. & N.R. Swanson (2005) Consistent estimation with a large number of weak instruments. *Econometrica* 73, 1673–1692.
Chao, J.C. & N.R. Swanson (2006) Asymptotic normality of single-equation estimators for the case with a large number of weak instruments. In D. Corbae, S.N. Durlauf, & B.E. Hansen (eds.), *Frontiers of Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, pp. 82–124. Cambridge University Press.
Davidson, R. & J.G. MacKinnon (2006) The case against JIVE. *Journal of Applied Econometrics* 21, 827–833.
Donald, S.G. & W.K. Newey (2001) Choosing the number of instruments. *Econometrica* 69, 1161–1191.
Fuller, W.A. (1977) Some properties of a modification of the limited information estimator. *Econometrica* 45, 939–954.
Han, C. & P.C.B. Phillips (2006) GMM with many moment conditions. *Econometrica* 74, 147–192.
Hansen, C., J.A. Hausman, & W.K. Newey (2008) Estimation with many instrumental variables. *Journal of Business & Economic Statistics* 26, 398–422.
Hausman, J.A., W.K. Newey, T. Woutersen, J. Chao, & N.R. Swanson (2007) IV Estimation with Heteroskedasticity and Many Instruments. Working paper, MIT.
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. & H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Wiley.
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. (1990) Efficient instrumental variable estimation of nonlinear models. *Econometrica* 58, 809–837.
Phillips, G.D.A. & 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 & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. 1, pp. 449–516. North-Holland.

Sawa, T. (1968) The exact sampling distribution of ordinary least squares and two-stage least squares estimators. *Journal of the American Statistical Association* 64, 923–937.

Staiger, D. & J.H. Stock (1997) Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stock, J.H. and M. Yogo (2005) Asymptotic distributions of instrumental variables statistics with many weak instruments. In D.W.K. Andrews & J.H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, pp. 109–120. Cambridge University Press.

APPENDIX A: Proofs of Theorems

We define a number of notations and abbreviations that will be used in Appendixes A and B. Let C denote a generic positive constant and let M, CS, and T denote the Markov inequality, the Cauchy–Schwarz inequality, and the triangle inequality, respectively. Also, for random variables W_i, Y_i, and η_i and for $\mathcal{Z} = (\Upsilon, Z)$, let $\bar{\omega}_i = \mathrm{E}[W_i|\mathcal{Z}]$, $\tilde{W}_i = W_i - \bar{\omega}_i$, $\bar{y}_i = \mathrm{E}[Y_i|\mathcal{Z}]$, $\tilde{Y}_i = Y_i - \bar{y}_i$, $\bar{\eta}_i = \mathrm{E}[\eta_i|\mathcal{Z}]$, $\tilde{\eta}_i = \eta_i - \bar{\eta}_i$, $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_n)'$, $\bar{\omega} = (\bar{\omega}_1, \ldots, \bar{\omega}_n)'$,

$$\bar{\mu}_W = \max_{1 \le i \le n} |\bar{\omega}_i|, \qquad \bar{\mu}_Y = \max_{1 \le i \le n} |\bar{y}_i|, \qquad \bar{\mu}_\eta = \max_{1 \le i \le n} |\bar{\eta}_i|,$$
$$\bar{\sigma}_W^2 = \max_{i \le n} \mathrm{Var}\left[W_i|\mathcal{Z}\right], \qquad \bar{\sigma}_Y^2 = \max_{i \le n} \mathrm{Var}\left[Y_i|\mathcal{Z}\right], \quad \text{and} \quad \bar{\sigma}_\eta^2 = \max_{i \le n} \mathrm{Var}\left[\eta_i|\mathcal{Z}\right],$$

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

We first give four lemmas that are useful in the proofs 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 $\mathcal{Z} = (\Upsilon, Z)$, $(W_i, Y_i)(i = 1, \ldots, n)$ are independent a.s., W_i and Y_i are scalars, and P is a symmetric, idempotent matrix of rank K, then for $\bar{\omega} = \mathrm{E}\left[(W_1, \ldots, W_n)'|\mathcal{Z}\right]$, $\bar{y} = \mathrm{E}\left[(Y_1, \ldots, Y_n)'|\mathcal{Z}\right]$, $\bar{\sigma}_{W_n} = \max_{i \le n} \mathrm{Var}(W_i|\mathcal{Z})^{1/2}$, $\bar{\sigma}_{Y_n} = \max_{i \le n} \mathrm{Var}(Y_i|\mathcal{Z})^{1/2}$, and $D_n = K\bar{\sigma}_{W_n}^2\bar{\sigma}_{Y_n}^2 + \bar{\sigma}_{W_n}^2\bar{y}'\bar{y} + \bar{\sigma}_{Y_n}^2\bar{\omega}'\bar{\omega}$, there exists a positive constant C such that*

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

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

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

Let $D_{1n} = \bar{\sigma}^2_{W_n} \bar{\sigma}^2_{Y_n}$. Note that for $i \neq j$ and $k \neq \ell$, $\mathrm{E}\left[\tilde{W}_i \tilde{Y}_j \tilde{W}_k \tilde{Y}_\ell | \mathcal{Z}\right]$ is zero unless $i = k$ and $j = \ell$ or $i = \ell$ and $j = k$. Then by CS and $\sum_j P_{ij}^2 = P_{ii}$,

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

Also, for $\tilde{W} = (\tilde{W}_1, \ldots, \tilde{W}_n)'$, 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 $\mathrm{E}\left[\tilde{W}\tilde{W}'|\mathcal{Z}\right] \leq \bar{\sigma}^2_{W_n} I_n$ a.s., so

$$\mathrm{E}[(\bar{y}'P\tilde{W})^2|\mathcal{Z}] = \bar{y}'PE[\tilde{W}\tilde{W}'|\mathcal{Z}]P\bar{y} \leq \bar{\sigma}^2_{W_n}\bar{y}'P\bar{y} \leq \bar{\sigma}^2_{W_n}\bar{y}'\bar{y},$$

$$\mathrm{E}\left[\left(\sum\nolimits_i P_{ii}\bar{y}_i \tilde{W}_i\right)^2 |\mathcal{Z}\right] = \sum_i P_{ii}^2 \mathrm{E}[\tilde{W}_i^2|\mathcal{Z}]\bar{y}_i^2 \leq \bar{\sigma}^2_{W_n}\bar{y}'\bar{y}.$$

Then by T we have

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

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

LEMMA A2. *Suppose that, conditional on $\mathcal{Z}$, the following conditions hold a.s.:*

(i) $P = P(\mathcal{Z})$ *is a symmetric, idempotent matrix with* $\mathrm{rank}(P) = K$ *and* $P_{ii} \leq C < 1$;
(ii) $(W_{1n}, U_1, \varepsilon_1), \ldots, (W_{nn}, U_n, \varepsilon_n)$ *are independent, and* $D_n = \sum_{i=1}^n \mathrm{E}\left[W_{in}W_{in}'|\mathcal{Z}\right]$ *satisfies* $\|D_n\| \leq C$ *a.s.n;*
(iii) $\mathrm{E}\left[W_{in}'|\mathcal{Z}\right] = 0$, $\mathrm{E}[U_i|\mathcal{Z}] = 0$, $\mathrm{E}[\varepsilon_i|\mathcal{Z}] = 0$, *and there exists a constant C such that* $\mathrm{E}[\|U_i\|^4|\mathcal{Z}] \leq C$ *and* $\mathrm{E}[\varepsilon_i^4|\mathcal{Z}] \leq C$;
(iv) $\sum_{i=1}^n \mathrm{E}\left[\|W_{in}\|^4|\mathcal{Z}\right] \xrightarrow{a.s.} 0$; *and*
(v) $K \to \infty$ *as* $n \to \infty$.

Then for

$$\bar{\Sigma}_n \stackrel{def}{=} \sum_{i\neq j} P_{ij}^2 \left(\mathrm{E}[U_i U_i'|\mathcal{Z}]\mathrm{E}[\varepsilon_j^2|\mathcal{Z}] + \mathrm{E}[U_i\varepsilon_i|\mathcal{Z}]\mathrm{E}[\varepsilon_j U_j'|\mathcal{Z}]\right)/K$$

and any sequences c_{1n} and c_{2n} depending on $\mathcal{Z}$ of conformable vectors with $\|c_{1n}\| \leq C$, $\|c_{2n}\| \leq C$, and $\Xi_n = c_{1n}'D_nc_{1n} + c_{2n}'\bar{\Sigma}_nc_{2n} > 1/C$ a.s.n, it follows that

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

i.e., $\mathrm{Pr}(Y_n \leq y|\mathcal{Z}) \xrightarrow{a.s.} \Phi(y)$ *for all y.*

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 $\mathcal{Z}$, $(W_i, Y_i)(i = 1, \ldots, n)$ are independent and W_i and Y_i are scalars, then there exists a positive constant C such that*

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

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

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

As before, for $i \neq j$ and $k \neq \ell$, $\mathrm{E}\left[\tilde{W}_i \tilde{Y}_j \tilde{W}_k \tilde{Y}_\ell | \mathcal{Z}\right]$ is zero unless $i = k$ and $j = \ell$ or $i = \ell$ and $j = k$. Also, $|P_{ij}| \leq P_{ii} < 1$ by CS and Assumption 1, so $P_{ij}^4 \leq P_{ij}^2$. Also, $\sum_j P_{ij}^2 = P_{ii}$, so

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

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 $\mathcal{Z}$, we have $\mathrm{E}[\tilde{W} \tilde{W}' | \mathcal{Z}] \leq \bar{\sigma}_{Wn}^2 I_n$, so

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

$$\mathrm{E}\left[\left(\sum_i P_{ii}^2 \bar{y}_i \tilde{W}_i\right)^2 | \mathcal{Z}\right] = \sum_i P_{ii}^4 \mathrm{E}[\tilde{W}_i^2 | \mathcal{Z}] \bar{y}_i^2 \leq K \bar{\sigma}_W^2 \bar{\mu}_Y^2 \quad \text{a.s.}$$

Then by T, we have $\left\| \sum_{i \neq j} P_{ij}^2 \tilde{W}_i \bar{y}_j \right\|_{L_2, \mathcal{Z}}^2 \leq \left\| \tilde{W}' \tilde{P} \bar{y} \right\|_{L_2, \mathcal{Z}}^2 + \left\| \sum_i P_{ii}^2 \bar{y}_i \tilde{W}_i \right\|_{L_2, \mathcal{Z}}^2 \leq C K \bar{\sigma}_W^2 \bar{\mu}_Y^2$ a.s. Interchanging the roles of Y_i and W_i gives $\left\| \sum_{i \neq j} P_{ij}^2 \bar{\omega}_i \tilde{Y}_j \right\|_{L_2, \mathcal{Z}}^2 \leq C K \bar{\mu}_W^2 \bar{\sigma}_Y^2$ a.s. The conclusion then follows by T. ■

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

LEMMA A4. *Suppose that there is* $C>0$ *such that, conditional on* $\mathcal{Z}$, $(W_1,Y_1,\eta_1),\ldots,(W_n,Y_n,\eta_n)$ *are independent with* $\mathrm{E}[W_i|\mathcal{Z}]=a_i/\sqrt{n}$, $\mathrm{E}[Y_i|\mathcal{Z}]=b_i/\sqrt{n}$, $|a_i|\le C$, $|b_i|\le C$, $\mathrm{E}[\eta_i^2|\mathcal{Z}]\le C$, $\mathrm{Var}(W_i|\mathcal{Z})\le C/r_n$, *and* $\mathrm{Var}(Y_i|\mathcal{Z})\le C/r_n$ *and there exists* π_n *such that* $\max_{i\le n}\left|a_i-Z_i'\pi_n\right|\overset{a.s.}{\to}0$ *and* $\sqrt{K}/r_n\to 0$. *Then*

$$A_n=\mathrm{E}\left[\sum_{i\neq j\neq k}W_iP_{ik}\eta_kP_{kj}Y_j|\mathcal{Z}\right]=O_p(1),\qquad \sum_{i\neq j\neq k}W_iP_{ik}\eta_kP_{kj}Y_j-A_n\overset{p}{\to}0.$$

Proof. Given in Appendix B. ■

LEMMA A5. *If Assumptions 1–3 are satisfied, then*

(i) $S_n^{-1}\tilde{H}S_n^{-1\prime}=\sum_{i\neq j}z_iP_{ij}(1-P_{jj})^{-1}z_j'/n+o_p(1)$,

(ii) $S_n^{-1}\sum_{i\neq j}X_iP_{ij}(1-P_{jj})^{-1}\varepsilon_j=O_p(1+\sqrt{K/r_n})$,

(iii) $S_n^{-1}\hat{H}S_n^{-1\prime}=\sum_{i\neq j}z_iP_{ij}z_j'/n+o_p(1)$,

(iv) $S_n^{-1}\sum_{i\neq j}X_iP_{ij}\varepsilon_j=O_p(1+\sqrt{K/r_n})$.

Proof. Let e_k denote the kth unit vector and 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. By Assumption 2, $\lambda_{\min}(S_n)\ge C\sqrt{r_n}$, implying $\left\|S_n^{-1}\right\|\le C/\sqrt{r_n}$. Therefore a.s.

$$\mathrm{E}[Y_i|\mathcal{Z}]=z_{ik}/\sqrt{n},\qquad \mathrm{Var}(Y_i|\mathcal{Z})\le C/r_n,$$
$$\mathrm{E}[W_i|\mathcal{Z}]=z_{i\ell}/\sqrt{n}(1-P_{ii}),\qquad \mathrm{Var}(W_i|\mathcal{Z})\le C/r_n.$$

Note that a.s.

$$\sqrt{K}\bar{\sigma}_{W_n}\bar{\sigma}_{Y_n}\le C\sqrt{K}/r_n\to 0,\quad \bar{\sigma}_{W_n}\sqrt{\bar{y}'\bar{y}}\le Cr_n^{-1/2}\sqrt{\sum_i z_{ik}^2/n}\to 0,$$
$$\bar{\sigma}_{Y_n}\sqrt{\bar{w}'\bar{w}}\le Cr_n^{-1/2}\sqrt{\sum_i z_{i\ell}^2(1-P_{ii})^{-2}/n}\le Cr_n^{-1/2}(1-\max_i P_{ii})^{-2}\sqrt{\sum_i z_{i\ell}^2/n}\to 0.$$

Because $e_k'S_n^{-1}\tilde{H}S_n^{-1\prime}e_\ell=e_k'S_n^{-1}\sum_{i\neq j}X_iP_{ij}X_j'S_n^{-1\prime}e_\ell/(1-P_{jj})=\sum_{i\neq j}Y_iP_{ij}W_j$ 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 $\upsilon>0$ and $A_n=\left\{\left|e_k'S_n^{-1}\tilde{H}S_n^{-1\prime}e_\ell-\sum_{i\neq j}e_k'z_iP_{ij}(1-P_{jj})^{-1}z_j'e_\ell/n\right|\ge\upsilon\right\}$, $P(A_n|\mathcal{Z})\overset{\text{a.s.}}{\to}0$. By the dominated convergence theorem, $P(A_n)=\mathrm{E}[P(A_n|\mathcal{Z})]\to 0$. The preceding argument establishes the first conclusion for the (k,ℓ)th element. Doing this for every element completes the proof of the first conclusion.

For the second conclusion, apply Lemma A1 with $Y_i=e_k'S_n^{-1}X_i$ as before and $W_i=\varepsilon_i/(1-P_{ii})$. Note that $\bar{w}_i=0$ and $\bar{\sigma}_{W_n}\le C$. Then by Lemma A1,

$$\mathrm{E}[\{e_k'S_n^{-1}\sum_{i\neq j}X_iP_{ij}(1-P_{jj})^{-1}\varepsilon_j\}^2|\mathcal{Z}]\le CK/r_n+C.$$

The conclusion then follows from the fact that $\mathrm{E}[A_n|\mathcal{Z}]\le C$ implies $A_n=O_p(1)$.

For the third conclusion, apply Lemma A1 with $Y_i = e_k' S_n^{-1} X_i$ as before and $W_i = e_\ell' S_n^{-1} X_i$, so a.s.

$$\sqrt{K}\bar{\sigma}_{W_n}\bar{\sigma}_{Y_n} \le C\sqrt{K}/r_n \to 0, \quad \bar{\sigma}_{W_n}\sqrt{\bar{y}'\bar{y}} \le Cr_n^{-1/2}\sqrt{\sum z_{ik}^2/n} \to 0, \quad \bar{\sigma}_{Y_n}\sqrt{\bar{w}'\bar{w}} \to 0.$$

The fourth conclusion follows similarly to the second conclusion. ■

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*

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

Proof. We use Lemma A5 and approximate the right-hand-side terms in Lemma A5 by $\bar{H}_n$ and H_n. Let $\bar{z}_i = \sum_{j=1}^n P_{ij} z_j$ be the ith element of Pz and note that

$$\sum_{i=1}^n \|z_i - \bar{z}_i\|^2/n = \|(I-P)z\|^2/n = \operatorname{tr}(z'(I-P)z/n) = \operatorname{tr}[(z - Z\pi_{Kn}')'(I-P)(z - Z\pi_{Kn}')/n]$$
$$\le \operatorname{tr}[(z - Z\pi_{Kn}')'(z - Z\pi_{Kn}')/n] = \sum_{i=1}^n \|z_i - \pi_{Kn} Z_i\|^2/n \to 0 \quad \text{a.s. } \mathbb{P}_{\mathcal{Z}}.$$

It follows that a.s.

$$\left\|\sum_i (\bar{z}_i - z_i)(1-P_{ii})^{-1} z_i'/n\right\| \le \sum_i \|\bar{z}_i - z_i\| \left\|(1-P_{ii})^{-1} z_i'\right\|/n$$
$$\le \sqrt{\sum_i \|\bar{z}_i - z_i\|^2/n}\sqrt{\sum_i \left\|(1-P_{ii})^{-1} z_i\right\|^2/n} \to 0.$$

Then

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

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

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

so the second conclusion follows similarly to the first. ■

Proof of Theorem 1. First, note that by $\lambda_{\min}\left(S_n S_n'/r_n\right) \ge \lambda_{\min}\left(\tilde{S}\tilde{S}'\right) \ge C$, we have

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

Therefore, $S_n'(\tilde{\delta}-\delta_0)/\sqrt{r_n} \xrightarrow{p} 0$ implies $\tilde{\delta} \xrightarrow{p} \delta_0$. Note that by Assumption 2, $\bar{H}_n$ is bounded and $\lambda_{\min}(\bar{H}_n) \geq C$ a.s.n. 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 equation (1) and Lemma A5,

$$r_n^{-1/2}S_n'(\tilde{\delta}-\delta_0) = (S_n^{-1}\tilde{H}S_n^{-1\prime})^{-1}S_n^{-1}\sum_{i\neq j}X_iP_{ij}\xi_j/\sqrt{r_n} = O_p(1)o_p(1) \xrightarrow{p} 0.$$

All of the previous statements are conditional on $\mathcal{Z} = (\Upsilon, Z)$ for a given sample size n, so for the random variable $R_n = r_n^{-1/2}S_n'(\tilde{\delta}-\delta_0)$, we have shown that for any constant $\upsilon > 0$, a.s. $\Pr(\|R_n\| \geq \upsilon|\mathcal{Z}) \to 0$. Then by the dominated convergence theorem, $\Pr(\|R_n\| \geq \upsilon) = \mathrm{E}[\Pr(\|R_n\| \geq \upsilon|\mathcal{Z})] \to 0$. Therefore, because υ is arbitrary, it follows that $R_n = r_n^{-1/2}S_n'(\tilde{\delta}-\delta_0) \xrightarrow{p} 0$.

Next note that $P_{ii} \leq C < 1$, so in the positive semidefinite sense in large enough samples a.s.,

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

Thus, by Assumption 2, H_n is bounded and bounded away from singularity a.s.n. Then the rest of the conclusion follows analogously with $\hat{\delta}$ replacing $\tilde{\delta}$ and H_n replacing $\bar{H}_n$. ■

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. Define

$$Y_n = \sum_i z_i(1-P_{ii})\xi_i\Big/\sqrt{n} + S_n^{-1}\sum_{i\neq j}U_iP_{ij}\xi_j.$$

By Assumptions 2–4,

$$\mathrm{E}\left[\left\|\sum_{i=1}^n(z_i-\bar{z}_i)\xi_i/\sqrt{n}\right\|^2|\mathcal{Z}\right]$$
$$= \sum_{i=1}^n\|z_i-\bar{z}_i\|^2\mathrm{E}\left[\xi_i^2|\mathcal{Z}\right]\Big/n \leq C\sum_{i=1}^n\|z_i-\bar{z}_i\|^2/n \xrightarrow{a.s.} 0.$$

Therefore, by M,

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

We now apply Lemma A2 to establish asymptotic normality of Y_n conditional on $\mathcal{Z}$. Let $\Gamma_n = \mathrm{Var}(Y_n|\mathcal{Z})$, so

$$\Gamma_n = \sum_{i=1}^n z_iz_i'(1-P_{ii})^2\mathrm{E}[\xi_i^2|\mathcal{Z}]/n + S_n^{-1}\sum_{i\neq j}P_{ij}^2$$
$$\times\left(\mathrm{E}[U_iU_i'|\mathcal{Z}]\mathrm{E}[\xi_j^2|\mathcal{Z}] + \mathrm{E}[U_i\xi_i|\mathcal{Z}]\mathrm{E}[U_j'\xi_j|\mathcal{Z}]\right)S_n^{-1\prime}.$$

Note that $\sqrt{r_n}S_n^{-1}$ is bounded by Assumption 2 and that $\sum_{i\neq j} P_{ij}^2/K \leq 1$, so by boundedness of K/r_n and Assumption 3, it follows that $\|\Gamma_n\| \leq C$ a.s.n. Also, $\mathrm{E}[\xi_i^2|\mathcal{Z}] \geq C > 0$, so

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

Therefore, by Assumption 2, $\lambda_{\min}(\Gamma_n) \geq C > 0$ a.s.n (for generic C that may be different from before). It follows that $\left\|\Gamma_n^{-1}\right\| \leq C$ a.s.n.

Let α be a $G \times 1$ nonzero vector. Let U_i be defined as in Lemma A2 and ξ_i be defined as ε_i in Lemma A2. In addition, let $W_{in} = z_i(1-P_{ii})\xi_i/\sqrt{n}$, $c_{1n} = \Gamma_n^{-1/2}\alpha$, 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 $\mathrm{E}[\xi_i^2|\mathcal{Z}]$ a.s.n, condition (ii) of Lemma A2 is satisfied; condition (iii) is satisfied by Assumptions 3 and 5. Also, by $(1-P_{ii})^{-1} \leq C$ and Assumption 5, $\sum_{i=1}^n \mathrm{E}\left[\|W_{in}\|^4|\mathcal{Z}\right] \leq C\sum_{i=1}^n \|z_i\|^4/n^2 \xrightarrow{a.s.} 0$, so condition (iv) is satisfied. Finally, condition (v) is satisfied by hypothesis. Note also that $c_{1n} = \Gamma_n^{-1/2}\alpha$ and $c_{2n} = \left(\sqrt{K/r_n}\right)\sqrt{r_n}S_n^{-1}\Gamma_n^{-1/2}\alpha$ satisfy $\|c_{1n}\| \leq C$ and $\|c_{2n}\| \leq C$ a.s.n. This follows from the boundedness of $\sqrt{K/r_n}$, $\sqrt{r_n}S_n^{-1}$, and Γ_n^{-1}. Moreover, the Ξ_n of Lemma A2 is

$$\Xi_n = \mathrm{Var}(c_{1n}'\sum_{i=1}^{n} W_{in} + c_{2n}'\sum_{i\neq j} U_i P_{ij}\xi_j/\sqrt{K}|\mathcal{Z}) = \mathrm{Var}(\alpha'\Gamma_n^{-1/2\prime}Y_n|\mathcal{Z}) = \alpha'\alpha$$

by construction. Then, applying Lemma A2, we have

$$\left(\alpha'\alpha\right)^{-1/2}\alpha'\Gamma_n^{-1/2}Y_n = \Xi_n^{-1/2}\left(\sum_{i=1}^{n} c_{1n}'W_{in} + c_{2n}'\sum_{i\neq j} U_i P_{ij}\xi_j/\sqrt{K}\right) \xrightarrow{d} N(0,1) \quad \text{a.s.}$$

It follows that $\alpha'\Gamma_n^{-1/2}Y_n \xrightarrow{d} N\left(0,\alpha'\alpha\right)$ a.s., so by the Cramér–Wold device, $\Gamma_n^{-1/2}$ $Y_n \xrightarrow{d} N(0, I_G)$ a.s.

Consider now the JIV1 estimator where $\xi_i = \varepsilon_i/(1-P_{ii})$. Plugging this into the expression for Γ_n, we find $\Gamma_n = \bar{\Omega}_n + \bar{\Psi}_n$ for $\bar{\Omega}_n$ and $\bar{\Psi}_n$ defined according to Assumption 5. Let $\bar{V}_n$ also be as defined following Assumption 5 and note that $B_n = \bar{V}_n^{-1/2}\bar{H}_n^{-1}\Gamma_n^{1/2}$ is an orthogonal matrix because $B_n B_n' = \bar{V}_n^{-1/2}\bar{V}_n\bar{V}_n^{-1/2\prime} = I$. Also, B_n is a function of only $\mathcal{Z}$, $\left\|\bar{V}_n^{-1/2}\right\| \leq C$ a.s.n because $\lambda_{\min}(\bar{V}_n) \geq C > 0$ a.s.n, and $\left\|\Gamma_n^{1/2}\right\| \leq C$ a.s.n. By Lemma A6, $(S_n^{-1}\tilde{H}S_n^{-1\prime})^{-1} = \bar{H}_n^{-1} + o_p(1)$. Note that if a random variable W_n satisfies $\|W_n\| \leq C$ a.s.n, then $W_n = O_p(1)$ (note that $1(\|W_n\| > C) \xrightarrow{a.s.} 0$ implies that $\mathrm{E}[1(\|W_n\| > C)] = \Pr(\|W_n\| > C) \to 0$). Therefore, we have

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

Note that because $\Gamma_n^{-1/2}Y_n \xrightarrow{d} N(0, I_G)$ a.s. and B_n is orthogonal to and a function only of $\mathcal{Z}$, we have $B_n\Gamma_n^{-1/2}Y_n \xrightarrow{d} N(0, I_G)$. Then by the Slutsky lemma and $\tilde{\delta} = \delta_0 +$

$\tilde{H}^{-1}\sum_{i\neq j} X_i P_{ij}\xi_j$, for $\xi_j = (1-P_{jj})^{-1}\varepsilon_j$, we have

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

which gives the first conclusion. The conclusion for JIV2 follows by a similar argument for $\xi_i = \varepsilon_i$. ■

Proof of Theorem 3. Under the hypotheses of Theorem 3, $r_n/K \to 0$, so following the proof of Theorem 2, we have $\sqrt{r_n/K}\sum_{i=1}^n z_i(1-P_{ii})\xi_i/\sqrt{n} \xrightarrow{p} 0$. Then similar to the proof of Theorem 2, for $Y_n = \sqrt{r_n}S_n^{-1}\sum_{i\neq j} U_i P_{ij}\xi_j/\sqrt{K}$, we have $\sqrt{r_n/K}S_n^{-1}\sum_{i\neq j} X_i P_{ij}\xi_j = Y_n + o_p(1)$. Here let

$$\Gamma_n = \mathrm{Var}(Y_n|\mathcal{Z}) = r_n S_n^{-1}\sum_{i\neq j} P_{ij}^2\left(\mathrm{E}[U_iU_i'|\mathcal{Z}]\mathrm{E}[\xi_j^2|\mathcal{Z}] + \mathrm{E}[U_i\xi_i|\mathcal{Z}]\mathrm{E}[U_j'\xi_j|\mathcal{Z}]\right)S_n^{-1\prime}/K.$$

Note that by Assumptions 2 and 3, $\|\Gamma_n\| \leq C$ a.s.n. Let $\bar{L}_n$ be any sequence of bounded matrices with $\lambda_{\min}(\bar{L}_n\Gamma_n\bar{L}_n') \geq C > 0$ a.s.n and let $\bar{Y}_n = \left(\bar{L}_n\Gamma_n\bar{L}_n'\right)^{-1/2}\bar{L}_nY_n$. Now let α be a nonzero vector and apply Lemma A2 with $W_{in} = 0$, $\varepsilon_i = \xi_i$, $c_{1n} = 0$, and $c_{2n} = \alpha'\left(\bar{L}_n\Gamma_n\bar{L}_n'\right)^{-1/2}\bar{L}_n\sqrt{r_n}S_n^{-1}$. We have $\mathrm{Var}\left(c_{2n}'\sum_{i\neq j} U_i P_{ij}\xi_j/\sqrt{K}|\mathcal{Z}\right) = \alpha'\alpha > 0$ by construction, and the other hypotheses of Lemma A2 can be verified as in the proof of Theorem 2. Then by the conclusion of Lemma A2, it follows that $\alpha'\bar{Y}_n \xrightarrow{d} N(0, \alpha'\alpha)$ a.s. By the Cramér–Wold device, a.s. $\bar{Y}_n \xrightarrow{d} N(0, I)$.

Consider now the JIV1 estimator and let L_n be specified as in the statement of the result such that $\lambda_{\min}\left(L_n\bar{V}_n^*L_n'\right) \geq C > 0$ a.s.n. Let $\bar{L}_n = L_n\bar{H}_n^{-1}$, so $L_n\bar{V}_n^*L_n' = \bar{L}_n\Gamma_n\bar{L}_n'$. Note that $\left\|\left(\bar{L}_n\Gamma_n\bar{L}_n'\right)^{-1/2}\right\| \leq C$ and $\left\|\Gamma_n^{1/2}\right\| \leq C$ a.s.n. By Lemma A6, $(S_n^{-1}\tilde{H}S_n^{-1\prime})^{-1} = \bar{H}_n^{-1} + o_p(1)$. Therefore, we have

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

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

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

The conclusion for JIV2 follows by a similar argument for $\xi_i = \varepsilon_i$. ■

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

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

$$\dot{\Sigma}_1 = \sum_{i\neq j\neq k} \dot{X}_i P_{ik}\xi_k^2 P_{kj}\dot{X}_j', \quad \dot{\Sigma}_2 = \sum_{i\neq j} P_{ij}^2\left(\dot{X}_i\dot{X}_i'\xi_j^2 + \dot{X}_i\xi_i\xi_j\dot{X}_j'\right).$$

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 use Lemma A4. Note that for $\dot{\delta} = \hat{\delta}$ and $X_i^P = X_i/(1-P_{ii})$ for JIV1 and $\dot{\delta} = \tilde{\delta}$ and $X_i^P = X_i$ for JIV2, we have $\dot{\delta} \xrightarrow{p} \delta_0$ 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 of $-2\xi_i X_i^{P\prime}$ or $X_i^P X_i^{P\prime}$. Note that $S_n/\sqrt{n}$ is bounded, so by CS, $\|\Upsilon_i\| = \left\|S_n z_i/\sqrt{n}\right\| \le C$. Then

$$\mathrm{E}[\eta_i^2|\mathcal{Z}] \le CE[\xi_i^2|\mathcal{Z}] + CE[\|X_i\|^2|\mathcal{Z}] \le C + C\|\Upsilon_i\|^2 + CE[\|U_i\|^2|\mathcal{Z}] \le C.$$

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

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

From the preceding expression for $\hat{\xi}_i^2 - \xi_i^2$, we see that $\hat{\Sigma}_1 - \dot{\Sigma}_1$ is a sum of terms of the form

$\hat{\Delta}\sum_{i\neq j\neq k} \dot{X}_i P_{ik}\eta_k P_{kj}\dot{X}_j'$, so T, $\hat{\Sigma}_1 - \dot{\Sigma}_1 \xrightarrow{p} 0$.

Let $d_i = C + |\varepsilon_i| + \|U_i\|$, $\hat{A} = (1+\left\|\hat{\delta}\right\|)$ 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, because P_{ii} is bounded away from 1, $(1-P_{ii})^{-1} \le C$ a.s. Hence, for both JIV1 and JIV2,

$$\|X_i\| \le C + \|U_i\| \le d_i, \quad \left\|\dot{X}_i\right\| \le Cr_n^{-1/2}d_i, \quad \left|\hat{\xi}_i - \xi_i\right| \le C\left|X_i'(\hat{\delta}-\delta_0)\right| \le Cd_i\hat{B},$$

$$\left|\hat{\xi}_i\right| \le C\left|X_i'(\delta_0 - \hat{\delta})\right| + |\xi_i| \le Cd_i\hat{A},$$

$$\left|\hat{\xi}_i^2 - \xi_i^2\right| \le \left(|\xi_i| + \left|\hat{\xi}_i\right|\right)\left|\hat{\xi}_i - \xi_i\right| \le Cd_i(1+\hat{A})d_i\hat{B} \le Cd_i^2\hat{A}\hat{B},$$

$$\left\|\dot{X}_i\left(\hat{\xi}_i - \xi_i\right)\right\| \le C\mu_n^{-1}d_i^2\hat{B}, \quad \left\|\dot{X}_i\hat{\xi}_i\right\| \le Cr_n^{-1/2}d_i^2\hat{A}, \quad \left\|\dot{X}_i\xi_i\right\| \le Cr_n^{-1/2}d_i^2.$$

Also note that because $\mathrm{E}[d_i^2|\mathcal{Z}] \le C$,

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

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

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

We also have

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

The second conclusion then follows from T. ∎

LEMMA A8. *If Assumptions 1–6 are satisfied, then*

$$\dot{\Sigma}_1 = \sum_{i\neq j\neq k} z_i P_{ik} \mathrm{E}[\xi_k^2|\mathcal{Z}] P_{kj} z_j'/n + o_p(1),$$
$$\dot{\Sigma}_2 = \sum_{i\neq j} P_{ij}^2 z_i z_i' \mathrm{E}[\xi_j^2|\mathcal{Z}]/n + S_n^{-1} \sum_{i\neq j} P_{ij}^2 \big(\mathrm{E}[U_i U_i'|\mathcal{Z}]\mathrm{E}[\xi_j^2|\mathcal{Z}] + \mathrm{E}[U_i \xi_i|\mathcal{Z}]\mathrm{E}[\xi_j U_j'|\mathcal{Z}]\big) S_n^{-1\prime} + o_p(K/r_n).$$

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, we use Lemma A3. Note that $\mathrm{Var}(\xi_i^2|\mathcal{Z}) \le C$ and $r_n \le Cn$, so for $u_{ki} = e_k' S_n^{-1} U_i$,

$$\mathrm{E}[(\dot{X}_{ik}\dot{X}_{i\ell})^2|\mathcal{Z}] \le C E[\dot{X}_{ik}^4 + \dot{X}_{i\ell}^4|\mathcal{Z}]$$
$$\le C\left\{z_{ik}^4/n^2 + \mathrm{E}\left[u_{ki}^4|\mathcal{Z}\right] + z_{i\ell}^4/n^2 + \mathrm{E}\left[u_{\ell i}^4|\mathcal{Z}\right]\right\} \le C/r_n^2,$$
$$\mathrm{E}[(\dot{X}_{ik}\xi_i)^2|\mathcal{Z}] \le C E\left[\left(z_{ik}^2\xi_i^2/n + u_{ki}^2\xi_i^2\right)|\mathcal{Z}\right] \le C/n + C/r_n \le C/r_n.$$

Also, if $\Omega_i = \mathrm{E}[U_i U_i'|\mathcal{Z}]$, then $\mathrm{E}[\dot{X}_i \dot{X}_i'|\mathcal{Z}] = z_i z_i'/n + S_n^{-1}\Omega_i S_n^{-1\prime}$ and $\mathrm{E}[\dot{X}_i \xi_i|\mathcal{Z}] = S_n^{-1}\mathrm{E}[U_i \xi_i|\mathcal{Z}]$. Next let W_i be $\dot{X}_{ik}\dot{X}_{i\ell}$ for some k and ℓ, so

$$\mathrm{E}[W_i|\mathcal{Z}] = e_k' S_n^{-1}\Omega_i S_n^{-1\prime} e_\ell + z_{ik} z_{i\ell}/n, \qquad |\mathrm{E}[W_i|\mathcal{Z}]| \le C/r_n,$$
$$\mathrm{Var}(W_i|\mathcal{Z}) \le \mathrm{E}[(\dot{X}_{ik}\dot{X}_{i\ell})^2|\mathcal{Z}] \le C/r_n^2.$$

Also let $Y_i = \xi_i^2$ and note that $|\mathrm{E}[Y_i|\mathcal{Z}]| \le C$ and $\mathrm{Var}(W_i|\mathcal{Z}) \le C$. Then in the notation of Lemma A3,

$$\sqrt{K}(\bar{\sigma}_{W_n}\bar{\sigma}_{Y_n} + \bar{\sigma}_{W_n}\bar{\mu}_{Y_n} + \bar{\mu}_{W_n}\bar{\sigma}_{Y_n}) \le \sqrt{K}(C/r_n + C/r_n + C/r_n) \le C\sqrt{K}/r_n.$$

By the conclusion of Lemma A3, for this W_i and Y_i we have

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

Consider also Lemma A3 with W_i and Y_i equal to $\dot{X}_{ik}\xi_i$ and $\dot{X}_{i\ell}\xi_i$, respectively, so $\bar{\sigma}_{W_n}\bar{\sigma}_{Y_n} + \bar{\sigma}_{W_n}\bar{\mu}_{Y_n} + \bar{\mu}_{W_n}\bar{\sigma}_{Y_n} \leq C/r_n$. Then, applying Lemma A3, we have

$$\sum_{i\neq j} P_{ij}^2 \dot{X}_{ik}\xi_i\xi_j\dot{X}_{j\ell} = e_k' S_n^{-1} \sum_{i\neq j} P_{ij}^2 \mathrm{E}[U_i\xi_i|\mathcal{Z}]\mathrm{E}[\xi_j U_j'|\mathcal{Z}] S_n^{-1\prime} e_\ell + O_p(\sqrt{K}/r_n).$$

Also, because $K \to \infty$, we have $O_p(\sqrt{K}/r_n) = o_p(K/r_n)$. The second conclusion then follows by T. ■

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

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

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

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

Adding this equation to the previous one gives

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

which yields the equality in Section 2.

Let $\dot{\sigma}_i^2 = \mathrm{E}\left[\xi_i^2|\mathcal{Z}\right]$ and $\bar{z}_i = \sum_j P_{ij}z_j = e_i'Pz$. Then following the same line of argument as at the beginning of this proof, with z_i replacing X_i and $\dot{\sigma}_k^2$ replacing $\hat{\xi}_k^2$,

$$\sum_{i\neq j\neq k} z_iP_{ik}\dot{\sigma}_k^2P_{kj}z_j'/n = \sum_i \dot{\sigma}_i^2\left(\bar{z}_i\bar{z}_i' - P_{ii}z_i\bar{z}_i' - P_{ii}\bar{z}_iz_i' + P_{ii}^2z_iz_i'\right)\Big/n - \sum_{i\neq j} P_{ij}^2z_iz_i'\dot{\sigma}_j^2/n.$$

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

$$\left\|\sum_i \dot{\sigma}_i^2 (\bar{z}_i\bar{z}_i' - z_i z_i')/n\right\| \le \sum_i \dot{\sigma}_i^2 (2\|z_i\|\|z_i-\bar{z}_i\| + \|z_i - \bar{z}_i\|^2)/n$$
$$\le C\left(\sum_i \|z_i\|^2/n\right)^{1/2}\left(\sum_i \|z_i-\bar{z}_i\|^2/n\right)^{1/2} + C\sum_i \|z_i - \bar{z}_i\|^2/n \to 0,$$
$$\left\|\sum_i \dot{\sigma}_i^2 P_{ii}(z_i\bar{z}_i' - z_i z_i')/n\right\| \le \left(\sum_i \dot{\sigma}_i^4 P_{ii}^2\|z_i\|^2/n\right)^{1/2}\left(\sum_i \|z_i - \bar{z}_i\|^2/n\right)^{1/2} \to 0.$$

It follows that

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

It then follows from Lemmas A7 and A8 and T that

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

because $\epsilon_n \to 0$. 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, because $\bar{H}_n^{-1}$, $\bar{\Omega}_n + \bar{\Psi}_n$, H_n^{-1}, and $\Omega_n + \Psi_n$ are all bounded a.s.n, Lemma A6 implies

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

which gives the first conclusion.

For the second result, consider the case where $K/r_n \to \infty$. 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

$$(r_n/K)\left(\hat\Sigma_1+\hat\Sigma_2\right) = (r_n/K)\bar\Omega_n + (r_n/K)\bar\Psi_n + (r_n/K)o_p(1) + o_p(1) = (r_n/K)\bar\Psi_n + o_p(1).$$

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

$$(r_n/K)\left(\hat\Sigma_1+\hat\Sigma_2\right) = (r_n/K)\Omega_n + (r_n/K)\Psi_n + (r_n/K)o_p(1) + o_p(1) = (r_n/K)\Psi_n + o_p(1).$$

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 bounded a.s.n and by Lemma A6,

$$\begin{aligned} S_n'\tilde V S_n &= \left(S_n^{-1}\tilde H S_n^{-1\prime}\right)^{-1}\left(\hat\Sigma_1+\hat\Sigma_2\right)\left(S_n^{-1}\tilde H' S_n^{-1\prime}\right)^{-1} \\ &= \left(\bar H_n^{-1}+o_p(1)\right)\left(r_n\bar\Psi_n/K_n+o_p(1)\right)\left(\bar H_n^{-1}+o_p(1)\right) = \bar V_n^* + o_p(1). \end{aligned}$$

Similarly, $S_n'\hat V S_n = V_n^* + o_p(1)$, which gives the second conclusion. ∎

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 element by element from $a(\delta)$. It follows from $\hat\delta \xrightarrow{p} \delta_0$ that $\bar\delta \xrightarrow{p} \delta_0$, so by condition (iii), $B_n\hat A S_n^{-1\prime} = B_n A S_n^{-1\prime} + o_p(1)$. Then multiplying by B_n and using Theorem 4, we have

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

for $F_n = B_nAS_n^{-1}\bar V_n^{1/2}$ and $\bar Y_n = \bar V_n^{-1/2}S_n'(\delta-\delta_0)$, where the third equality in the preceding display follows from the Slutsky theorem given the continuity of the square root matrix. By Theorem 2, $\bar Y_n \xrightarrow{d} N(0, I_G)$. Also, from the proof of Theorem 2, it follows that this convergence is a.s. conditional on $\mathcal{Z}$. Then because $L_n = (F_nF_n')^{-1/2}F_n$ satisfies $L_nL_n' = I$, it follows from the Slutsky theorem and standard convergence in distribution results that

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

giving the conclusion. ∎

Proof of Corollary 1. Let $a(\delta) = c'\delta$, so $\bar{A} = A = c'$. Note that condition (i) of Theorem 5 is satisfied. Let $B_n = b_n$. Then $B_n A S_n^{-1\prime} = b_n c' S_n^{-1\prime}$ 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, $\lambda_{\min}(B_n A S_n^{-1\prime} \bar{V}_n S_n^{-1} A' B_n') = b_n^2 c' S_n^{-1\prime} \bar{V}_n S_n^{-1} c \geq C$; or if $K/r_n \to \infty$, then $\lambda_{\min}(B_n A S_n^{-1\prime} \bar{V}_n^* S_n^{-1} A' B_n') = b_n^2 c' S_n^{-1\prime} \bar{V}_n^* S_n^{-1} c \geq C$, which gives the first conclusion. The second conclusion follows similarly. ■

Proof of Corollary 2. We will show the result for $\hat{\delta}$; the result for $\tilde{\delta}$ follows analogously. Let $\gamma = \lim_{n\to\infty}(r_n/n)$, so γ exists and $\gamma \in \{0, 1\}$ by Assumption 2. Also,

$$\sqrt{r_n} S_n^{-1\prime} = \sqrt{r_n} \tilde{S}_n^{-1\prime} \text{diag}\left(1/\sqrt{n}, \ldots, 1/\sqrt{n}, 1/\sqrt{r_n}\right) \to R = \begin{bmatrix} \sqrt{\gamma} I & -\pi_1' \\ 0 & 1 \end{bmatrix}.$$

Consider first the case where $r_n = n$ so that $\gamma = 1$. Take $b_n = \sqrt{r_n}$ and note that $b_n c' S_n^{-1\prime} = c'(\sqrt{r_n} S_n^{-1\prime})$ is bounded. Also, $c'R \neq 0$ because R is nonsingular and $\|V_n\| \leq C$ a.s.n implying that $b_n^2 c' S_n^{-1\prime} V_n S_n^{-1} c = c' R V_n R' c + o_{\text{a.s.}}(1)$. Also $\Psi_n = S_n^{-1} \text{E}[\left(\sum_{i\neq j} P_{ij} U_i \varepsilon_j\right) \left(\sum_{i\neq j} P_{ij} U_i \varepsilon_j\right)' |\mathcal{Z}] S_n^{-1\prime}$ is positive semidefinite, so $V_n \geq H_n^{-1} \Omega_n H_n^{-1}$. Also, by Assumptions 2 and 4, there is $C > 0$ with $\lambda_{\min}(H_n^{-1} \Omega_n H_n^{-1}) \geq C$ a.s.n. Therefore, a.s.n,

$$b_n^2 c' S_n^{-1\prime} V_n S_n^{-1} c \geq c' R H_n^{-1} \Omega_n H_n^{-1} R' c + o(1) \geq C + o(1) \geq C. \quad \textbf{(A.1)}$$

The conclusion then follows from Corollary 1.

For $\gamma = 0$, let $a = (-\pi_1, 1)c$ and note that $c'R = (0, a) \neq 0$. If K/r_n is bounded, let $b_n = \sqrt{r_n}$. Then, as before, $b_n c' S_n^{-1\prime}$ is bounded and equation (A.1) is satisfied, and the conclusion follows. If $K/r_n \to \infty$, let $b_n = r_n/\sqrt{K}$. Note that $b_n c' S_n^{-1\prime} = \sqrt{r_n/K} c'(\sqrt{r_n} S_n^{-1\prime}) \to 0$, so $b_n c' S_n^{-1\prime}$ is bounded. Also, note that

$$\sqrt{r_n} S_n^{-1} e_G = \text{diag}(\sqrt{r_n/n}, \ldots, \sqrt{r_n/n}, 1) \begin{bmatrix} I & 0 \\ -\pi_1 & 1 \end{bmatrix} e_G = e_G.$$

Furthermore, a constant sign of $\text{E}[\varepsilon_i U_{iG}|Z]$ implies $\text{E}[\varepsilon_i U_{iG}|Z]\text{E}[\varepsilon_j U_{jG}|Z] \geq 0$, so by $P_{ii} \leq C < 1$,

$$\begin{aligned} \sum_{i\neq j} P_{ij}^2 \left(\text{E}[U_{iG}^2|\mathcal{Z}]\sigma_j^2 + \text{E}[\varepsilon_i U_{iG}|\mathcal{Z}]\text{E}[\varepsilon_j U_{jG}|\mathcal{Z}]\right) \Big/ K &\geq \sum_{i\neq j} P_{ij}^2 \text{E}[U_{iG}^2|\mathcal{Z}]\sigma_j^2/K \\ &\geq C \sum_{i\neq j} P_{ij}^2/K = C\left(\sum_{i,j} P_{ij}^2 - \sum_i P_{ii}^2\right) \Big/ K = C\left(1 - \sum P_{ii}^2/K\right) \geq C. \end{aligned}$$

Therefore, we have, a.s.,

$$\begin{aligned} (r_n/K)\Psi_n = \sqrt{r_n} S_n^{-1} e_G \Bigg[\sum_{i\neq j} P_{ij}^2 \Big(&\text{E}[U_{iG}^2|\mathcal{Z}]\sigma_j^2 \\ &+ \text{E}[\varepsilon_i U_{iG}|\mathcal{Z}]\text{E}[\varepsilon_j U_{jG}|\mathcal{Z}]\Big)/K \Bigg] e_G' \sqrt{r_n} S_n^{-1\prime} \\ \geq C e_G e_G'. \end{aligned}$$

Also, H_n is a.s. bounded so that $\lambda_{\min}(H_n^{-1}) = 1/\lambda_{\max}(H_n) \geq C + o_{\text{a.s.}}(1)$. It then follows from $c'R = ae_G'$ that

$$b_n^2 c' S_n^{-1\prime} \bar{V}_n^* S_n^{-1} c = r_n c' S_n^{-1\prime} H_n^{-1} (r_n/K) \Psi_n H_n^{-1} S_n^{-1} c \geq C r_n c' S_n^{-1\prime} H_n^{-1} e_G e_G' H_n^{-1} S_n^{-1} c$$
$$= a^2 C (e_G' H_n^{-1} e_G)^2 + o_{\text{a.s.}}(1) \geq C + o_{\text{a.s.}}(1).$$

The conclusion then follows from Corollary 1. ∎

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. *Under Assumption 1 and for any subset I_2 of the set $\left\{(i,j)_{i,j=1}^n\right\}$ and any subset I_3 of $\left\{(i,j,k)_{i,j,k=1}^n\right\}$, (i) $\sum_{\mathcal{I}_2} P_{ij}^4 \leq K$; (ii) $\sum_{\mathcal{I}_3} P_{ij}^2 P_{jk}^2 \leq K$; and (iii) $\sum_{\mathcal{I}_3} \left| P_{ij}^2 P_{ik} P_{jk} \right| \leq K$, a.s.n.*

Proof. By Assumption 1, $Z'Z$ is nonsingular a.s.n. Also, because P is idempotent, $\text{rank}(P) = \text{tr}(P) = K$, $0 \leq P_{ii} \leq 1$, and $\sum_{j=1}^n P_{ij}^2 = P_{ii}$. Therefore, a.s.$n$,

$$\sum_{\mathcal{I}_2} P_{ij}^4 \leq \sum_{i,j=1}^n P_{ij}^2 = \sum_{i=1}^n P_{ii} = K,$$

$$\sum_{\mathcal{I}_3} P_{ij}^2 P_{jk}^2 \leq \sum_{j=1}^n \left(\sum_{i=1}^n P_{ij}^2\right)\left(\sum_{k=1}^n P_{jk}^2\right) = \sum_{j=1}^n P_{jj}^2 \leq \sum_{j=1}^n P_{jj} = K,$$

$$\sum_{\mathcal{I}_3} \left| P_{ij}^2 P_{ik} P_{jk} \right| \leq \sum_{i,j} P_{ij}^2 \sum_k \left| P_{ik} P_{jk} \right| \leq \sum_{i,j} P_{ij}^2 \sqrt{\sum_k P_{ik}^2} \sqrt{\sum_k P_{jk}^2}$$
$$\leq \sum_{i,j} P_{ij}^2 \sqrt{P_{ii} P_{jj}} \leq \sum_{i,j} P_{ij}^2 = K.$$
∎

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

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

Proof. To show part (i), note that

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

Note that $\text{tr}(A') = \text{tr}(A)$ and $\text{tr}(AB) = \text{tr}(BA)$ for any square matrices A and B. Then, $\text{tr}\left[(P-D)^4\right] = \text{tr}(P) - 4\text{tr}(PD) + 4\text{tr}(PD^2) + 2\text{tr}(PDPD) - 4\text{tr}(PD^3) + \text{tr}(D^4)$. By $0 \leq$

$P_{ii} \leq 1$ we have $D^j \leq I$ for any positive integer j and $\text{tr}(PD^j) = \text{tr}(PD^j P) \leq \text{tr}(P) = K$ a.s.n. Also, a.s.n, $\text{tr}(PDPD) = \text{tr}(PDPDP) \leq \text{tr}(PD^2 P) \leq \text{tr}(P) = K$ and $\text{tr}(D^4) = \sum_i P_{ii}^4 \leq K$. Therefore, by T we have $\left|\text{tr}\left[(P-D)^4\right]\right| \leq 16K$, giving conclusion (i).

Next, let L be the lower triangular matrix with $L_{ij} = P_{ij}1(i > j)$. 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 positive integer j, $[(L')^j]' = L^j$. Then using $\text{tr}(AB) = \text{tr}(BA)$ and $\text{tr}(A') = \text{tr}(A)$,

$$\text{tr}((P-D)^4) = 2\text{tr}(L^4) + 8\text{tr}(L^3L') + 4\text{tr}(L^2L'^2) + 2\text{tr}(L'LL'L).$$

Next, compute each of the terms. Note that

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

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

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

and

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

$$+\sum_{j<i<k} P_{ij}P_{jk}P_{kj}P_{ji} + \sum_{j<\ell<i} P_{ij}P_{ji}P_{i\ell}P_{\ell i}$$

$$+\sum_{\ell<j<i} P_{ij}P_{ji}P_{i\ell}P_{\ell i} + \left(\sum_{\ell<j<k<i} + \sum_{j<\ell<k<i} + \sum_{\ell<j<i<k} + \sum_{j<\ell<i<k}\right)$$

$$\times P_{ij}P_{jk}P_{k\ell}P_{\ell i}$$

$$= \sum_{i<j} P_{ij}^4 + 2\sum_{i<j<k}\left(P_{ij}^2P_{ik}^2 + P_{ik}^2P_{jk}^2\right) + 4\sum_{i<j<k<\ell} P_{ik}P_{kj}P_{j\ell}P_{\ell i}.$$

Summing up gives the result $\text{tr}((P-D)^4) = 2\sum_{i<j} P_{ij}^4 + 4\sum_{i<j<k}(P_{ij}^2P_{jk}^2 + P_{ik}^2P_{jk}^2 + P_{ij}^2P_{ik}^2) + 8S_n$. Then by T and Lemma B1, we have

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

a.s.n, thus giving part (iii). That is, $S_n = O_{\text{a.s.}}(K)$.

To show part (ii), take $\{\varepsilon_i\}$ to be a sequence of independent and identically distributed 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

$$\Delta_1 = \sum_{i<j<k}\left[P_{ij}P_{ik}\varepsilon_j\varepsilon_k + P_{ij}P_{jk}\varepsilon_i\varepsilon_k + P_{ik}P_{jk}\varepsilon_i\varepsilon_j\right],$$

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

Note that by Lemma A1,

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

Also, note that

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

and

$$\begin{aligned}
\text{E}\left[\Delta_2^2|\mathcal{Z}\right] = \text{E}\Big[&\left(\textstyle\sum_{i<j<k} P_{ij}P_{ik}\varepsilon_j\varepsilon_k + P_{ij}P_{jk}\varepsilon_i\varepsilon_k\right) \\
&\times\left(\textstyle\sum_{\ell<m<q} P_{\ell m}P_{\ell q}\varepsilon_m\varepsilon_q + P_{\ell m}P_{mq}\varepsilon_\ell\varepsilon_q\right)|\mathcal{Z}\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\{i,\ell\}<j<k} P_{ij}P_{ik}P_{\ell j}P_{\ell k} + \sum_{i<\{j,m\}<k} P_{ij}P_{jk}P_{im}P_{mk} \\
&\quad + \sum_{i<j<m<k} P_{ij}P_{ik}P_{jm}P_{mk} + \sum_{\ell<i<j<k} P_{ij}P_{jk}P_{\ell i}P_{\ell k} \\
&= \sum_{i<j<k} P_{ij}^2P_{ik}^2 + \sum_{i<j<k} P_{ij}^2P_{jk}^2 + 2\sum_{i<\ell<j<k} P_{ij}P_{ik}P_{\ell j}P_{\ell k} \\
&\quad +2\sum_{i<j<m<k} P_{ij}P_{jk}P_{im}P_{mk} \\
&\quad + \sum_{i<j<k<\ell} P_{ij}P_{i\ell}P_{jk}P_{k\ell} + \sum_{i<j<k<\ell} P_{jk}P_{k\ell}P_{ij}P_{i\ell} \\
&= \sum_{i<j<k} P_{ij}^2P_{ik}^2 + \sum_{i<j<k} P_{ij}^2P_{jk}^2 + 2S_n = O_{\text{a.s.}}(K).
\end{aligned}
$$

Because $\Delta_1 = \Delta_2 + \Delta_3$, it follows that $\mathrm{E}\left[\Delta_1^2|\mathcal{Z}\right] = \mathrm{E}\left[\Delta_2^2|\mathcal{Z}\right] + \mathrm{E}\left[\Delta_3^2|\mathcal{Z}\right] + 2\mathrm{E}\left[\Delta_2\Delta_3|\mathcal{Z}\right] = O_{\text{a.s.}}(K) + 2S_n = O_{\text{a.s.}}(K)$. Therefore, by T, the expression for $\mathrm{E}\left[\Delta_3^2|\mathcal{Z}\right]$ given previously, and $\Delta_3 = \Delta_1 - \Delta_2$,

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

■

LEMMA B3. *Let L be the lower triangular matrix with $L_{ij} = P_{ij}1(i>j)$. Then, under Assumption 2, $\|LL'\| \leq C\sqrt{K}$ a.s.n, where $\|A\| = \left[\mathrm{Tr}\left(A'A\right)\right]^{1/2}$.*

Proof. From the proof of Lemma B2 and by Lemma B1 and Lemma B2(ii), we have a.s.n

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

Taking square roots gives the answer. ■

For Lemma B4, which follows, let $\phi_i = \phi_i(\mathcal{Z})$ $(i = 1, \ldots, n)$ denote some sequence of measurable functions. In applications of this lemma, we will take $\phi_i(\mathcal{Z})$ to be either conditional variances or conditional covariances given $\mathcal{Z}$. Also, to set some notation, let $\sigma_i^2 = \sigma_i^2(\mathcal{Z}) = \mathrm{E}[\varepsilon_i^2|\mathcal{Z}]$, $\omega_i^2 = \omega_{in}^2(\mathcal{Z}) = \mathrm{E}[u_i^2|\mathcal{Z}]$, and $\gamma_i = \gamma_{in}(\mathcal{Z}) = \mathrm{E}[u_i\varepsilon_i|\mathcal{Z}]$, where 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. Let the following results apply.

LEMMA B4. *Suppose that (a) P is a symmetric, idempotent matrix with rank $(P) = K$ and $P_{ii} \leq C < 1$; (b) $(u_1, \varepsilon_1), \ldots, (u_n, \varepsilon_n)$ are independent conditional on $\mathcal{Z}$; (c) there exists a constant C such that, a.s., $\sup_i \mathrm{E}\left(u_i^4|\mathcal{Z}\right) \leq C$, $\sup_i \mathrm{E}\left(\varepsilon_i^4|\mathcal{Z}\right) \leq C$, and $\sup_i|\phi_i| = \sup_i|\phi_i(\mathcal{Z})| \leq C$. Then, a.s.,* *(i)* $\mathrm{E}\left[\left(\frac{1}{K}\sum_{i<k} P_{ki}^2\phi_k\left(u_i\varepsilon_i-\gamma_i\right)\right)^2|\mathcal{Z}\right]\to 0$

(ii) $\mathrm{E}\left[\left(\frac{1}{K}\sum_{i<k} P_{ki}^2\phi_k\left(\varepsilon_j^2-\sigma_j^2\right)\right)^2|\mathcal{Z}\right]\to 0$;

(iii) $\mathrm{E}\left[\left(\frac{1}{K}\sum_{i<k} P_{ki}^2\phi_k\left(u_j^2-\omega_j^2\right)\right)^2|\mathcal{Z}\right]\to 0$

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

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

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

Proof. To show part (i), note that

$$
\begin{aligned}
&\mathrm{E}\left[\left(\frac{1}{K}\sum_{i<k\le n} P_{ki}^2\phi_k u_i\varepsilon_i-\gamma_i\right)^2|\mathcal{Z}\right]\\
&\quad=\frac{1}{K^2}\sum_{i<k\le n} P_{ki}^4\phi_k^2\left\{\mathrm{E}\left(u_i^2\varepsilon_i^2|\mathcal{Z}\right)-\gamma_i^2\right\}\\
&\qquad+\frac{2}{K^2}\sum_{1\le i<k<l\le n} P_{ki}^2P_{li}^2\phi_k\phi_l\left\{\mathrm{E}\left(u_i^2\varepsilon_i^2|\mathcal{Z}\right)-\gamma_i^2\right\}\\
&\quad\le\frac{1}{K^2}\sum_{1\le i<k\le n} P_{ki}^4\phi_k^2\left\{\sqrt{\mathrm{E}\left(u_i^4|\mathcal{Z}\right)\mathrm{E}\left(\varepsilon_i^4|\mathcal{Z}\right)}+\mathrm{E}\left(u_i^2|\mathcal{Z}\right)\mathrm{E}\left(\varepsilon_i^2|\mathcal{Z}\right)\right\}\\
&\qquad+\frac{2}{K^2}\sum_{1\le i<k<l\le n} P_{ki}^2P_{li}^2|\phi_k||\phi_l|\left\{\sqrt{\mathrm{E}\left(u_i^4|\mathcal{Z}\right)\mathrm{E}\left(\varepsilon_i^4|\mathcal{Z}\right)}+\mathrm{E}\left(u_i^2|\mathcal{Z}\right)\mathrm{E}\left(\varepsilon_i^2|\mathcal{Z}\right)\right\}\\
&\quad\le C\left\{\frac{1}{K^2}\sum_{1\le i<k\le n} P_{ki}^4+\frac{2}{K^2}\sum_{1\le i<k<l\le n} P_{ki}^2P_{li}^2\right\}\to 0,
\end{aligned}
$$

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 a.s. follows from applying Lemma B1(i) and (ii). Parts (ii) and (iii) can be proved in essentially the same way as part (i); hence, to avoid redundancy, we do not give detailed arguments for these parts.

To show part (iv), first let L be a lower triangular matrix with (i,j)th element $L_{ij}=P_{ij}1(i>j)$ as in Lemma B3 and define $D_\gamma=\mathrm{diag}(\gamma_1,\ldots,\gamma_n)$, $D_\phi=\mathrm{diag}(\phi_1,\ldots,\phi_n)$, $u=(u_1,\ldots,u_n)'$, and $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)'$. It then follows by direct multiplication that

$$
\begin{aligned}
\varepsilon'L'D_\phi Lu-\mathrm{tr}\left\{L'D_\phi LD_\gamma\right\}=&\sum_{1\le i<k\le n} P_{ki}^2\phi_k\left(u_i\varepsilon_i-\gamma_i\right)\\
&+\sum_{1\le i<j<k\le n} P_{ki}P_{kj}\phi_k\left(u_i\varepsilon_j+u_j\varepsilon_i\right)
\end{aligned}
$$

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

$$\begin{aligned}\frac{1}{K^2}\,\mathrm{E}\left[\left(\sum_{1\le i<j<k\le n} P_{ki}P_{kj}\,\phi_k\left(u_i\varepsilon_j+u_j\varepsilon_i\right)\right)^2 \mid \mathcal{Z}\right] \\ \le 2\frac{1}{K^2}\mathrm{E}\left[\left(u'L'D_\phi L\varepsilon-\mathrm{tr}\left\{L'D_\phi LD_\gamma\right\}\right)^2|\mathcal{Z}\right] \\ +\,2\frac{1}{K^2}\mathrm{E}\left[\left(\sum_{1\le i<k\le n} P_{ki}^2\,\phi_k\left(u_i\,\varepsilon_i-\gamma_i\right)\right)^2|\mathcal{Z}\right].\end{aligned} \tag{B.1}$$

It has already been shown in the proof of part (i) that $(1/K^2)\,\mathrm{E}\left[\left(\sum_{1\le i<k\le n} P_{ki}^2\phi_k\left(u_i\,\varepsilon_i-\gamma_i\right)\right)^2|\mathcal{Z}\right]\to 0$ a.s. $\mathbb{P}_{\mathcal{Z}}$, so what remains to be shown is that $\left(1/K^2\right)$ $\mathrm{E}\left[\left(u'L'D_\phi L\varepsilon-tr\left\{L'D_\phi LD_\gamma\right\}\right)^2|\mathcal{Z}\right]\to 0$ a.s. $\mathbb{P}_{\mathcal{Z}}$. To show the latter, note first that, by straightforward calculations, we have

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

Next, note that, by straightforward calculation, we have

$$\begin{aligned}&\mathrm{E}\left[\varepsilon u'\otimes\varepsilon u'|\mathcal{Z}\right] \\ &=\begin{pmatrix}\sigma_1^2\omega_1^2e_1e_1' & \sigma_1^2\omega_2^2e_1e_2' & \cdots & \sigma_1^2\omega_n^2e_1e_n' \\ \sigma_2^2\omega_1^2e_2e_1' & \sigma_2^2\omega_2^2e_2e_2' & \cdots & \sigma_2^2\omega_n^2e_2e_n' \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n^2\omega_1^2e_ne_1' & \sigma_n^2\omega_2^2e_ne_2' & \cdots & \sigma_n^2\omega_n^2e_ne_n'\end{pmatrix}+\begin{pmatrix}\gamma_1^2e_1e_1' & \gamma_1\gamma_2e_2e_1' & \cdots & \gamma_1\gamma_ne_ne_1' \\ \gamma_2\gamma_1e_1e_2' & \gamma_2^2e_2e_2' & \cdots & \gamma_2\gamma_ne_ne_2' \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_n\gamma_1e_1e_n' & \gamma_n\gamma_2e_2e_n' & \cdots & \gamma_n^2e_ne_n'\end{pmatrix} \\ &+\begin{pmatrix}\vartheta_1e_1e_1' & \underset{n\times n}{0} & \cdots & \underset{n\times n}{0} \\ \underset{n\times n}{0} & \vartheta_2e_2e_2' & \cdots & \underset{n\times n}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{n\times n}{0} & \underset{n\times n}{0} & \cdots & \vartheta_ne_ne_n'\end{pmatrix}+\begin{pmatrix}\gamma_1\otimes D_\gamma & \underset{n\times n}{0} & \cdots & \underset{n\times n}{0} \\ \underset{n\times n}{0} & \gamma_2\otimes D_\gamma & \cdots & \underset{n\times n}{0} \\ \vdots & \vdots & \ddots & \vdots \\ \underset{n\times n}{0} & \underset{n\times n}{0} & \cdots & \gamma_n\otimes D_\gamma\end{pmatrix} \\ &=\left(D_\sigma\otimes I_n\right)\mathrm{vec}\left(I_n\right)\mathrm{vec}\left(I_n\right)'\left(D_\omega\otimes I_n\right) \\ &\quad+\left(D_\gamma\otimes I_n\right)\underline{K}_{nn}\left(D_\gamma\otimes I_n\right)+\underline{\mathrm{E}}'D_\vartheta\underline{\mathrm{E}}+\left(D_\gamma\otimes D_\gamma\right),\end{aligned} \tag{B.3}$$

where $\underline{K}_{nn}$ is an $n^2\times n^2$ commutation matrix such that, for any $n\times n$ matrix A, $\underline{K}_{nn}$ $\mathrm{vec}\,(A)=\mathrm{vec}\left(A'\right)$. (See Magnus and Neudecker, 1988, pp. 46–48, for more on commutation matrices.) Also, here, $D_\gamma=\mathrm{diag}\,(\gamma_1,\ldots,\gamma_n)$, $D_\sigma=\mathrm{diag}\left(\sigma_1^2,\ldots,\sigma_n^2\right)$, $D_\omega=\mathrm{diag}\left(\omega_1^2,\ldots,\omega_n^2\right)$, $D_\vartheta=\mathrm{diag}\,(\vartheta_1,\ldots,\vartheta_n)$ with $\vartheta_i=\mathrm{E}\left[\varepsilon_i^2u_i^2|\mathcal{Z}\right]-\sigma_i^2\omega_i^2-2\gamma_i^2$ for $i=1,\ldots,n$, $\underline{\mathrm{E}}=\left(e_1\otimes e_1\vdots e_2\otimes e_2\vdots\cdots\vdots e_n\otimes e_n\right)'$, and e_i is the ith column of an $n\times n$ identity matrix. It follows from (B.2) and (B.3) and by straightforward calculations that

$$
\begin{aligned}
&\frac{1}{K^2}\mathrm{E}\left[\left(u'L'D_\phi L\varepsilon - \mathrm{tr}\left\{L'D_\phi LD_\gamma\right\}\right)^2 \mid \mathcal{Z}\right] \\
&\quad = \frac{1}{K^2}\mathrm{tr}\left\{\left(L'D_\phi L \otimes L'D_\phi L\right)\mathrm{E}\left[\varepsilon u' \otimes \varepsilon u' |\mathcal{Z}\right]\right\} - \frac{1}{K^2}\left[\mathrm{tr}\left\{L'D_\phi LD_\gamma\right\}\right]^2 \\
&\quad = \frac{1}{K^2}\mathrm{vec}\left(I_n\right)'\left(D_\omega L'D_\phi LD_\sigma \otimes L'D_\phi L\right)\mathrm{vec}\left(I_n\right) \\
&\qquad + \frac{1}{K^2}\mathrm{tr}\left\{\left(D_\gamma L'D_\phi LD_\gamma \otimes L'D_\phi L\right)\underline{K}_{nn}\right\} \\
&\qquad + \frac{1}{K^2}\mathrm{tr}\left\{\left(L'D_\phi L \otimes L'D_\phi L\right)\underline{\mathrm{E}}'D_\vartheta\underline{\mathrm{E}}\right\} + \frac{1}{K^2}\mathrm{tr}\left\{\left(L'D_\phi LD_\gamma \otimes L'D_\phi LD_\gamma\right)\right\} \\
&\qquad - \frac{1}{K^2}\left[\mathrm{tr}\left\{L'D_\phi LD_\gamma\right\}\right]^2 \\
&\quad = \frac{1}{K^2}\mathrm{tr}\left\{L'D_\phi LD_\omega L'D_\phi LD_\sigma\right\} + \frac{1}{K^2}\mathrm{tr}\left\{\left(D_\gamma L'D_\phi LD_\gamma \otimes L'D_\phi L\right)\underline{K}_{nn}\right\} \\
&\qquad + \frac{1}{K^2}\mathrm{tr}\left\{\left(L'D_\phi L \otimes L'D_\phi L\right)\underline{\mathrm{E}}'D_\vartheta\underline{\mathrm{E}}\right\}. \qquad \textbf{(B.4)}
\end{aligned}
$$

Focusing first on the first term of (B.4), and letting $\overline{\omega}^2 = \max_{1\le i\le n}\omega_i^2$, $\overline{\sigma}^2 = \max_{1\le i\le n}\sigma_i^2$, and $\overline{\phi}^2 = \max_{1\le i\le n}\phi_i^2$, we get

$$
\begin{aligned}
\frac{1}{K^2}\mathrm{tr}\left\{L'D_\phi LD_\omega L'D_\phi LD_\sigma\right\} &\le \overline{\omega}^2\overline{\sigma}^2\,\overline{\phi}^2\frac{1}{K^2}\mathrm{tr}\left\{L'LL'L\right\} \\
&\le C\frac{1}{K^2}\mathrm{tr}\left\{L'LL'L\right\} = \frac{C}{K^2}\left\|LL'\right\|^2 \quad \text{a.s. } \mathbb{P}_\mathcal{Z}, \qquad \textbf{(B.5)}
\end{aligned}
$$

where the first inequality follows by repeated application of CS and of the simple inequality

$$
\mathrm{tr}\left\{A'\Lambda A\right\} \le \max_{1\le i\le n}\lambda_i\,\mathrm{tr}\left(A'A\right), \qquad \textbf{(B.6)}
$$

which holds for $n\times n$ matrices A and $\Lambda = \mathrm{diag}\left(\lambda_1,\ldots,\lambda_n\right)$ such that $\lambda_i \ge 0$ for all i, and where the second inequality follows in light of the assumptions of the lemma.

Turning our attention now to the second term of (B.4), we make use of the fact that, for $n\times n$ matrices A and B, $\mathrm{tr}\left\{(A\otimes B)\underline{K}_{nn}\right\} = \mathrm{tr}\{AB\}$ (a specialization of the result given by Abadir and Magnus, 2005, p. 304) to obtain $K^{-2}\mathrm{tr}\left\{\left(D_\gamma L'D_\phi LD_\gamma \otimes L'D_\phi L\right)\underline{K}_{nn}\right\} = K^{-2}\mathrm{tr}\left\{L'D_\phi LD_\gamma L'D_\phi LD_\gamma\right\}$. As in (B.5), by repeated use of CS and the inequality (B.6), we obtain

$$
\frac{1}{K^2}\mathrm{tr}\left\{\left(D_\gamma L'D_\phi LD_\gamma \otimes L'D_\phi L\right)\underline{K}_{nn}\right\} \le \frac{C}{K^2}\left\|LL'\right\|^2 \quad \text{a.s. } \mathbb{P}_\mathcal{Z}. \qquad \textbf{(B.7)}
$$

Finally, to analyze the third term of (B.4), we note that

$$\frac{1}{K^2}\left|\mathrm{tr}\left\{\left(L'D_\phi L\otimes L'D_\phi L\right)\underline{\mathrm{E}}'D_\vartheta\underline{\mathrm{E}}\right\}\right|$$

$$\leq\frac{1}{K^2}\sum_{i=1}^{n}|\vartheta_i|\left(e_i'L'D_\phi Le_i\right)^2\leq\frac{1}{K^2}\sum_{i=1}^{n}|\vartheta_i|\left(e_i'L'D_\phi^2 Le_i\right)\left(e_i'L'Le_i\right)$$

$$\leq\overline{\phi}^2\frac{1}{K^2}\sum_{i=1}^{n}|\vartheta_i|\left(e_i'L'Le_i\right)^2$$

$$\leq C\frac{1}{K^2}\sum_{i=1}^{n}\left(e_i'L'Le_i\right)^2\leq C\frac{1}{K^2}\sum_{i=1}^{n}\left(e_i'P'Pe_i\right)^2=C\frac{1}{K^2}\sum_{i=1}^{n}P_{ii}^2$$

$$\leq C\frac{1}{K^2}\sum_{i=1}^{n}P_{ii}=\frac{C}{K}\ \text{ a.s. }\mathbb{P}_{\mathcal{Z}},\tag{B.8}$$

where the first inequality follows from T, the second inequality follows from CS, the third inequality makes use of (B.6), the fourth inequality uses CS and T and follows in light of the assumptions of the lemma, and the last inequality holds because $P_{ii} < 1$.

In light of (B.4), it follows from (B.5), (B.7), and (B.8) and Lemma B3 that $(1/K^2)$ $\mathrm{E}[(u'L'D_\phi L\varepsilon-\mathrm{tr}\{L'D_\phi LD_\gamma\})^2\mid\mathcal{Z}]\leq 2C(1/K^2)\,\|LL'\|^2+C(1/K)\leq C/K$ a.s. $\mathbb{P}_{\mathcal{Z}}$, which shows part (iv).

It is easily seen that parts (v) and (vi) can be proved in essentially the same way as part (iv) (by taking $u_i=\varepsilon_i$); hence, to avoid redundancy, we do not give detailed arguments for these parts. ∎

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 because 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, $\mathrm{E}\left[\|w_{1n}\|^4|\mathcal{Z}\right]\leq\sum_i\mathrm{E}\left[\|w_{in}\|^4|\mathcal{Z}\right]\leq C\sum_i\mathrm{E}\left[\|W_{in}\|^4|\mathcal{Z}\right]\to 0$ a.s., so by a conditional version of M, we deduce that for any $v>0$, $P(|w_{1n}|\geq v\mid\mathcal{Z})\to 0$. Moreover, note that $\sup_n\mathrm{E}\left[|P(|w_{1n}|\geq v\mid\mathcal{Z})|^2\right]<\infty$. It follows that, by Theorem 25.12 of Billingsley (1986), $P(|w_{1n}|\geq v)=\mathrm{E}\left[P(|w_{1n}|\geq v\mid\mathcal{Z})\right]\to 0$ as $n\to\infty$; i.e., $w_{1n}\xrightarrow{p}0$ 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 $\mathcal{Z}$, $\sum_{i=2}^n y_{in}\xrightarrow{d}N(0,1)$, a.s. To proceed, let $\mathcal{X}_i=(W_{in}',U_i',\varepsilon_i)'$ for $i=1,\ldots,n$. Define the σ-fields $F_{i,n}=\sigma(\mathcal{X}_1,\ldots,\mathcal{X}_i)$ for $i=1,\ldots,n$. Note that, by construction, $F_{i-1,n}\subseteq F_{i,n}$. Moreover, it is straightforward to verify that, conditional on $\mathcal{Z}$, $\{y_{in},\mathcal{F}_{i,n},1\leq i\leq n,n\geq 2\}$ is a martingale difference array, and we can apply the martingale central limit theorem. As before, let $\sigma_i^2=\mathrm{E}[\varepsilon_i^2|\mathcal{Z}]$, $\omega_i^2=\omega_{in}^2(\mathcal{Z})=\mathrm{E}[u_i^2|\mathcal{Z}]$, and $\gamma_i=\gamma_{in}(\mathcal{Z})=\mathrm{E}[u_i\varepsilon_i|\mathcal{Z}]$, where 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, note that $\mathrm{E}[w_{in}\bar{y}_{jn}|\mathcal{Z}]=0$ for all i and j and that

$$\mathrm{E}\left[(\bar{y}_{in})^2\,|\mathcal{Z}\right] = \sum_{j<i}\sum_{k<i}\mathrm{E}\left[(u_j P_{ij}\varepsilon_i + u_i P_{ij}\varepsilon_j)(u_k P_{ik}\varepsilon_i + u_i P_{ik}\varepsilon_k)|\mathcal{Z}\right]/K$$

$$= \sum_{j<i} P_{ij}^2\left[\omega_j^2\sigma_i^2 + \omega_i^2\sigma_j^2 + 2\gamma_i\gamma_j\right]\Big/K.$$

Thus,

$$s_n^2(\mathcal{Z}) = \mathrm{E}\left[\left(\textstyle\sum_{i=2}^n y_{in}\right)^2|\mathcal{Z}\right] = \sum_{i=2}^n\left(\mathrm{E}\left[w_{in}^2|\mathcal{Z}\right] + \mathrm{E}\left[\bar{y}_{in}^2|\mathcal{Z}\right]\right)$$

$$= b_{1n}' D_n b_{1n} - \mathrm{E}\left[w_{1n}^2|\mathcal{Z}\right] + \sum_{i\neq j} P_{ij}^2\left[\omega_j^2\sigma_i^2 + \omega_i^2\sigma_j^2 + 2\gamma_i\gamma_j\right]\Big/K$$

$$= b_{1n}' D_n b_{1n} + b_{2n}'\bar{\Sigma}_n b_{2n} + o_{\text{a.s.}}(1)$$

$$= \Xi_n^{-1/2}\left(c_{1n}' D_n c_{1n} + c_{2n}'\bar{\Sigma}_n c_{2n}\right)\Xi_n^{-1/2} + o_{\text{a.s.}}(1)$$

$$= \Xi_n^{-1/2}\,\Xi_n\,\Xi_n^{-1/2} + o_{\text{a.s.}}(1) = 1 + o_{\text{a.s.}}(1) \to 1 \ \text{ a.s.},$$

where $D_n = D_n(\mathcal{Z}) = \sum_{i=1}^n \mathrm{E}\left[W_{in}W_{in}'|\mathcal{Z}\right]$ and

$$\bar{\Sigma}_n = \bar{\Sigma}_n(\mathcal{Z}) = \sum_{i\neq j} P_{ij}^2\left(\mathrm{E}[U_i U_i'|\mathcal{Z}]\mathrm{E}[\varepsilon_j^2|\mathcal{Z}] + \mathrm{E}[U_i\varepsilon_i|\mathcal{Z}]\mathrm{E}[\varepsilon_j U_j'|\mathcal{Z}]\right)\Big/K.$$

Thus, $s_n^2(\mathcal{Z})$ is bounded and bounded away from zero a.s. Also, $\sum_{i=2}\mathrm{E}\left[y_{in}^4|\mathcal{Z}\right] \le C\sum_{i=2}^n \mathrm{E}\left[\|W_{in}\|^4|\mathcal{Z}\right] + C\sum_{i=2}^n\mathrm{E}\left[\bar{y}_{in}^4|\mathcal{Z}\right]$. By condition (iv), $\sum_{i=2}^n\mathrm{E}\left[\|W_{in}\|^4|\mathcal{Z}\right]\to 0$. Let $\bar{y}_{in}^{\varepsilon} = \sum_{j<i} u_j P_{ij}\varepsilon_i/\sqrt{K}$ and $\bar{y}_{in}^{u} = \sum_{j<i} u_i P_{ij}\varepsilon_j/\sqrt{K}$. By $\left|P_{ij}\right| < 1$ and $\sum_j P_{ij}^2 = P_{ii}$, we have that a.s.

$$\sum_{i=2}^n \mathrm{E}\left[\left(\bar{y}_{in}^{\varepsilon}\right)^4|\mathcal{Z}\right] \le \frac{C}{K^2}\sum_{i=2}^n\sum_{j,k,\ell,m<i} P_{ij}P_{ik}P_{i\ell}P_{im}\mathrm{E}\left[\varepsilon_i^4|\mathcal{Z}\right]\mathrm{E}\left[u_j u_k u_\ell u_m|\mathcal{Z}\right]$$

$$\le \frac{C}{K^2}\sum_{i=2}^n\left(\sum_{j<i}P_{ij}^4 + \sum_{j,k<i}P_{ij}^2P_{ik}^2\right) \le CK/K^2 \to 0.$$

Similarly, $\sum_{i=2}^n\mathrm{E}\left[\left(\bar{y}_{in}^{u}\right)^4|\mathcal{Z}\right]\to 0$ a.s., so that

$$\sum_{i=2}^n\mathrm{E}\left[\bar{y}_{in}^4|\mathcal{Z}\right] \le C\sum_{i=2}^n\left\{\mathrm{E}\left[\left(\bar{y}_{in}^{\varepsilon}\right)^4|\mathcal{Z}\right] + \mathrm{E}\left[\left(\bar{y}_{in}^{u}\right)^4|\mathcal{Z}\right]\right\}\to 0$$

Then by T we have $\sum_{i=2}^n\mathrm{E}\left[y_{in}^4|\mathcal{Z}\right]\to 0$ a.s.

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

$$P\left(\left|\textstyle\sum_{i=2}^n\mathrm{E}\left[y_{in}^2|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}\right] - s_n^2(\mathcal{Z})\right| \ge \epsilon \mid \mathcal{Z}\right)\to 0. \tag{B.9}$$

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

$$\sum_{i=2}^{n}\mathrm{E}\left[y_{in}^2|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}\right]-s_n^2(\mathcal{Z})=\sum_{i=2}^{n}\left(\mathrm{E}[w_{in}^2|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}]-\mathrm{E}[w_{in}^2|\mathcal{Z}]\right)$$
$$+\sum_{i=2}^{n}\mathrm{E}\left[w_{in}\bar{y}_{in}|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}\right]+\sum_{i=2}^{n}\left(\mathrm{E}[\bar{y}_{in}^2|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}]-\mathrm{E}[\bar{y}_{in}^2|\mathcal{Z}]\right). \quad \textbf{(B.10)}$$

We will show that each term on the right-hand side of (B.10) converges to zero a.s. To proceed, note first that by independence of $W_{1n},\ldots,W_{nn}$ conditional on $\mathcal{Z}$, $\mathrm{E}[w_{in}^2|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}]=\mathrm{E}[w_{in}^2|\mathcal{Z}]$ a.s. Next, note that $\mathrm{E}\left[w_{in}\bar{y}_{in}|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}\right]=\mathrm{E}[w_{in}u_i|\mathcal{Z}]\sum_{j<i}P_{ij}\varepsilon_j/\sqrt{K}+\mathrm{E}[w_{in}\varepsilon_i|\mathcal{Z}]\sum_{j<i}P_{ij}u_j/\sqrt{K}$. Let $\delta_i=\delta_i(\mathcal{Z})=\mathrm{E}[w_{in}u_i|\mathcal{Z}]$ 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 $\bar{P}_{ij}=0$, $j\le i$, and let $\delta=(\delta_1,\ldots,\delta_n)$. Then, $\sum_{i=2}^{n}\sum_{j<i}\delta_i P_{ij}\varepsilon_j/\sqrt{K}=\delta'\bar{P}'\varepsilon/\sqrt{K}$. By CS, $\delta'\delta=\sum_{i=1}^{n}\left(\mathrm{E}\left[w_{in}u_i|\mathcal{Z}\right]\right)^2\le\sum_{i=1}^{n}\mathrm{E}[w_{in}^2|\mathcal{Z}]\mathrm{E}[u_i^2|\mathcal{Z}]\le C$ a.s. By Lemma B3, $\left\|\bar{P}'\bar{P}\right\|\le C\sqrt{K}$ a.s., which in turn implies that $\lambda_{\max}\left(\bar{P}'\bar{P}\right)\le C\sqrt{K}$ a.s. It then follows given $\mathrm{E}\left[u_j^2|\mathcal{Z}\right]\le C$ a.s. that $\mathrm{E}[(\delta'\bar{P}'\varepsilon/\sqrt{K})^2|\mathcal{Z}]\le C\delta'\bar{P}'\bar{P}\delta/K\le C\|\delta\|^2/\sqrt{K}\le C/\sqrt{K}\to 0$ a.s., so that by M we have for any $\epsilon>0$, $P\left(\left|\delta(\mathcal{Z})'\bar{P}'\varepsilon/\sqrt{K}\right|\ge\epsilon|\mathcal{Z}\right)\to 0$ a.s. Similarly, we have $\sum_{i=2}^{n}\mathrm{E}\left[w_{in}\varepsilon_i|\mathcal{Z}\right]\sum_{j<i}P_{ij}u_j/\sqrt{K}\to 0$ a.s. Therefore, it follows by T that, for any $\epsilon>0$, $P\left(\left|\sum_{i=2}^{n}\mathrm{E}\left[w_{in}\bar{y}_{in}|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}\right]\right|\ge\epsilon|\mathcal{Z}\right)\to 0$ a.s.

To finish showing that equation (B.9) is satisfied, it only remains to show that, for any $\epsilon>0$,

$$P\left(\left|\sum_{i=2}^{n}\left(\mathrm{E}\left[\bar{y}_{in}^2|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}\right]-\mathrm{E}[\bar{y}_{in}^2|]\mathcal{Z}\right)\right|\ge\epsilon|\mathcal{Z}\right)\to 0 \quad \text{a.s.} \quad \textbf{(B.11)}$$

Now, write

$$\sum_{i=2}^{n}\mathrm{E}\left(\left[\bar{y}_{in}^2|\mathcal{X}_1,\ldots,\mathcal{X}_{i-1},\mathcal{Z}\right]-\mathrm{E}[\bar{y}_{in}^2|\mathcal{Z}]\right)$$
$$=\sum_{j<i}\omega_i^2P_{ij}^2\left(\varepsilon_j^2-\sigma_j^2\right)\Big/K+2\sum_{j<k<i}\omega_i^2P_{ij}P_{ik}\varepsilon_j\varepsilon_k/K$$
$$+\sum_{j<i}\sigma_i^2P_{ij}^2\left(u_j^2-\omega_j^2\right)\Big/K+2\sum_{j<k<i}\sigma_i^2P_{ij}P_{ik}u_ju_k/K$$
$$+2\sum_{j<i}\gamma_iP_{ij}^2\left(u_j\varepsilon_j-\gamma_j\right)\Big/K+2\sum_{j<k<i}\gamma_iP_{ij}P_{ik}(u_j\varepsilon_k+u_k\varepsilon_j)/K. \quad \textbf{(B.12)}$$

By applying parts (i)–(iii) of Lemma B4 with $\phi_i=\gamma_i$, ω_i^2, and σ_i^2, respectively, we obtain, a.s., $\mathrm{E}\left[\left(\sum_{j<i}\gamma_iP_{ij}^2[u_j\varepsilon_j-\gamma_j]/K\right)^2|\mathcal{Z}\right]\to 0$, $\mathrm{E}\left[\left(\sum_{j<i}\omega_i^2P_{ij}^2[\varepsilon_j^2-\sigma_j^2]/K\right)^2|\,\mathcal{Z}\right]\to 0$, and $\mathrm{E}\left[\left(\sum_{j<i}\sigma_i^2\,P_{ij}^2[u_j^2-\omega_j^2]/K\right)^2|\,\mathcal{Z}\right]\to 0$. Moreover, applying part (iv) of Lemma B4 with $\phi_i=\gamma_i$, we obtain $\mathrm{E}\left[\left(\sum_{j<k<i}\gamma_iP_{ij}\;P_{ik}\left[u_j\varepsilon_k+u_k\varepsilon_i\right]/K\right)^2|\,\mathcal{Z}\right]\to 0$ a.s. $\mathbb{P}_{\mathcal{Z}}$. Similarly, conditional on $\mathcal{Z}$, all of the remaining terms in equation (B.12) converge in mean square to zero a.s. by parts (v) and (vi) of Lemma B4.

The preceding argument shows that as $n \to \infty$, $P(Y_n \le y|\mathcal{Z}) \to \Phi(y)$ a.s. $\mathbb{P}_{\mathcal{Z}}$, for every real number y, where $\Phi(y)$ denotes the cumulative distribution function of a standard normal distribution. Moreover, it is clear that, for some $\epsilon > 0$, $\sup_n \mathrm{E}\left[|P(Y_n \le y|\mathcal{Z})|^{1+\epsilon}\right] < \infty$ (take, e.g., $\epsilon = 1$). Hence, by a version of the dominated convergence theorem, as given by Theorem 25.12 of Billingsley (1986), we deduce that $P(Y_n \le y) = \mathrm{E}[P(Y_n \le y|\ \mathcal{Z})] \to \mathrm{E}[\Phi(y)] = \Phi(y)$, which gives the desired conclusion. ■

Proof of Lemma A4. Let $\bar{\omega}_i = \mathrm{E}[W_i|\mathcal{Z}]$, $\tilde{W}_i = W_i - \bar{\omega}_i$, $\bar{y}_i = \mathrm{E}[Y_i|\mathcal{Z}]$, $\tilde{Y}_i = Y_i - \bar{y}_i$, $\bar{\eta}_i = \mathrm{E}[\eta_i|\mathcal{Z}]$, $\tilde{\eta}_i = \eta_i - \bar{\eta}_i$,

$$\bar{\mu}_W^2 = \max_{i\le n} \bar{\omega}_i^2 \le C/n, \qquad \bar{\mu}_Y^2 = \max_{i\le n} \bar{y}_i^2 \le C/n, \qquad \bar{\mu}_\eta^2 = \max_{i\le n} \bar{\eta}_i^2 \le C,$$
$$\bar{\sigma}_W^2 = \max_{i\le n} \mathrm{Var}(W_i|\mathcal{Z}) \le C/r_n, \qquad \bar{\sigma}_Y^2 = \max_{i\le n} \mathrm{Var}(Y_i|\mathcal{Z}) \le C/r_n,$$
$$\bar{\sigma}_\eta^2 = \max_{i\le n} \mathrm{Var}(\eta_i|\mathcal{Z}) \le C.$$

Also, let $\breve{y}_i = \sum_j P_{ij}\bar{y}_j$, $\breve{\omega}_i = \sum_j P_{ij}\bar{\omega}_j$, be predicted values from projecting $\bar{y}$ and $\bar{\omega}$ on P and note that

$$\sum_i \breve{y}_i^2 \le \sum_i \bar{y}_i^2 \le C, \sum_i \breve{\omega}_i^2 \le \sum_i \bar{\omega}_i^2 \le C.$$

By adding and subtracting terms similar to the beginning of the proof of Theorem 4,

$$\begin{aligned} A_n &= \sum_{i\ne j}\sum_{k\notin\{i,j\}} \bar{\omega}_i P_{ik}\bar{\eta}_k P_{kj}\bar{y}_j \\ &= \sum_i \bar{\eta}_i\left(\breve{\omega}_i\breve{y}_i - P_{ii}\bar{\omega}_i\breve{y}_i - P_{ii}\breve{\omega}_i\bar{y}_i + 2P_{ii}^2\bar{\omega}_i\bar{y}_i\right)\Big/ n - \sum_{i,j}\bar{\omega}_i\bar{y}_i P_{ij}^2\bar{\eta}_j. \end{aligned}$$

By T, CS, and $\bar{\eta}_k \le C$,

$$\left|\sum_k \breve{\omega}_k\bar{\eta}_k\breve{y}_k\right| \le C\sqrt{\sum_k \breve{\omega}_k^2}\sqrt{\sum_k \breve{y}_k^2} \le C, \qquad \left|\sum_i \bar{\omega}_i P_{ii}\bar{\eta}_i\breve{y}_i\right| \le \sqrt{\sum_i \bar{\omega}_i^2 P_{ii}^2\bar{\eta}_i^2}\sqrt{\sum_i \breve{y}_i^2} \le C,$$

and it follows similarly that $\sum_i \breve{\omega}_i P_{ii}\bar{\eta}_i\bar{y}_i$ is bounded. By Lemma B1, $\left|\sum_{i,k}\bar{\omega}_i\bar{y}_i P_{ik}^2\bar{\eta}_k\right| \le Cn^{-1}\left|\sum_{i,k} P_{ik}^2\right| \le CK/n \le C$. Also, $\left|\sum_i \bar{\omega}_i\bar{y}_i P_{ii}^2\bar{\eta}_i\right| \le Cn/n = C$. Thus, $|A_n| \le C$ holds by T.

For the remainder of this proof we let $\mathrm{E}[\bullet]$ denote the conditional expectation given $\mathcal{Z}$. Note that

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

Summing and subtracting the last term gives

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

where

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

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

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 for when each of the three indexes i,j,k is equal to one of the three indexes r,s,t. There are six ways this can happen, leading to six terms in

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

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

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

Similarly, by CS,

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

Next, by Lemma B1 and CS

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

Similarly,

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

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

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

T then gives $\mathrm{E}[\hat{\psi}_1^2] \to 0$, 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 $\mathrm{E}[\tilde{W}_i P_{ik}\tilde{\eta}_k P_{kj}\bar{y}_j \tilde{W}_r P_{rs}\tilde{\eta}_s P_{st}\bar{y}_t] = 0$, except when $i = r$ and $j = s$ or $i = s$ and $j = r$. Then by $(A + B + C)^2 \leq 3(A^2 + B^2 + C^2)$ and for fixed k, $\sum_{i\neq k} P_{ik}^2 \leq P_{kk}$, $\sum_{i\neq k} P_{ik}^4 \leq P_{kk}$, it follows that

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

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

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

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 $\mathrm{E}[\tilde{W}_i P_{ik}\bar{\eta}_k P_{kj}\tilde{Y}_j \tilde{W}_r P_{rs}\bar{\eta}_s P_{st}\tilde{Y}_t]=0$, except when $i=r$ and $j=t$ or $i=t$ and $j=r$. Thus,

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

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

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

It therefore follows that

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

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

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

From the previous expression for $\mathrm{E}[\hat{\psi}_3^2]$, we then have $\mathrm{E}[\hat{\psi}_3^2] \le C\bar{\sigma}_W^2\bar{\sigma}_Y^2\bar{\mu}_\eta^2 K \le Cr_n^{-2}K \to 0$. Then $\hat{\psi}_3 \xrightarrow{p} 0$ by M.

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

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

Note that for $i \neq j$,

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

Therefore, for fixed i,

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

Note that because P is idempotent, we have $\sum_j \sum_k P_{jk} \bar{\eta}_j \breve{y}_j \bar{\eta}_k \breve{y}_k \leq \sum_j \bar{\eta}_j^2 \breve{y}_j^2 \leq \bar{\mu}_\eta^2 \sum_j \breve{y}_j^2 \leq \bar{\mu}_\eta^2 \sum_j \bar{y}_j^2 \leq n \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \leq C$. Then it follows that

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

Also, using similar reasoning,

$$\sum_i (P_{ii} \bar{\eta}_i \breve{y}_i)^2 \leq \sum_i \bar{\eta}_i^2 \breve{y}_i^2 \leq n \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \leq C,$$

$$\sum_i \left(\sum_j P_{ij} \bar{\eta}_j P_{jj} \bar{y}_j \right)^2 \leq \sum_i \bar{\eta}_i^2 P_{ii}^2 \bar{y}_i^2 \leq \sum_i \bar{\eta}_i^2 \bar{y}_i^2 \leq C,$$

$$\sum_i \left(\bar{y}_i \sum_k P_{ik}^2 \bar{\eta}_k \right)^2 \leq \bar{\mu}_Y^2 \sum_{i,k,\ell} P_{ik}^2 P_{i\ell}^2 \bar{\eta}_k \bar{\eta}_\ell \leq \bar{\mu}_Y^2 \bar{\mu}_\eta^2 \sum_{i,k,\ell} P_{ik}^2 P_{i\ell}^2 \leq K \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \leq C,$$

$$\sum_i P_{ii}^4 \bar{\eta}_i^2 \bar{y}_i^2 \leq n \bar{\mu}_\eta^2 \bar{\mu}_Y^2 \leq C.$$

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

Next, consider $\hat{\psi}_6$. Note that for $i \neq k$, $\sum_{j \notin \{i,k\}} \bar{\omega}_i P_{ik} P_{kj} \bar{y}_j = \bar{\omega}_i P_{ik} \breve{y}_k - \bar{\omega}_i P_{ik}^2 \bar{y}_i - \bar{\omega}_i P_{ik} P_{kk} \bar{y}_k$. Then for fixed k,

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

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

$$\mathrm{E}[\hat{\psi}_6^2] = \sum_k \mathrm{E}[\tilde{\eta}_k^2](\sum_{i\neq k}\sum_{j\notin\{i,k\}} \bar{\varpi}_i P_{ik} P_{kj} \bar{y}_j)^2$$

$$\leq 5\bar{\sigma}_\eta^2 \sum_k \left(\breve{\varpi}_k^2 \breve{y}_k^2 + \sum_{i,j} P_{kj}^2 P_{ki}^2 \bar{\varpi}_i \bar{y}_i \bar{\varpi}_j \bar{y}_j + \breve{\varpi}_k^2 P_{kk}^2 \bar{y}_k^2 + \bar{\varpi}_k^2 P_{kk}^2 \breve{y}_k^2 + 4\bar{\varpi}_k^2 P_{kk}^4 \bar{y}_k^2 \right)$$

$$\leq 5\bar{\sigma}_\eta^2 \left(\sum_k \breve{\varpi}_k^2 \breve{y}_k^2 + \bar{\mu}_W^2 \bar{\mu}_Y^2 \sum_{i,j,k} P_{kj}^2 P_{ki}^2 + \bar{\mu}_Y^2 \sum_k \breve{\varpi}_k^2 + \bar{\mu}_W^2 \sum_k \breve{y}_k^2 + n4\bar{\mu}_W^2 \bar{\mu}_Y^2 \right)$$

$$\leq 5\bar{\sigma}_\eta^2 \left(\sum_k \breve{\varpi}_k^2 \breve{y}_k^2 + 7n\bar{\mu}_W^2 \bar{\mu}_Y^2 \right) \leq C\sum_k \breve{\varpi}_k^2 \breve{y}_k^2 + Cn/n^2 \leq C\sum_k \breve{\varpi}_k^2 \breve{y}_k^2 + o(1).$$

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

$$|\bar{\varpi}_i - \breve{\varpi}_i| = \left|\bar{\varpi}_i - Z_i'(Z'Z)^{-1}Z'\bar{\varpi}\right| = \left|\bar{\varpi}_i - Z_i'\alpha_n - Z_i'(Z'Z)^{-1}Z'(\bar{\varpi} - Z\alpha_n)\right|$$

$$\leq \Delta_n/\sqrt{n} + \left(\sum_j P_{ij}^2\right)^{1/2} \left(\sum_j \left[\bar{\varpi}_j - Z_j'\alpha_n\right]^2\right)^{1/2}$$

$$\leq \Delta_n + P_{ii}^{1/2}\sqrt{n}\max_{i\leq n}\left|\bar{\varpi}_i - Z_i'\alpha_n\right| = \Delta_n + P_{ii}^{1/2}\Delta_n \leq C\Delta_n.$$

Then by T, $\max_{i\leq n}|\breve{\varpi}_i| \leq \max_{i\leq n}|\bar{\varpi}_i| + \Delta_n \to 0$, so that

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

Then we have $\mathrm{E}[\hat{\psi}_6^2] \to 0$, so by M, $\hat{\psi}_6 \xrightarrow{p} 0$. The conclusion then follows by T. ■