The random matrix machinery in many instruments asymptotics.

Pavel Yaskov.

PRELIMINARY AND INCOMPLETE.

March 28, 2014

Abstract

This paper proposes a test of a key condition on the instrument projection matrix in the Bekker (1994) framework of an instrumental variables regression with many instruments. The validity of this condition has two consequences. First, it implies that the limited-information maximum likelihood (LIML) estimator is optimal in a broad class of estimators considered by Anderson et al.(2010). Second, asymptotic variances for many popular estimators (see Hausman et al.(2012), van Hasselt (2010)) have much simpler forms under this condition. The latter could be used to improve finite sample properties of tests. Another goal of the paper is to show how universality results from the random matrix theory could be used in econometrics.

1 Introduction

This paper contributes to the literature on many instruments in several directions. First, it resolves a recent question posed in Anderson et al.(2010), Kunitomo (2012), Anatolyev and Gospodinov (2011) and Anatolyev (2013) on the validity of the condition

$$\frac{1}{n}\sum_{i=1}^{n}(P_{ii}-\alpha)^2 \xrightarrow{p} 0 \quad \text{as } n\to\infty \text{ and } l/n\to\alpha\in[0,1), \tag{$*$}$$

where P_{ii} are diagonal elements of the instrument projection matrix $P_Z = Z(Z'Z)^{-1}Z'$ with a $p\times n$ random instrument matrix Z ($l < n$). This is a key condition implying the weak heteroscedasticity assumption considered by Anderson et al. (2010) and Kunitomo (2012). The latter, in turn, guarantees that LIML estimator is well-behaved, optimal in a certain sense (see Anderson et al. (2010)) and has the same asymptotic distribution as in the case of normal errors (see van Hasselt (2010) and Bekker (1994)). Condition ($*$) also gives a much simpler forms for asymptotic variances for the bias-corrected two stage least squares estimator (see van Hasselt (2010)), the jackknife LIML estimator and the heteroscedasticity robust Fuller estimator (see Hausman et al. (2012)). This could be used to improve finite sample properties of many standard tests including t-test and specification tests (see Okui and Lee (2012) and Anatolyev (2013)).

Second, the paper shows how to use universality results from the random matrix theory in econometrics.

The paper is structured as follows. Section 2 contains main results. Section 3 deals with examples and counterexamples. All proofs and auxiliary results are relegated to the Appendix.

2 Main results

Let Z be a $n\times l$ random matrix with IID rows $z_1',\dots,z_n'$ and $l\leqslant n$. Denote by $\lambda_{\min}(A)$ the smallest eigenvalue of a square matrix A. Since the object of our study is the orthogonal projector $P_Z = Z(Z'Z)^{-1}Z'$ associated with Z, we may assume that $Ez_iz_i' = I_l$ (after a proper normalization).

Assumption 1. *For any matrices A_l of the size $l\times l$ ($l=1,2,\dots$) and such that $\|A_l\|$ is uniformly bounded over l, $(z_1'A_lz_1 - \mathrm{tr}A_l)/l \xrightarrow{p} 0$ as $l\to\infty$.*

Assumption 1*. *For any matrices A_l of the size $l\times l$ ($l=1,2,\dots$) and such that $\|A_l\|$ is uniformly bounded over l, $P(|z_1'A_lz_1 - \mathrm{tr}A_l| > \delta l) = o(1/l)$ as $l\to\infty$ for all $\delta>0$.*

As far as we know, Assumption 1 is the most general assumption which implies that quantities like $\mathrm{tr}(Z'Z+\varepsilon nI_l)^{-1}$, $\varepsilon>0$, behave as if Z were a Gaussian matrix. The latter is called universality in the random matrix theory. The formal statement is given in the following Proposition.

Proposition 1. *Under Assumption 1,*

$$\mathrm{tr}(Z'Z+\varepsilon nI_l)^{-1} - \mathrm{tr}(W'W+\varepsilon nI_l)^{-1} \xrightarrow{p} 0, \quad n\to\infty,$$

for all $\varepsilon > 0$ and $l = O(n)$, where W is a $n \times l$ matrix with IID standard normal entries.

Proposition 1*. *Under Assumption 1*,*

$$\max_{1 \leqslant j \leqslant n} \left| \operatorname{tr}(Z_{-j}' Z_{-j} + \varepsilon n I_l)^{-1} - \operatorname{tr}(W'W + \varepsilon n I_l)^{-1} \right| \xrightarrow{p} 0, \quad n \to \infty,$$

for all $\varepsilon > 0$ and $l = \alpha n + o(n)$, where W is a $(n-1) \times l$ matrix with IID standard normal entries.

Proposition 2. *Let Assumption 1 hold and $P(\lambda_{\min}(Z'Z) > Cn) \to 1$ for some $C > 0$ as $n \to \infty$ and $l/n = \alpha + o(1)$ with $\alpha \in [0, 1)$. Then $n^{-1} \sum_{i=1}^n (P_{ii} - \alpha)^2 \xrightarrow{p} 0$ as $n \to \infty$ and $l/n \to \alpha$.*

Proposition 2*. *Let Assumption 1* hold and $P(\lambda_{\min}(Z'Z) > Cn) \to 1$ for some $C > 0$ as $n \to \infty$ and $l/n = \alpha + o(1)$ with $\alpha \in [0, 1)$. Then $\max_i |P_{ii} - \alpha| \xrightarrow{p} 0$ as $n \to \infty$ and $l/n \to \alpha$.*

Assumption that $\lambda_{\min}(Z'Z)/n$ is separated from zero with high probability is rather technical and is hard for theoretical verification. However, see Theorem* in Appendix B (cf. Yaskov(2013)).

Now we discuss a question how to test Condition (*) in practice. We need one more assumption.

Assumption 2. *For any matrices A_l of the size $l \times l$ ($l = 1, 2, \ldots$) and such that $\|A_l\|$ is uniformly bounded over l, $E|z_1' A_l z_1 - \operatorname{tr} A_l|^2 = O(l)$ as $l \to \infty$.*

Assumption 2 holds if instruments are linear combinations of weakly dependent factors (see Proposition 4 below).

Proposition 3. *Let Assumption 2 hold and $P(\lambda_{\min}(Z'Z) > Cn) \to 1$ for some $C > 0$ as $n \to \infty$ and $l/n = \alpha + o(1)$ with $\alpha \in [0, 1)$. Then $\sum_{i=1}^n |P_{ii} - l/n|^2 = O_p(1)$ as $n \to \infty$.*

As a result, we see that, under some reasonable assumptions, Condition $(*)$ reduces to

$$\sum_{i=1}^{n} |P_{ii} - l/n|^2 = O_p(1). \qquad (**)$$

In particular, if the instrument matrix Z is a Gaussian random matrix then results[1] of Bai and Silverstein (2004) and the central limit theorem for quadratic forms imply that $\sum_{i=1}^n |P_{ii} - l/n|^2$ should have a certain asymptotic distribution. Therefore, the rule of thumb test of Condition $(**)$ against $\sum_{i=1}^n |P_{ii} - l/n|^2 \xrightarrow{p} \infty$ could have the following form. If $\sum_{i=1}^n |P_{ii} - l/n|^2 > q$ then put the validity of Condition $(**)$ in question; here q is a certain (e.g., 0.01) quantile of the random variable $\sum_{i=1}^n |P_{ii} - l/n|^2$ in the case of jointly normal instruments (q could be found by simulations). Formal proofs of these assertions are rather technical and are postponed for the future research.

Examples and counterexamples.

Example 1. Instruments that are sums of weakly dependent random variables.

[1] Namely, central limit theorem for Stieltjes transform in Bai and Silverstein (2004), Lemma 1.1.

Suppose $z = \Gamma\varepsilon$, where Γ is a non-random $l \times \infty$ matrix and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots)$ is a random sequence which components are orthonormal and weak dependent in a way that

$$|cov(\varepsilon_i^2, \varepsilon_j^2)| \leqslant \varphi_{j-i} \quad \text{and} \quad |E\varepsilon_i\varepsilon_j\varepsilon_p\varepsilon_q| \leqslant \min\{\varphi_{j-i}, \varphi_{p-j}, \varphi_{q-p}\}, \quad i < j < p < q,$$

with φ_p decreasing to 0 as $p \to \infty$ and $\sum_{p\geqslant 1} p\varphi_p < \infty$. In particular, the last bounds take place if variables ε_n have bounded moments of order $2\delta > 4$ and are strongly mixing with mixing coefficients proportional to $\varphi_p^{(\delta-2)/\delta}$.

Proposition 4. *If $z = \Gamma\varepsilon$ and $Ezz' = I_l$, then, for any $a \in \mathbb{R}^l$ and all positive-semidefinite symmetric matrices A of size $l \times l$,*

$$E|z'a|^4 \leqslant K|a'a|^2 \quad \textit{and} \quad E|z'Az - \mathrm{tr}A|^2 \leqslant C\mathrm{tr}A^2$$

for some $C, K > 0$ depending on φ.

Proposition 4 implies that Assumption 1 (with other assumptions of Theorem*) holds. Propositions 3 and 4 imply that $n^{-1}\sum_{i=1}^n |P_{ii} - \alpha|^2 \xrightarrow{p} 0$ for each fixed i and the given structure of instruments.

Proposition 4 also implies that

$$P(|z_1'A_lz_1 - \mathrm{tr}A_l| > \delta l) \leqslant \frac{E|z_1'A_lz_1 - \mathrm{tr}A_l|^2}{\delta^2 l^2} \leqslant \frac{C\|A_l\|^2}{\delta^2 l} = O(1/l)$$

as $l \to \infty$ for all $\delta > 0$ and all symmetric $(l \times l)$-matrices A_l such that $\|A_l\|$ is uniformly bounded over l. The latter is, of course, not enough but close to Assumption 1^*.

Example 2. Instruments that are sums of independent random variables.

Let, in Example 1, $\varepsilon_1, \varepsilon_2, \ldots$ be independent random variables with zero mean and unit variance. Assume also that $\sup_i E\varepsilon_i^p = \nu < \infty$ for some $p > 2$. Therefore, by Lemma B.26 in Bai& Silverstein (2011),

$$E|z'Az - \mathrm{tr}A|^p \leqslant C(|\mathrm{tr}A^2|^{p/2} + \mathrm{tr}A^{2p}) \leqslant C\|A_l\|^p(l + l^{p/2})$$

for any symmetric $(l \times l)$-matrix and some $C > 0$ depending only on p and ν. This bound guarantees that Assumption 1^* holds since

$$P(|z_1'A_lz_1 - \mathrm{tr}A_l| > \delta l) \leqslant \frac{E|z_1'A_lz_1 - \mathrm{tr}A_l|^p}{\delta^p l^p} \leqslant \frac{C\|A_l\|^p(l + l^{p/2})}{\delta^p l^p} = o(1/l)$$

as $l \to \infty$ for all $\delta > 0$ and all symmetric $(l \times l)$-matrices A_l such that $\|A_l\|$ is uniformly bounded over l.

Using Proposition 4 (for independent ε_i) we get that

$$\sup_{a\in\mathbb{R}^l:\, a'a=1} E|z_1'a|^4 \leqslant K$$

for some $K > 0$ not depending on l. Thus Theorem* and Proposition 2^* hold. As a result, $\max_i |P_{ii} - \alpha| \to 0$.

Example 3. Instruments that are weakly dependent.

This reduces to Example 1 with $\varepsilon_i = z_i$, $i = 1, \ldots, n$.

Counterexample 1. Instruments interacted with dummy variable.

Suppose $z = du$, where $d \in \{0, 1\}$ is a dummy variable with $P(d = 0) = P(d = 1) = 1/2$ and $u = \Gamma\varepsilon$ with ε defined in Example 1 and $Euu' = I_l$ (here we allow $Ezz' \neq I_l$).

Let calculate the limit of P_{ii} in this case. First we note that, by CLT, $\sum_{i=1}^n d_i = n/2 + O_P(\sqrt{n})$. Hence, the rank of the matrix

$$Z'Z = \sum_{i=1}^n d_i u_i u_i'$$

is non greater than $n/2 + O_P(\sqrt{n})$ and it size is $l \times l$. Therefore, $Z'Z$ is degenerate with large probability under the scheme $l/n = \alpha + o(1)$ with $\alpha > 1/2$. So let us suppose that $\alpha < 1/2$.

Proposition 5. *Under given assumptions, if d is independent of u, then $|P_{ii} - 2\alpha d_i| \xrightarrow{p} 0$ for any fixed i.*

By Proposition 5,

$$E\Big|\frac{1}{n}\sum_{i=1}^n P_{ii}^2 - \frac{1}{n}\sum_{i=1}^n (2\alpha d_i)^2\Big| \leqslant \frac{1}{n}\sum_{i=1}^n E|P_{ii}^2 - (2\alpha d_i)^2| = E|P_{11}^2 - (2\alpha d_1)^2| \leqslant (1+2\alpha)E|P_{11} - 2\alpha d_1| \to 0.$$

In addition, the law of large numbers implies that

$$\frac{1}{n}\sum_{i=1}^n (2\alpha d_i)^2 \xrightarrow{p} E(2\alpha d_1)^2 = 2\alpha^2.$$

Hence, $n^{-1}\sum_{i=1}^n P_{ii}^2 \xrightarrow{p} 2\alpha^2$.

Counterexample 2.

Suppose $z = (1, v, v^2, v^3, v^4, vd_1, vd_2, \ldots, vd_{l-5})$, where $d_j \in \{0, 1\}$ are IID dummy variables with $P(d_j = 1) = 1/2$ and $v \sim \mathcal{N}(0, 1)$ does not depend on $(d_j)_{j=1}^\infty$. As was argued by Hausman et al.(2012), P_{ii} could not be asymptotically constant in this case.

References.

Anatolyev, S. (2012). Inference in regression models with many regressors. Journal of Econometrics 170, 368-382.

Anatolyev S. (2013), Instrumental variables estimation and inference in the presence of many exogenous regressors. The Econometrics journal 16(1), 27-72.

Anatolyev S. and A. Abutaliev (2013), Asymptotic variance under many instruments: Numerical computations. Economics Letters 118, 272-274.

Anatolyev, S. and N. Gospodinov (2011). Specification testing in models with many instruments. Econometric Theory 27, 427-441.

Anderson, T.W., N. Kunimoto and Y. Matsushita (2010). On the asymptotic optimality of the LIML estimator with possibly many instruments. Journal of Econometrics, 157, 191-204.

Bai, Z. D. and J. W. Silverstein (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. Ann. Probab. 32, 553-605.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. Econometrica, 54, 657-682.

Hansen, C., J. Hausman and W. K. Newey (2008). Estimation with many instrumental variables. Journal of Business and Economics Statistics 26, 398-422.

Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012). Instrumental variable estimation with heteroskedasticity and many instruments. Quantitative Economics 3, 211-255.

Kunitomo, N. (2012). An optimal modification of the liml estimation for many instruments and persistent heteroskedasticity. Annals of the Institute of Statistical Mathematics 64, 881-910.

Lee, Y. and R. Okui (2012). Hahn-Hausman test as a specification test. Journal of Econometrics 167, 133-139.

Najim, J. and J. Yao (2014). Gaussian fluctuations for linear spectral statistics of large random covariance matrices. Unpublished manuscript.

van Hasselt, M. (2010). Many instruments asymptotic approximations under nonnormal error distributions. Econometric Theory 26, 633-645.

Yaskov, P.A. (2013). On the behaviour of the smallest eigenvalue of a high-dimensional sample covariance matrix. Russian mathematical surveys 68:4, 187-188.

Appendix A.

Theorem 0. *If $\lambda_{\min}(Z'Z)/\sqrt{n} \xrightarrow{p} \infty$ as $n \to \infty$ for some given $l = l(n)$, then*

$$\frac{1}{n}\sum_{i=1}^{n} f(P_{ii}) - Ef(P_{11}) \xrightarrow{p} 0$$

for any continuous function f, where P_{ii}, $i = 1, \dots, n$, are diagonal elements of P_Z.

Proof of Theorem 0. Any continuous function on $[0, 1]$ could be approximated by a smooth function. Therefore, we may consider only smooth functions f. The rest of the proof consists in the verification of several claims.

Claim 1. There are λ_n such that $\lambda_n \xrightarrow{p} \infty$ and $n^{-1} \sum_{i=1}^n [f(P_{ii}) - f_i] \xrightarrow{p} 0$, where $f_i = f(z_i'(Z'Z + \lambda_n I_l)^{-1} z_i)$.

Since $\lambda_{\min}(Z'Z) \xrightarrow{p} \infty$, there are λ_n that grow to infinity slower than $\lambda_{\min}(Z'Z)$ (i.e. $\lambda_n/\lambda_{\min}(Z'Z) \xrightarrow{p} 0$). Using the formula $P_{ii} = z_i'(Z'Z)^{-1} z_i$, the smoothness of f and the inequality

$$|z_i'(Z'Z)^{-1} z_i - z_i'(Z'Z + \lambda_n I_l)^{-1} z_i| = \lambda_n |z_i'(Z'Z)^{-1}(Z'Z + \lambda_n I_l)^{-1} z_i| \leqslant \lambda_n/\lambda_{\min}(Z'Z),$$

we prove Claim 1.

Claim 2. $n^{-1} \sum_{i=1}^n [f_i - E_{-i} f_i] \xrightarrow{p} 0$.

Since $|f_i|$ is bounded, we have

$$E\Big|\frac{1}{n} \sum_{i=1}^n [f_i - E_{-i} f_i]\Big|^2 = O(n^{-1}) + O(1) \cdot E[f_1 - E_{-1} f_1][f_2 - E_{-2} f_2].$$

Hence, we only need to show that $E[f_1 - E_{-1} f_1][f_2 - E_{-2} f_2] = o(1)$.

Recall the Sherman-Morrison-Woodbury formula (SMW)

$$(A + uu')^{-1} = A^{-1} - \frac{A^{-1} uu' A^{-1}}{1 + u' A^{-1} u}.$$

By the SMW formula,

$$z_i'(Z'Z + \lambda_n I_l)^{-1} z_i = g(z_i(Z_{-i}' Z_{-i} + \lambda_n I_l)^{-1} z_i)$$

with $g(x) = x/(1 + x)$, $x \geqslant 0$. In addition, the function $h(x) = f(g(x))$ is second-order smooth on $\mathbb{R}_+$ and there is $C_0 > 0$ such that $|h^{(k)}(x)|^2 \leqslant C_0$ on $\mathbb{R}_+$ for each $k = 0, 1$. Put $f_{ij} = h(z_i'(Z_{-ij}' Z_{-ij} + \lambda_n I_l)^{-1} z_i)$ for $i \neq j$. Since

$$E[f_{12} - E_{-12} f_{12}][f_{21} - E_{-12} f_{12}] = E\big[E_{-12}[f_{12} - E_{-12} f_{12}][f_{21} - E_{-12} f_{21}]\big] = 0$$

and $E_{-1} f_{12} = E_{-12} f_{12} = E_{-12} f_{21} = E_{-2} f_{21}$, the equality $E[f_1 - E_{-1} f_1][f_2 - E_{-2} f_2] = o(1)$ (as well as Claim 2) follows from Claim 3 below.

Claim 3. $E|f_i - f_{ij}| \to 0$ and $E|E_{-i} f_i - E_{-i} f_{ij}| \to 0$ for any fixed i, j, $i \neq j$.

The SMW formula yields

$$\Delta_{ij} = z_i'[(Z_{-i}'Z_{-i} + \lambda_n I_l)^{-1} - (Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-1}]z_i = \frac{|z_i'(Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-1} z_j|^2}{1 + z_j'(Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-1} z_j}.$$

If $|\Delta_{ij}| \leqslant 1$, then $|f_i - f_{ij}| \leqslant C_0\, |\Delta_{ij}|$. else if $|\Delta_{ij}| > 1$, then $|f_i - f_{ij}| \leqslant 2C_0$. By conditional Jensen's inequality,

$$E|E_{-i}(f_i - f_{ij})| \leqslant E|f_i - f_{ij}| \leqslant 2C_0\, E \min\{|\Delta_{ij}|, 1\}$$

and

$$E \min\{|\Delta_{ij}|, 1\} = EE_{-i} \min\{|\Delta_{ij}|, 1\} \leqslant E \min\{E_{-i}|\Delta_{ij}|, 1\}.$$

It follows from the inequality $E_{-i} z_i z_i' = I_l$ that

$$E_{-i}|\Delta_{ij}| = E_{-i} \frac{z_j'(Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-1} z_i z_i' (Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-1} z_j}{1 + z_j'(Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-1} z_j} =$$

$$= \frac{z_j'(Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-2} z_j}{1 + z_j'(Z_{-ij}'Z_{-ij} + \lambda_n I_l)^{-1} z_j} \leqslant \frac{1}{\lambda_n} = o(1).$$

Hence, Claim 3 is obtained.

Claim 4. $E|n^{-1}\sum_{i=1}^n E_{-i}f_i - E_{-1}f_1| \to 0$.

Using that $|f_i| \leqslant C_0$ and $E_{-1}f_{12} = E_{-12}f_{12} = E_{-12}f_{21} = E_{-2}f_{21}$, we derive from Claim 3 that

$$E\Big|\frac{1}{n}\sum_{i=1}^n E_{-i}f_i - E_{-1}f_1\Big| \leqslant \frac{1}{n}\sum_{i=1}^n E|E_{-i}f_i - E_{-1}f_1| \leqslant E|E_{-1}f_1 - E_{-2}f_2| =$$

$$= E|E_{-1}f_1 - E_{-1}f_{12} + E_{-2}f_{21} - E_{-2}f_2| \leqslant E|E_{-1}f_1 - E_{-1}f_{12}| + E|E_{-2}f_21 - E_{-2}f_2| = o(1).$$

Thus, Claim 4 is proven.

Claim 5. If $\lambda_{\min}(Z'Z) \xrightarrow{p} \infty$, then $n^{-1}\sum_{i=1}^n f(P_{ii}) - E_{-1}f(P_{11}) \xrightarrow{p} 0$.

This follows from Claims 1-4.

Claim 6. $E|E_{-1}f_1 - Ef_1|^2 \to 0$.

To prove Claim 6 we need the assumption $\lambda_{\min}(Z'Z)/\sqrt{n} \xrightarrow{p} \infty$. Going back to the definition of λ_n in the proof of Claim 1 we can initially take λ_n such that λ_n growing faster than $\sqrt{n}$ and slower than $\lambda_{\min}(Z'Z)$ (i.e. $\lambda_n/\lambda_{\min}(Z'Z) \xrightarrow{p} 0$). Let $E_i = E(\,\cdot|z_2, \ldots, z_i)$ and

$E_1 = E$. Using that $E_i(E_{-1}f_{1i}) = E_{i-1}(E_{-1}f_{1i})$ we represent $E_{-1}f_1 - Ef_1$ as the sum of martingale differences

$$E_{-1}f_1 - Ef_1 = \sum_{i=2}^{n}(E_i - E_{i-1})E_{-1}f_1 = \sum_{i=2}^{n}(E_i - E_{i-1})E_{-1}(f_1 - f_{1i}),$$

where, by the SMW formula and the inequalities given in the proof of Claim 3,

$$|E_{-1}(f_1 - f_{1i})| \leqslant E_{-1}|f_1 - f_{1i}| \leqslant 2C_0 E_{-1}\min\{|\Delta_{1i}|, 1\} \leqslant 2C_0 \min\{E_{-1}|\Delta_{1i}|, 1\} \leqslant \frac{2C_0}{\lambda_n}.$$

Claim 6 now follows from

$$E|E_{-1}f_1 - Ef_1|^2 = \sum_{i=2}^{n} E|(E_i - E_{i-1})E_{-1}(f_1 - f_{1i})|^2 \leqslant \frac{4C_0^2 n}{\lambda_n^2} = o(1).$$

Proposition*. *If $\lambda_{\min}(Z'Z) \xrightarrow{p} \infty$ as $n \to \infty$ and $l/n \xrightarrow{p} \alpha \in (0,1)$, then the following assumptions are equivalent*

1. $n^{-1}\sum_{j=1}^{n} P_{jj}^2 \xrightarrow{p} \alpha^2$,
2. $P_{ii} \xrightarrow{p} \alpha$ *for each fixed* i,
3. $P_{ii} = \dfrac{z_i'(Z_{-i}'Z_{-i})^{-1}z_i}{1 + z_i'(Z_{-i}'Z_{-i})^{-1}z_i}$ *does not asymptotically depend on* z_i *for each fixed* i.[2]
4. $n^{-1}\sum_{j=1}^{n}(P_{jj} - \alpha)^2 \xrightarrow{p} \alpha^2$,

Proof of Proposition*. W.l.o.g. we consider $i = 1$. By Claim 5 in the proof of Theorem 1, if $\lambda_{\min}(Z'Z) \xrightarrow{p} \infty$, then

$$\frac{l}{n} - E_{-1}P_{11} = \frac{1}{n}\sum_{i=1}^{n} P_{ii} - E_{-1}P_{11} \xrightarrow{p} 0 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} P_{ii}^2 - E_{-1}P_{11}^2 \xrightarrow{p} 0.$$

In particular, $E_{-1}P_{11} \xrightarrow{p} \alpha$. Suppose 1 holds. Then

$$E_{-1}[P_{11} - E_{-1}P_{11}]^2 = E_{-1}P_{11}^2 - [E_{-1}P_{11}]^2 \xrightarrow{p} 0$$

and $EE_{-1}[P_{11} - E_{-1}P_{11}]^2 = E[P_{11} - E_{-1}P_{11}]^2 \to 0$. The latter yields 2 since $E_{-1}P_{11} \xrightarrow{p} \alpha$. Obviously, 2 implies 3 and we only need to show that 3 implies 1. Assume that 3 holds, i.e. $z_1'(Z_{-1}'Z_{-1})^{-1}z_1 - f_n(Z_{-1}) \xrightarrow{p} 0$ for some nonnegative functions f_n. Then

$$P_{11} - \frac{f_n(Z_{-1})}{1 + f_n(Z_{-1})} = \frac{z_1'(Z_{-1}'Z_{-1})^{-1}z_1}{1 + z_1'(Z_{-1}'Z_{-1})^{-1}z_1} - \frac{f_n(Z_{-1})}{1 + f_n(Z_{-1})} \xrightarrow{p} 0$$

[2]By this, we mean that $z_i'(Z_{-i}'Z_{-i})^{-1}z_i - f_n(Z_{-i}) \xrightarrow{p} 0$ for each fixed i and some functions f_n.

and, by conditional Jensen's inequality,

$$E\Big|E_{-1}P_{11}-\frac{f_n(Z_{-1})}{1+f_n(Z_{-1})}\Big|\leqslant E\Big|P_{11}-\frac{f_n(Z_{-1})}{1+f_n(Z_{-1})}\Big|\to 0,$$

$$E\Big|E_{-1}P_{11}^2-\frac{f_n(Z_{-1})^2}{[1+f_n(Z_{-1})]^2}\Big|\leqslant E\Big|P_{11}^2-\frac{f_n(Z_{-1})^2}{[1+f_n(Z_{-1})]^2}\Big|\to 0.$$

Therefore, we conclude that $E_{-1}P_{11}^2-[E_{-1}P_{11}]^2\xrightarrow{p}0$ and $E_{-1}P_{11}^2\xrightarrow{p}\alpha^2$. Thus, we get 1.

Finally, 1 is obviously equivalent to 4 since $n^{-1}\sum_{i=1}^n iP_{ii}=l/n\to\alpha$. Q.e.d.

Proof of Proposition 1. Here we proof a little more general version, i.e.

$$\operatorname{tr}\Big(\sum_{i=1}^n d_iz_iz_i'+\varepsilon nI_l\Big)^{-1}-\operatorname{tr}\Big(\sum_{i=1}^n d_iw_iw_i'+\varepsilon nI_l\Big)^{-1}\xrightarrow{p}0,\quad n\to\infty,$$

for all $\varepsilon>0$ and $l=O(n)$, where (d_i,z_i) are IID, d_i are a bounded non-negative scalar random variable, and w_i are IID $l\times 1$ standard normal vectors independent of everything else.

Using so called Lindeberg's method and the Sherman-Morrison-Woodbury formula (see Claim 2 in the proof of Theorem 1) we get

$$\begin{aligned}
&\Big|\operatorname{tr}(\sum_{i=1}^n d_iz_iz_i'+\varepsilon nI_l\Big)^{-1}-\operatorname{tr}\Big(\sum_{i=1}^n d_iw_iw_i'+\varepsilon nI_l\Big)^{-1}\Big|\leqslant\\
&\qquad\leqslant\frac{1}{n}\sum_{k=1}^n\Big|\operatorname{tr}\Big(C_k+d_k\frac{z_kz_k'}{n}+\varepsilon I_l\Big)^{-1}-\operatorname{tr}\Big(C_k+\varepsilon I_l\Big)^{-1}+\\
&\qquad\qquad+\operatorname{tr}\Big(C_k+\varepsilon I_l\Big)^{-1}-\operatorname{tr}\Big(C_k+d_k\frac{w_kw_k'}{n}+\varepsilon I_l\Big)^{-1}\Big|\\
&\qquad=\frac{1}{n}\sum_{k=1}^n\left|\frac{d_kz_k\big(C_k+\varepsilon I_l\big)^{-2}z_k/n}{1+d_kz_k\big(C_k+\varepsilon I_l\big)^{-1}z_k/n}-\frac{d_kw_k\big(C_k+\varepsilon I_l\big)^{-2}w_k/n}{1+d_kw_k\big(C_k+\varepsilon I_l\big)^{-1}w_k/n}\right|\\
&\qquad=\frac{1}{n}\sum_{k=1}^n|\Delta_k|
\end{aligned}$$

where $C_1=\sum_{i=2}^n d_iz_iz_i'/n$, $C_n=\sum_{i=1}^{n-1}d_iw_iw_i'/n$ and

$$C_k=\sum_{i=1}^{k-1}d_iw_iw_i'/n+\sum_{i=k+1}^n d_iz_iz_i'/n,\quad 1<k<n.$$

By exchangeability,

$$E\frac{1}{n}\sum_{k=1}^n|\Delta_k|=E|\Delta_1|$$

and we only need to show that $\Delta_1\xrightarrow{p}0$ (since $|\Delta_1|\leqslant 1/\varepsilon$).

Arguing as in the proof of Proposition 4 and using Assumption 1 (and its analog for Gaussian vectors w_k) we easily get for $j = 1, 2$

$$\frac{1}{n}[z_1\big(C_1+\varepsilon I_l\big)^{-j}z_1 - \mathrm{tr}\big(C_1+\varepsilon I_l\big)^{-j}] \xrightarrow{p} 0,$$

$$\frac{1}{n}[w_1\big(C_1+\varepsilon I_l\big)^{-j}w_1 - \mathrm{tr}\big(C_1+\varepsilon I_l\big)^{-j}] \xrightarrow{p} 0.$$

Finally we see that

$$\Delta_1 = \frac{d_1\big(C_1+\varepsilon I_l\big)^{-2}/n + o_P(1)}{1+d_1\mathrm{tr}\big(C_1+\varepsilon I_l\big)^{-1}/n + o_P(1)} - \frac{d_1\mathrm{tr}\big(C_1+\varepsilon I_l\big)^{-2}/n + o_P(1)}{1+d_1\mathrm{tr}\big(C_1+\varepsilon I_l\big)^{-1}/n + o_P(1)} \xrightarrow{p} 0.$$

Here the notation is the same as in the proof of Proposition 1 and $d_i \equiv 1$. Let

$$z_{i,j} = z_i I(i \neq j) \quad \text{and} \quad w_{i,j} = w_i I(i<j) + w_{i-1}I(i>j).$$

As in the proof of Proposition 1, we derive

$$\begin{aligned}\max_{1\leqslant j\leqslant n}\Big|\mathrm{tr}\Big(\sum_{i=1}^{n} z_{i,j}z'_{i,j}+\varepsilon n I_l\Big)^{-1} - \mathrm{tr}\Big(\sum_{i=1}^{n} w_{i,j}w'_{i,j}+\varepsilon n I_l\Big)^{-1}\Big| &\leqslant \\ \leqslant \frac{1}{n}\max_{1\leqslant j\leqslant n}\sum_{k=1}^{n}\Bigg|\frac{z_{k,j}\big(C_{k,j}+\varepsilon I_l\big)^{-2}z_{k,j}/n}{1+z_{k,j}\big(C_{k,j}+\varepsilon I_l\big)^{-1}z_{k,j}/n} &- \frac{w_{k,j}\big(C_{k,j}+\varepsilon I_l\big)^{-2}w_{k,j}/n}{1+w_{k,j}\big(C_{k,j}+\varepsilon I_l\big)^{-1}w_{k,j}/n}\Bigg| \\ \leqslant \frac{1}{n}\sum_{k=1}^{n}(\max_{1\leqslant j\leqslant n}|\Delta_{k,j}| + \max_{1\leqslant j\leqslant n}|D_{k,j}|)\end{aligned}$$

where $C_{1,j} = \sum_{i=2}^{n} z_{i,j}z'_{i,j}/n$, $C_{n,j} = \sum_{i=1}^{n-1} w_{i,j}w'_{i,j}/n$ and

$$C_{k,j} = \sum_{i=1}^{k-1} w_{i,j}w'_{i,j}/n + \sum_{i=k+1}^{n} z_{i,j}z'_{i,j}/n, \quad 1<k<n,$$

$$\Delta_{k,j} = \frac{z_{k,j}\big(C_{k,j}+\varepsilon I_l\big)^{-2}z_{k,j}/n}{1+z_{k,j}\big(C_{k,j}+\varepsilon I_l\big)^{-1}z_{k,j}/n} - \frac{\mathrm{tr}\big(C_{k,j}+\varepsilon I_l\big)^{-2}/n}{1+\mathrm{tr}\big(C_{k,j}+\varepsilon I_l\big)^{-1}/n}, \quad 1\leqslant k\leqslant n,$$

and $D_{k,j}$ defined as $\Delta_{k,j}$ with $z_{k,j}$ replaced by $w_{k,j}$. Now we only need to show that

$$\max_{1\leqslant k\leqslant n}(E\max_{1\leqslant j\leqslant n}|\Delta_{k,j}| + E\max_{1\leqslant j\leqslant n}|D_{k,j}|) \to 0.$$

Let us first prove that $\max_{1\leqslant k\leqslant n} E\max_{1\leqslant j\leqslant n}|\Delta_{k,j}| \to 0$. Since $|\Delta_{k,j}|$ is bounded (by 2ε), we have (for all $\delta > 0$)

$$\max_{1\leqslant k\leqslant n} E\max_{1\leqslant j\leqslant n}|\Delta_{k,j}| \leqslant \delta + 2\varepsilon\max_{1\leqslant k\leqslant n} P(\max_{1\leqslant j\leqslant n}|\Delta_{k,j}| > \delta).$$

Using inequality
$$\Big|\frac{x}{1+y}-\frac{x_0}{1+y_0}\Big|\leqslant|x-x_0|+\varepsilon^2|y-y_0|$$
for $0\leqslant x_0\leqslant\varepsilon^2$ and $y,y_0\geqslant 0$, we derive

$$\max_{1\leqslant k\leqslant n}P(\max_{1\leqslant j\leqslant n}|\Delta_{k,j}|>\delta)\leqslant n\max_{1\leqslant k,j\leqslant n}P(|\Delta_{k,j}|>\delta)\leqslant n\max_{1\leqslant k,j\leqslant n}\sum_{m=1}^{2}P(|M^m_{k,j}|>\delta_0),$$

where $\delta_0=\delta\min\{1,\varepsilon^{-2}\}/2$,

$$M^m_{k,j}=(z_{k,j}\big(C_{k,j}+\varepsilon I_l\big)^{-m}z_{k,j}-\mathrm{tr}\big(C_{k,j}+\varepsilon I_l\big)^{-m})/n,\quad m=1,2.$$

Moreover, by the law of iterated mathematical expectations and the independence of $z_{k,j}$ and $C_{k,j}$, for all k,j,m,

$$P(|M^m_{k,j}|>\delta_0)=EP(|z'_{k,j}A_lz_{k,j}-\mathrm{tr}A_l|>\delta_0 n)|_{A_l=(C_{k,j}+\varepsilon I_l)^{-m}}\leqslant S_n,$$

where $S_n=\sup P(|z_1'A_lz_1-\mathrm{tr}A_l|>\delta_0 n)$ is taken over all $(l\times l)$-matrices A_l with $\|A_l\|\leqslant\max\{\varepsilon^{-1},\varepsilon^{-2}\}$. By Assumption 1^*, $nS_n\to 0$ if $l=\alpha n+o(n)$ with $\alpha\in(0,1)$. Hence,

$$\max_{1\leqslant k\leqslant n}E\max_{1\leqslant j\leqslant n}|\Delta_{k,j}|\leqslant\delta+o(1)\quad\text{for all }\delta>0.$$

As a result, we get that $\max\limits_{1\leqslant k\leqslant n}E\max\limits_{1\leqslant j\leqslant n}|\Delta_{k,j}|\to 0$.

Now let us prove that $\max\limits_{1\leqslant k\leqslant n}E\max\limits_{1\leqslant j\leqslant n}|D_{k,j}|\to 0$. This could be done as above if we prove that the following version of Assumption 1^* holds:

> *For any matrices A_l of the size $l\times l$ $(l=1,2,\dots)$ and such that $\|A_l\|$ is uniformly bounded over l, $P(|w_1'A_lw_1-\mathrm{tr}A_l|>\delta l)=o(1/l)$ as $l\to\infty$ for all $\delta>0$.*

We may consider w.l.o.g. only symmetric matrices A_l. For each symmetric $(l\times l)$-matrix A_l, there is an orthonormal basis $e_1,\dots,e_l$ in $\mathbb{R}^l$ such that $A_l=\sum_{k=1}^l\lambda_ke_ke_k'$ and $w_1'A_lw_1=\sum_{k=1}^l\lambda_k(w_1'e_k)^2$. Noting that $Ew_1'A_lw_1=\mathrm{tr}A_l$ and $\{(w_1'e_k)^2\}_{k=1}^l$ are IID random variables distributed as $\xi\sim\chi^2_1$, we see that

$$\begin{aligned}P(|w_1'A_lw_1-\mathrm{tr}A_l|>\delta l)&\leqslant\frac{E|w_1'A_lw_1-\mathrm{tr}A_l|^4}{\delta^4l^4}=\frac{E|\sum\limits_{k=1}^{l}\lambda_k[(w_1'e_k)^2-E(w_1'e_k)^2]|^4}{\delta^4l^4}\\&\leqslant\frac{C\sum\limits_{k=1}^{l}\lambda_k^4E[(w_1'e_k)^2-E(w_1'e_k)^2]^4}{\delta^4l^4}\\&\leqslant C\|A_l\|^4E|\xi-E\xi|^4\cdot\frac{1}{l^3}=o(1/l)\end{aligned}$$

for some universal constant $C > 0$ (by the Marcinkiewicz-Zygmund inequality) when $\|A_l\|$ is uniformly bounded over l. Thus the version of Assumption 1* holds and

$$\max_{1\leqslant k\leqslant n} E \max_{1\leqslant j\leqslant n} |D_{k,j}| \to 0.$$

Q.e.d.

Proof of Proposition 2. W.l.o.g. we consider $i = 1$. To avoid technicalities we prove the proposition under the assumption that $P(\lambda_{\min}(Z'_{-1}Z_{-1}) > Cn) \to 1$ instead of $P(\lambda_{\min}(Z'Z) > Cn) \to 1$ (in the second case, one should proceed as in Claim 1 in the proof of Theorem 1).

Using that $P(\lambda_{\min}(Z'_{-1}Z_{-1}) > Cn) \to 1$ we obtain via Assumption 1

$$P(|z'_1 A_l z_1 - \mathrm{tr} A_l| > \varepsilon)\big|_{A_l=(Z'_{-1}Z_{-1})^{-1}} \xrightarrow{p} 1$$

for any $\varepsilon > 0$. Therefore, letting $w_1 = z_1(Z'_{-1}Z_{-1})^{-1}z_1$ we derive

$$P\big(|w_1 - \mathrm{tr}(Z'_{-1}Z_{-1})^{-1}| > \varepsilon\big) = EP(|z'_1 A_l z_1 - \mathrm{tr} A_l| > \varepsilon)\big|_{A_l=(Z'_{-1}Z_{-1})^{-1}} \to 1$$

for any $\varepsilon > 0$. Writing $z'_1(Z'Z)^{-1}z_1 = g(w_1)$ with $g(x) = x/(1 + x)$ as in Claim 2 in the proof of Theorem 0we get

$$z'_1(Z'Z)^{-1}z_1 - g(\mathrm{tr}(Z'_{-1}Z_{-1})^{-1}) \xrightarrow{p} 0 \quad \text{and} \quad E_{-1}P_{11} - g(\mathrm{tr}(Z'_{-1}Z_{-1})^{-1}) \xrightarrow{p} 0,$$

where $E_{-1} = E[\,\cdot\,|Z_{-1}]$. Now the desired result follows from Proposition* (see point 3). Q.e.d.

Proof of Proposition 2*. First note that

$$\max_i |P_{ii} - z'_i(Z'Z + \varepsilon n I_l)^{-1}z_i| \leqslant \varepsilon n/\lambda_{\min}(Z'Z)$$

(see Claim 1 in the proof of Theorem 1). Assumption 1* yields

$$\begin{aligned} P(\max_i |z'_i(Z'_{-i}Z_{-i} + \varepsilon n I_l)^{-1}z_i - \mathrm{tr}(Z'_{-i}Z_{-i} + \varepsilon n I_l)^{-1}| > \delta) \leqslant \\ &\leqslant nP(|z'_1(Z'_{-1}Z_{-1} + \varepsilon n I_l)^{-1}z_1 - \mathrm{tr}(Z'_{-1}Z_{-1} + \varepsilon n I_l)^{-1}| > \delta) \\ &= nEP(|z'_1 A_l z_1 - \mathrm{tr} A_l| > \delta n)\big|_{A_l=(Z'_{-1}Z_{-1}/n+\varepsilon I_l)^{-1}} \\ &\leqslant n \sup P(|z'_1 A_l z_1 - \mathrm{tr} A_l| > \delta n) \to 0 \end{aligned}$$

for any $\delta > 0$ and $l = \alpha n + o(n)$ with $\alpha \in (0, 1)$, where sup is taken over all A_l with $\|A_l\| \leqslant \varepsilon^{-1}$. Therefore we get

$$\max_i |z'_i(Z'_{-i}Z_{-i} + \varepsilon n I_l)^{-1}z_i - \mathrm{tr}(Z'_{-i}Z_{-i} + \varepsilon n I_l)^{-1}| \xrightarrow{p} 0.$$

Using Proposition 1*, we derive that

$$\max_i |z'_i(Z'_{-i}Z_{-i} + \varepsilon n I_l)^{-1}z_i - \mathrm{tr}(W'W + \varepsilon n I_l)^{-1}| \xrightarrow{p} 0$$

where W is $(n-1)\times l$ matrix with IID standard normal entries. Note that

$$|\mathrm{tr}(W'W+\varepsilon nI_l)^{-1}-\mathrm{tr}(W'W)^{-1}|=\varepsilon n\cdot\mathrm{tr}[(W'W+\varepsilon nI_l)^{-1}(W'W)^{-1}]\leqslant\frac{\varepsilon n^2}{\lambda_{\min}(W'W)^2}.$$

Moreover, we have $\mathrm{tr}(W'W)^{-1}\xrightarrow{p}\alpha/(1-\alpha)$ that could be verified as in the proof of Proposition 2 with z_i replaced by standard normal vectors w_i (in the proof, we established that $g(\mathrm{tr}(Z'_{-1}Z_{-1})^{-1})\to\alpha$ for $g(x)=x/(1+x)$) . Combining the above estimates we see that

$$\begin{aligned}\max_i|P_{ii}-\alpha|\leqslant&\frac{\varepsilon n}{\lambda_{\min}(Z'Z)}+\max_i|z_i'(Z'Z+\varepsilon nI_l)^{-1}z_i-\alpha|\\=&\frac{\varepsilon n}{\lambda_{\min}(Z'Z)}+\max_i\left|g(z_i'(Z'_{-i}Z_{-i}+\varepsilon nI_l)^{-1}z_i)-g\Big(\frac{\alpha}{1-\alpha}\Big)\right|\\\leqslant&\frac{\varepsilon n}{\lambda_{\min}(Z'Z)}+\max_i\left|z_i'(Z'_{-i}Z_{-i}+\varepsilon nI_l)^{-1}z_i-\frac{\alpha}{1-\alpha}\right|\\\leqslant&\frac{\varepsilon n}{\lambda_{\min}(Z'Z)}+\frac{\varepsilon n^2}{\lambda_{\min}(W'W)^2}+o_P(1).\end{aligned}$$

Since there is a constant $C>0$ such that

$$P(\lambda_{\min}(Z'Z)>Cn)\to 1,\quad P(\lambda_{\min}(W'W)>Cn)\to 1,$$

we have

$$\max_i|P_{ii}-\alpha|\leqslant\frac{\varepsilon}{C}+\frac{\varepsilon}{C^2}+o_p(1).$$

The latter holds for any $\varepsilon>0$. Hence, we get the desired convergence $\max_i|P_{ii}-\alpha|\xrightarrow{p}0$. Q.e.d.

Proof of Proposition 3. To be proven.

Proof of Proposition 4. Since $z=\Gamma\varepsilon$ and $I_l=Ezz'=\Gamma E\varepsilon\varepsilon'\Gamma'=\Gamma\Gamma'$, we obtain $z'a=\varepsilon'b$ with $b=\Gamma'a$ satisfying $b'b=a'\Gamma\Gamma'a=a'a$. Therefore, the first inequality

$$E|z'a|^4=E|\varepsilon'b|^4\leqslant K|b'b|^2=K|a'a|^2 \tag{$*$}$$

is contained in Theorem 3a of Gaposhkin(1972) (for some $K>0$ not depending of a). Let us verify the second inequality. Since $I_l=Ezz'=\Gamma E\varepsilon\varepsilon'\Gamma'=\Gamma\Gamma'$, putting $B=(b_{ij})=\Gamma'A\Gamma$ we have $\mathrm{tr}B=\mathrm{tr}(A\Gamma\Gamma')=\mathrm{tr}A$, $z'Az-\mathrm{tr}A=\varepsilon'B\varepsilon-\mathrm{tr}B$,

$$\mathrm{tr}B^2=\mathrm{tr}(\Gamma'A\Gamma\Gamma'A\Gamma)=\mathrm{tr}(A^2\Gamma\Gamma')=\mathrm{tr}A^2$$

and, by the inequality $(x+y)^2\leqslant 2x^2+2y^2$,

$$\begin{aligned}E|\varepsilon'B\varepsilon-\mathrm{tr}B|^2=&E\Big|\sum_i b_{ii}(\varepsilon_i^2-E\varepsilon_i^2)+\sum_{i\neq j}b_{ij}\varepsilon_i\varepsilon_j\Big|^2\\&\leqslant 2E\Big|\sum_i b_{ii}(\varepsilon_i^2-E\varepsilon_i^2)\Big|^2+2E\Big|2\sum_{i<j}b_{ij}\varepsilon_i\varepsilon_j\Big|^2.\end{aligned}$$

Applying the Cauchy-Schwartz inequality, we see that

$$
\begin{aligned}
E\Big|\sum_i b_{ii}(\varepsilon_i^2 - E\varepsilon_i^2)\Big|^2 \leqslant & \sum_i b_{ii}^2 Var(\varepsilon_i^2) + \sum_{i\neq j} b_{ii}b_{jj} cov(\varepsilon_i^2, \varepsilon_j^2) \\
\leqslant & \varphi(0)\sum_i b_{ii}^2 + \sum_{i\neq j} \frac{b_{ii}^2 + b_{jj}^2}{2}\varphi(|i-j|) \\
\leqslant & \varphi_0 \sum_i b_{ii}^2 + \sum_i b_{ii}^2 \sum_{j:j\neq i} \varphi_{|i-j|} \\
\leqslant & 2\mathrm{tr}B^2 \sum_p \varphi_p = 2\mathrm{tr}A^2 \sum_p \varphi_p.
\end{aligned}
$$

Let us now deal with the second term

$$
\begin{aligned}
E\Big|\sum_{i<j} b_{ij}\varepsilon_i\varepsilon_j\Big|^2 =& 4\sum_{i<j} b_{ij}^2 E\varepsilon_i^2\varepsilon_j^2 + 4\sum_{i<j<p} b_{ij}b_{ip}E\varepsilon_i^2\varepsilon_j\varepsilon_p + 4\sum_{i<j<p} b_{ij}b_{jp}E\varepsilon_i\varepsilon_j^2\varepsilon_p \\
&+ 4\sum_{i<p<j} b_{ij}b_{pj}E\varepsilon_i\varepsilon_j^2\varepsilon_p + 4\sum_{i<j<p<q} b_{ij}b_{pq}E\varepsilon_i\varepsilon_j\varepsilon_p\varepsilon_q \\
&+ 4\sum_{i<p<j<q} b_{ij}b_{pq}E\varepsilon_i\varepsilon_p\varepsilon_j\varepsilon_q + 4\sum_{i<p<q<j} b_{ij}b_{pq}E\varepsilon_i\varepsilon_p\varepsilon_q\varepsilon_j
\end{aligned}
$$

Let us step by step control all terms in the left hand side of the last inequality.

Control of $\sum_{i<j<p} b_{ij}b_{ip}E\varepsilon_i^2\varepsilon_j\varepsilon_p$.

By the Cauchy-Schwartz inequality and inequality $(*)$,

$$
2\sum_{i<j<p} b_{ij}b_{ip}E\varepsilon_i^2\varepsilon_j\varepsilon_p + \sum_{i<j} b_{ij}^2 E\varepsilon_i^2\varepsilon_j^2 = \sum_i E\varepsilon_i^2\Big(\sum_{j:j>i} b_{ij}\varepsilon_j\Big)^2 \leqslant
$$

$$
\leqslant \sum_i \sqrt{E\varepsilon_i^4}\Big[E\Big(\sum_{j:\,j>i} b_{ij}\varepsilon_j\Big)^4\Big]^{1/2} \leqslant M\sqrt{K}\sum_i\sum_{j:\,j>i} b_{ij}^2 \leqslant M\sqrt{K}\mathrm{tr}A^2,
$$

where $M = \sup\sqrt{E\varepsilon_i^4}$.

Control of $\sum_{i<j<p} b_{ij}b_{jp}E\varepsilon_i\varepsilon_j^2\varepsilon_p$.

By the Cauchy-Schwartz inequality and inequality $(*)$,

$$
\Big|2\sum_{i<j<p} b_{ij}b_{ip}E\varepsilon_i\varepsilon_j^2\varepsilon_p\Big| = \Big|\sum_j E\varepsilon_j^2\Big(\sum_{i:\,i<j} b_{ij}\varepsilon_i\Big)\Big(\sum_{p:\,p>j} b_{jp}\varepsilon_p\Big)\Big| \leqslant
$$

$$
\leqslant \sum_j \sqrt{E\varepsilon_j^4}\Big[E\Big(\sum_{i:\,i<j} b_{ij}\varepsilon_i\Big)^4\Big]^{1/4}\Big[E\Big(\sum_{p:\,p>j} b_{jp}\varepsilon_p\Big)^4\Big]^{1/4} \leqslant
$$

$$
\leqslant M\Big[\sum_j\Big[E\Big(\sum_{i:\,i<j} b_{ij}\varepsilon_i\Big)^4\Big]^{1/2}\Big]^{1/2}\Big[\sum_j E\Big(\sum_{p:\,p>j} b_{jp}\varepsilon_p\Big)^4\Big]^{1/2}\Big]^{1/2} \leqslant M\sqrt{K}\mathrm{tr}A^2.
$$

Control of $\sum_{i<p<j} b_{ij}b_{pj}E\varepsilon_i\varepsilon_p\varepsilon_j^2$.

This could be done similarly to $\sum_{i<j<p} b_{ij}b_{ip}E\varepsilon_i^2\varepsilon_j\varepsilon_p$.

Control of $\sum_{i<p<j<q} b_{ij}b_{pq}E\varepsilon_i\varepsilon_p\varepsilon_j\varepsilon_q$.

Since

$$|E\varepsilon_i\varepsilon_p\varepsilon_j\varepsilon_q| \leqslant \min\{\varphi_{p-i}, \varphi_{q-j}\}, \quad i<p<j<q,$$

we have

$$|\sum_{i<p<j<q} b_{ij}b_{pq}E\varepsilon_i\varepsilon_p\varepsilon_j\varepsilon_q\Big| \leqslant \sum_{i<p<j<q} |b_{ij}b_{pq}| \min\{\varphi_{p-i}, \varphi_{q-j}\} \leqslant \sqrt{I_1 I_2}$$

with

$$I_1 = \sum_{i<p<j<q} b_{ij}^2 \min\{\varphi_{p-i}, \varphi_{q-j}\}, \quad I_2 = \sum_{i<p<j<q} b_{pq}^2 \min\{\varphi_{p-i}, \varphi_{q-j}\}.$$

In addition,

$$\begin{aligned} I_1 &= \sum_{i<j} b_{ij}^2 \sum_{p:\, i<p<j} \Big[(p-i)\varphi_{p-i} + \sum_{q:\, q-j>p-i} \varphi_{q-j}\Big] \\ &\leqslant \sum_{i<j} b_{ij}^2 \Big[\sum_p p\varphi_p + \sum_q \sum_{q:\, q>p} \varphi_q\Big] \leqslant 2\mathrm{tr}A^2 \sum_p p\varphi_p. \end{aligned}$$

and similarly

$$I_2 = \sum_{p<q} b_{pq}^2 \sum_{j:\, p<j<q} \Big[(q-j)\varphi_{q-j} + \sum_{i:\, p-i>q-j} \varphi_{p-i}\Big] \leqslant 2\mathrm{tr}A^2 \sum_p p\varphi_p.$$

Control of $\sum_{i<p<q<j} b_{ij}b_{pq}E\varepsilon_i\varepsilon_p\varepsilon_q\varepsilon_j$.

Since

$$|E\varepsilon_i\varepsilon_p\varepsilon_q\varepsilon_j| \leqslant \min\{\varphi_{p-i}, \varphi_{j-q}\}, \quad i<p<q<j,$$

we have

$$|\sum_{i<p<q<j} b_{ij}b_{pq}E\varepsilon_i\varepsilon_p\varepsilon_q\varepsilon_j\Big| \leqslant \sum_{i<p<q<j} |b_{ij}b_{pq}| \min\{\varphi_{p-i}, \varphi_{j-q}\} \leqslant \sqrt{I_3 I_4}$$

with

$$I_3 = \sum_{i<p<q<j} b_{ij}^2 \min\{\varphi_{p-i}, \varphi_{j-q}\}, \quad I_2 = \sum_{i<p<q<j} b_{pq}^2 \min\{\varphi_{p-i}, \varphi_{j-q}\}.$$

Since $I_3 = \sum_{i<j} L_{ij} b_{ij}^2$ with

$$L_{ij} = \sum_{p:\, i<p<j} \Big[(p-i)\varphi_{p-i}\, I\Big(p-i \leqslant \frac{j-i}{2}\Big) + \sum_{q:\, q>p,\, j-q>p-i} \varphi_{j-q}\Big] \leqslant 2\sum_p p\varphi_p,$$

we conclude that $I_3 \leqslant 2\mathrm{tr}A^2 \sum_p p\varphi_p$. Similarly we get that

$$I_4 = \sum_{p<q} b_{pq}^2 \sum_{i:\, i<p} \Big[(p-i)\varphi_{p-i} + \sum_{j:\, j-q>p-i} \varphi_{j-q}\Big] \leqslant 2\mathrm{tr}A^2 \sum_p p\varphi_p.$$

Control of $\sum_{i<j<p<q} b_{ij}b_{pq}E\varepsilon_i\varepsilon_j\varepsilon_p\varepsilon_q$.

Since

$$|E\varepsilon_i\varepsilon_j\varepsilon_p\varepsilon_q| \leqslant \min\{\varphi_{j-i}, \varphi_{p-j}, \varphi_{q-p}\}, \quad i<j<p<q,$$

we conclude that

$$\Big|\sum_{i<j<p<q} b_{ij}b_{pq}E\varepsilon_i\varepsilon_j\varepsilon_p\varepsilon_q\Big| \leqslant \sum_{i<j<p<q} |b_{ij}b_{pq}| \min\{\varphi_{j-i}, \varphi_{p-j}, \varphi_{q-p}\} \leqslant \sqrt{I_5 I_6},$$

where

$$I_5 = \sum_{i<j<p<q} b_{ij}^2 \min\{\varphi_{p-j}, \varphi_{q-p}\}, \quad I_6 = \sum_{i<j<p<q} b_{pq}^2 \min\{\varphi_{j-i}, \varphi_{p-j}\}.$$

Additionally,

$$I_5 = \sum_{i<j} b_{ij}^2 \sum_{p\colon p>j} \Big[(p-j)\varphi_{p-j} + \sum_{q\colon q-p>p-j} \varphi_{q-p}\Big] \leqslant 2\mathrm{tr}A^2 \sum_p p\varphi_p$$

and similarly

$$I_6 = \sum_{p<q} b_{pq}^2 \sum_{j\colon j<p} \Big[(p-j)\varphi_{p-j} + \sum_{i\colon j-i>p-j} \varphi_{j-i}\Big] \leqslant 2\mathrm{tr}A^2 \sum_p p\varphi_p.$$

Proof of Proposition 5. Assume w.l.o.g. that $i=1$. First note that

$$|P_{11} - z_1'(Z'Z + \varepsilon n I_l)^{-1} z_1| \leqslant \varepsilon n / \lambda_{\min}(Z'Z)$$

(see Claim 1 in the proof of Theorem 1). Applying Theorem* for u_i instead of z_i (as well as Proposition 4) we get that $P(\lambda_{\min}(U_m'U_m) > Cm) \to 1$ for some constant $C>0$, whenever $l = 2\alpha m + o(m)$ and U_m is a $m \times l$ matrix with rows u_{i+1}, $i=1,\dots,m$. In addition,

$$P(\lambda_{\min}(Z'Z) > Cn/3) = P\Big(\lambda_{\min}\Big(\sum_{i=1}^n d_i u_i u_i'\Big) > Cn/3\Big) =$$

$$= EP\Big(\lambda_{\min}\Big(\sum_{i=1}^n d_i u_i u_i'\Big) > Cn/3\Big|\sum_{i=1}^n d_i\Big) = EP(\lambda_{\min}(U_m'U_m) > Cn/3)|_{m=\sum_{i=1}^n d_i}$$

By the law of large numbers, $\sum_{i=1}^n d_i = n/2 + o(n)$ a.s. Therefore,

$$P(\lambda_{\min}(Z'Z) > Cn/3) \to 1$$

and

$$|P_{11} - z_1'(Z'Z + \varepsilon n I_l)^{-1} z_1| \leqslant \frac{3\varepsilon}{C} + o_P(1).$$

By the Sherman-Morrison-Woodbury formula,

$$z_1'(Z'Z + \varepsilon n I_l)^{-1} z_1 = g(z_1'(Z_{-1}'Z_{-1} + \varepsilon n I_l)^{-1} z_1),$$

where $g(x) = x/(x+1)$. By Proposition 4,

$$E|z_1'(Z_{-1}'Z_{-1} + \varepsilon n I_l)^{-1} z_1 - d_1 \mathrm{tr}(Z_{-1}'Z_{-1} + \varepsilon n I_l)^{-1}|^2 =$$

$$= E|u_1' A_l u_1 z_1 - \mathrm{tr} A_l|^2|_{A_l = (Z_{-1}'Z_{-1} + \varepsilon n I_l)^{-1}} \leqslant C E \mathrm{tr}(Z_{-1}'Z_{-1} + \varepsilon n I_l)^{-2} \leqslant \frac{Cl}{\varepsilon^2 n^2} = o(1).$$

Note that

$$|\mathrm{tr}(U_m'U_m + \varepsilon n I_l)^{-1} - \mathrm{tr}(U_m'U_m)^{-1}| \leqslant \frac{\varepsilon n m}{\lambda_{\min}(U_m'U_m)^2} \leqslant \frac{2\varepsilon}{C^2} + o_p(1)$$

whenever $n = 2m + o(m)$ since $P(\lambda_{\min}(U_m'U_m)^2 > Cm) \to 1$ for some constant $C > 0$. As in the Proof of Proposition 2 one could show that $g(\mathrm{tr}(U_m'U_m)^{-1}) \to 2\alpha$ if $l = 2\alpha m + o(m)$. The latter yields We have

$$\begin{aligned} E\big|g(\mathrm{tr}(Z_{-1}'Z_{-1} + \varepsilon n I_l)^{-1}) - 2\alpha\big| &= E\Big[\big|g(\mathrm{tr}(Z_{-1}'Z_{-1} + \varepsilon n I_l)^{-1}) - 2\alpha\big|\Big| \sum_{i=2}^n d_i\Big] \\ &= E\big|g(\mathrm{tr}(U_m'U_m + \varepsilon n I_l)^{-1}) - 2\alpha\big|\Big|_{m=\sum_{i=2}^n d_i} \\ &\leqslant E \min\{2\varepsilon/C^2 + o_P(1), 2\} + E|g(\mathrm{tr}(U_m'U_m)^{-1} - 2\alpha|\Big|_{m=\sum_{i=2}^n d_i} \\ &\leqslant \min\{2\varepsilon/C^2 + o_P(1), 2\} + o(1), \end{aligned}$$

where we take into account that g is bounded function with $|g(x) - g(y)| \leqslant |x-y|$, $x, y \geqslant 0$.

Combining all above estimates together we arrive at

$$|P_{11} - 2\alpha d_1| \leqslant \frac{3\varepsilon}{C} + \frac{2\varepsilon}{C^2} + o_P(1).$$

Q.e.d.

Appendix B.

Theorem*. *Let Assumption 1 hold and*

$$\sup_{a \in \mathbb{R}^l :\, a'a = 1} E|z_1'a|^4 \leqslant K$$

for some $K > 0$ not depending on l. If $l = l(n) = \alpha n + o(n)$ for some $\alpha \in [0,1)$, then $P(\lambda_{\min}(Z'Z) > Cn) \to 1$ as $n \to \infty$ for some $C = C(\alpha, K) > 0$.

Assumptions like $\sup E|z_1'a|^4 \leqslant K$ are only needed to guarantee that there is a large enough constant $L > 0$ such that averages $E \min\{|z_1'a|^2, L\}$ are uniformly (over a in the unit sphere) close to $E|z_1'a|^2 = 1$. That is, linear combinations $z_1'a$ don't explode on average in this sense.