# Off-diagonal elements of projection matrices and dimension asymptotics

STANISLAV ANATOLYEV* CERGE-EI and NES

MAKSIM SMIRNOV CERGE-EI

**Abstract**

We provide insights on asymptotic behavior of the off-diagonal elements of projection matrices in settings, where the dimension of underlying vectors grows with the sample size. Under designs favorable to application of the random matrix theory, the off-diagonal elements are asymptotically normal with a simple variance expression. We also discuss the robustness of the result to deviations of the design from the ideal setup.

**Keywords**: projection matrix, many instruments/regressors, dimension asymptotics

**JEL classification codes**: C21, C26, C46

*Corresponding author. Address: CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic. E-mail: stanislav.anatolyev@cerge-ei.cz. This research was supported by the grant 24-12720S from the Czech Science Foundation. We thank an anonymous expert referee for useful suggestions and Pavel Yaskov for helpful discussions.

# 1 Introduction

Consider $n \times n$ projection matrix $P = Z\left(Z'Z\right)^{-1}Z'$ of rank $\ell$ associated with the $n \times \ell$ instrument (or regressor, or covariate) matrix $Z$ containing $n$ independent copies $z_i$, $i = 1, ..., n$ of $\ell$-vector $z$. The elements of matrix $P$ enter various formulas that are used to derive asymptotics of estimators and tests in the many instrument, many regressor and many covariate literatures (e.g., van Hasselt 2010, Chao, Swanson, Hausman, Newey, and Woutersen 2012, and Cattaneo, Jansson, and Newey 2018, to mention only a few). The qualifier "many" refers to the dimension asymptotics, where $\ell$ increases to infinity proportionately to $n$, with the 'aspect ratio' $\alpha = \lim_{\ell,n\to\infty} \ell/n \in (0, 1)$ . In particular, most important is the asymptotic behavior of the diagonal elements $p_{ii}$ of $P$, particularly, the question of whether the limit of each equals $\alpha$ or these limits exhibit variability across $i$. This issue is analyzed at length in Anatolyev and Yaskov (2017), where, in particular, various examples – idealistic and realistic – of both phenomena are given.

The off-diagonal elements $p_{ij}$, $i \neq j$, however, also play a role – they figure into definitions of some estimators and test statistics, they participate in various derivations and proofs, sometimes they are involved in assumptions about $P$. Hence, their asymptotic behavior may also be of interest – for example, in the context of correlated observations with randomly assigned instruments, or in the case of leave-out estimators that lack the terms associated with the diagonal of $P$. While the asymptotic properties of $p_{ii}$'s were studied in Anatolyev and Yaskov (2017), the asymptotics of $p_{ij}$'s have not been considered yet, and the present note fills this gap.

When the instruments are deterministic, one can observe various patterns for elements of $P$, off-diagonal ones in particular, depending on the instrument design. For example, the group instrument setup (Bekker and van der Ploeg 2005) implies $p_{ij} = \mathbb{I}_{\{i,j\in g \text{ for some group } g\}}/|g|$, so the majority of $p_{ij}$'s are exact zeros, while the minority are strictly positive and may be asymptotically fixed or converge to 0 depending on the design of groups. We instead are interested in obtaining a more systematic pattern of the asymptotics of $p_{ij}$, $i \neq j$, under a random instrument design. The linear regression literature provides very loose bounds for $p_{ij}$ in general, such as $\frac{1}{n} - \frac{1}{2} \le p_{ij} \le \frac{1}{2}$ with an intercept included and $-\frac{1}{2} \le p_{ij} \le \frac{1}{2}$ without an intercept (e.g., Mohammadi 2016), which are not helpful.

Intuition may suggest that, because the total 'mass' of $P$ exactly (with a constant included) or approximately (without a constant) equals $n$, and the trace of $P$ equals $\ell$, by the symmetry arguments, the 'mass' of each off-diagonal element is $(n - \ell) / (n^2 - n) = O(1/n)$, and hence each $p_{ij}$ must perhaps converge to zero with rate $n$. The asymptotic results we obtain below show that the randomness in $p_{ij}$ in fact dominates these biases and drives the convergence rate down to $\sqrt{n}$, and the asymptotic distribution is normal with a simple variance formula, at least for favorable instrument designs.

# 2 Leave-two-out connection

A starting point is a two-permutation extension of the celebrated Sherman-Morrison formula, which may also be of interest on its own. Let $a, b \in \mathbb{R}^p$, and $S \in \mathbb{R}^{p \times p}$ be symmetric and invertible. Then

$$(S + aa' + bb')^{-1} = S^{-1} + S^{-1} \frac{a'S^{-1}b\,(ab' + ba') - (1 + b'S^{-1}b)\,aa' - (1 + a'S^{-1}a)\,bb'}{(1 + a'S^{-1}a)\,(1 + b'S^{-1}b) - (a'S^{-1}b)^2} S^{-1},$$

which can be verified directly. Taking a bilinear form for the pair $(a, b)$ with respect to both sides yields after simplifications

$$a'\,(S + aa' + bb')^{-1}\,b = \frac{a'S^{-1}b}{(1 + a'S^{-1}a)\,(1 + b'S^{-1}b) - (a'S^{-1}b)^2}. \tag{1}$$

Denote a leave-two-out version of $Z'Z$ by $(Z'Z)_{(ij)} = \sum_{k \notin \{i,j\}} z_k z_k'$. Let

$$p_{k_1 k_2}^{(ij)} \equiv z_{k_1}' (Z'Z)_{(ij)}^{-1} z_{k_2}$$

be a corresponding leave-two-out counterpart of $p_{k_1 k_2}$. Applying formula (1) with $a = z_i$, $b = z_j$ and $S = (Z'Z)_{(ij)}$ for $i \neq j$, one obtains

$$p_{ij} = \frac{p_{ij}^{(ij)}}{\left(1 + p_{ii}^{(ij)}\right)\left(1 + p_{jj}^{(ij)}\right) - \left(p_{ij}^{(ij)}\right)^2}. \tag{2}$$

This relation is helpful in establishing asymptotic properties of $p_{ij}$, as all the elements on the right side are bilinear or quadratic forms in vectors $z_i$ and/or $z_j$ that are statistically independent of the associated matrix $(Z'Z)_{(ij)}^{-1}$.

# 3 Asymptotics

We derive an elegant asymptotic result, when the composition of instrument is ideal for application of basic results of the random matrix theory. Then, we discuss how robust the obtained results are to deviations of instrument design from the ideal structure. Note that $P$ is invariant to non-singular linear transformations of $z$, hence one can impose isotropy, i.e. set $var\,(z) = I_\ell$, without loss of generality

Fix a pair of indices $1 \leq i, j \leq n$, $i \neq j$.

**Theorem**. Let the $\ell$ elements of $z$ be IID with zero mean, unit variance and finite fourth moments. Suppose $z_i$ are IID across $i = 1, 2, ..., n$. Then, as $n \to \infty$, $\ell \to \infty$ and $\ell/n \to \alpha \in (0, 1)$, we have

$$\sqrt{n}\, p_{ij} \xrightarrow{d} \mathcal{N}\big(0, \alpha\,(1 - \alpha)\big). \tag{3}$$

While it is assumed that $E[z] = 0$, adding unity to be an element of $z$ does not alter (3): in that case, $p_{ij}$ gains an additional $O_P(1/n)$ term.[1] This also means that the elements of $z$ may not be centered at zero, without changing the conclusion.

# 4 Robustness

The invariance of $P$ to linear transformations of $z$ implies that the Theorem can handle non-independent settings with a "transformed independence design," as named in the high-dimensional statistics literature (see Athey, Imbens, and Wager 2018), and even time series, such as strong autoregressive, settings (e.g., Bai and Zhou 2008). A natural question, however, arises, of how robust this result is to deviations from the transformed independence design and distributional homogeneity.

The statement (3) is based on three ingredients: asymptotic normality for a linear combination of elements of $z_i$, validity of the Marchenko-Pastur law for a large covariance matrix $n^{-1}Z'Z$ associated with $Z$, and asymptotic constancy of the denominator of (2). We will now discuss them one by one.

Asymptotic normality for a linear combination of elements of $z_i$ is natural to hold as long as there is sufficient amount of mixing across them, while the asymptotic variance implied by this combination stays valid because the elements are uncorrelated by design. Arguably, this is the most robust ingredient in obtaining (3).

Even though the Marchenko-Pastur law is most often formulated for the IID case, distributional homogeneity across instruments is not necessary as long as an additional Lindeberg-type condition holds (Bai and Silverstein 2010, Theorem 3.10). Some relaxations in the literature away from independence describe pretty realistic configurations of instrument/covariate sets. In particular, Bryson, Vershynin, and Zhao (2022) consider two interesting setups. In one, all instruments are partitioned into blocks, with tight dependence within blocks but independence across blocks, which is a metaphor for inclusion of a number of basic instruments along with their nonlinear functions (e.g., sieves). In the other, the instrument set is filled with a number of basic instruments and various interactions among them. Subject to a rate restriction on the degree of induced dependence, in both setups the Marchenko-Pastur law is valid, with, perhaps surprisingly, the same aspect ratio $\alpha$ as under coordinate independence.

[1] Suppose that the matrix of included instruments $Z$ is appended by $\iota$, an $n$-vector of ones. By the partitioned matrix inverse formula, $p_{ij}$ gains an additional term

$$\frac{1}{n}\frac{\left(1-e_i'P\iota\right)\left(1-e_j'P\iota\right)}{1-\left(\iota'P\iota\right)/n}$$

(where $e_i'$ and $e_i'$ are $i^{th}$ and $j^{th}$ unit $n$-vectors), which is $O_P\left(1/n\right)$ as long as the constant is not asymptotically perfectly explained by the instruments already included in $Z$.

The third ingredient behind (3) is asymptotic homogeneity of the diagonal elements of $P$. Anatolyev and Yaskov (2017) present analysis of this property and give examples of instrument setups when it does (see examples in Section 4) or does not (see examples in Section 5) take place. When the diagonal of $P$ is asymptotically heterogeneous, the diagonal elements have random limits, which complicates the asymptotic distribution on the right side of (3) in a complex manner according to the transformation (2).

An intermediate result of the Theorem is that $p_{ij} = O_P\left(1/\sqrt{n}\right)$. This conclusion is far more robust than the ultimate result, and does not require the aforementioned ingredients to hold. This asymptotic order is driven by separation, asymptotically, of the eigenvalues of $n^{-1}Z'Z$ from zero, which holds under much weaker conditions than those in the Theorem – in particular, for all the random design examples in Anatolyev and Yaskov (2017). Yaskov (2014) gives quite weak conditions for such separation, which allow distributional heterogeneity, thick tails, and a wide scope of dependence within $z_i$. The result $p_{ij} \asymp 1/\sqrt{n}$ under random instrument design deems invalid the assumption of row-wise absolute summability of elements of $P$ made in the earlier many instruments literature (e.g., van Hasselt 2010, Assumption 3(c)), even though this assumption does hold in the group instrument setup mentioned in the Introduction.

# Appendix: proof of Theorem

Let $\underline{\alpha} = \left(1-\sqrt{\alpha}\right)^2$ and $\bar{\alpha} = \left(1+\sqrt{\alpha}\right)^2$. Denote by $\lambda_k$, $1 \le k \le \ell$, the $\ell$ positive eigenvalues of the symmetric positive definite matrix $\left(Z'Z\right)_{(ij)}/\left(n-2\right)$. By the Marchenko-Pastur law (Bai and Silverstein 2010, Theorem 3.6),

$$\frac{1}{\ell}\sum_{k=1}^{\ell}\delta_{\lambda_k} \xrightarrow{d} \mathcal{MP},$$

whose density is $\left(2\pi x\alpha\right)^{-1}\sqrt{(\bar{\alpha}-x)(x-\underline{\alpha})}\,\mathbb{I}_{\{x\in[\underline{\alpha},\bar{\alpha}]\}}$. By the almost sure convergence of the extreme eigenvalues to $\underline{\alpha}$ and $\bar{\alpha}$ (Bai and Silverstein 2010, Theorem 5.11), we have that $[\min_{1\le k\le \ell}\lambda_k, \max_{1\le k\le \ell}\lambda_k] \subset \left[\frac{1}{2}\underline{\alpha}, \bar{\alpha}+1\right]$ almost surely for all large $n$, while by the second Helly-Bray theorem (Smith 2006) with the function $u \mapsto u^{-2}$ continuous on $\left[\frac{1}{2}\underline{\alpha}, \bar{\alpha}+1\right]$, we have

$$\frac{1}{\ell}\sum_{k=1}^{\ell}\lambda_k^{-2} \xrightarrow{a.s.} \int_{\underline{\alpha}}^{\bar{\alpha}}\frac{1}{x^2}\frac{1}{2\pi x\alpha}\sqrt{(\bar{\alpha}-x)(x-\underline{\alpha})}dx = \frac{1}{(1-\alpha)^3}.$$

This, together with uniform integrability of the sequence, implies that

$$E\left[\frac{n^2}{\ell}z_i'\left(Z'Z\right)_{(ij)}^{-2}z_i\right] = \frac{n^2}{\ell}E\big[\mathrm{tr}\big((Z'Z)_{(ij)}^{-2}\big)\big] = \frac{n^2}{\left(n-2\right)^2}E\left[\frac{1}{\ell}\sum_{k=1}^{\ell}\lambda_k^{-2}\right] \to \frac{1}{(1-\alpha)^3},$$

and also

$$var\left(\mathrm{tr}\left((Z'Z)^{-2}_{(ij)}\right)\right) = \frac{\ell^2}{(n-2)^4}\, var\left(\frac{1}{\ell}\sum_{k=1}^{\ell}\lambda_k^{-2}\right) = o\left(\frac{\ell^2}{n^4}\right).$$

Similarly, by using the second Helly-Bray theorem with the function $u \mapsto u^{-4}$, we can determine the order of

$$E\left[z_i'(Z'Z)^{-4}_{(ij)} z_i\right] = E\left[\mathrm{tr}\left((Z'Z)^{-4}_{(ij)}\right)\right] = \frac{\ell}{(n-2)^4} E\left[\frac{1}{\ell}\sum_{k=1}^{\ell}\lambda_k^{-4}\right] = O\left(\frac{\ell}{n^4}\right).$$

By independence of $z_i$ from $(Z'Z)^{-2}_{(ij)}$ and across $i$, zero mean, unit variance and finiteness of the fourth moments of $z_i$, using the inequality on quadratic forms with power 2 (Bai and Silverstein 2010, Lemma B.26) applied to the conditional variance and then taking unconditional expectations, we have

$$E\left[var\left(z_i'(Z'Z)^{-2}_{(ij)} z_i \middle| (Z'Z)_{(ij)}\right)\right] \leq \bar{C} E\left[z_{ik}^4\right] E\left[\mathrm{tr}\left((Z'Z)^{-4}_{(ij)}\right)\right]$$

for an absolute constant $\bar{C}$. By the analysis of variance formula, $var\left(z_i'(Z'Z)^{-2}_{(ij)} z_i\right)$ equals

$$\begin{aligned} &E\left[var\left(z_i'(Z'Z)^{-2}_{(ij)} z_i \middle| (Z'Z)_{(ij)}\right)\right] + var\left(E\left[\mathrm{tr}\left(z_i'(Z'Z)^{-2}_{(ij)} z_i\right) \middle| (Z'Z)_{(ij)}\right]\right) \\ &\leq \bar{C} E\left[z_{ik}^4\right] E\left[\mathrm{tr}\left((Z'Z)^{-4}_{(ij)}\right)\right] + var\left(\mathrm{tr}\left((Z'Z)^{-2}_{(ij)}\right)\right). \end{aligned}$$

The first term is $O(\ell/n^4)$, while the second is $o(\ell^2/n^4)$, thus

$$var\left(\frac{n^2}{\ell} z_i'(Z'Z)^{-2}_{(ij)} z_i\right) = o(1),$$

implying that

$$\frac{n^2}{\ell} z_i'(Z'Z)^{-2}_{(ij)} z_i \xrightarrow{p} \frac{1}{(1-\alpha)^3}.$$

Define vector $g_{ij}$ by

$$g_{ij} = \sqrt{\ell}\frac{(Z'Z)^{-1}_{(ij)} z_i}{\sqrt{z_i'(Z'Z)^{-2}_{(ij)} z_i}}.$$

Then, conditional on $Z\backslash\{z_j\}$, $\ell^{-1/2} g_{ij}' z_j \xrightarrow{d} \mathcal{N}(0,1)$, as $\ell^{-1} g_{ij}' g_{ij} = 1$, by a version of the central limit theorem (Pötscher and Prucha 2003, Theorem 29). As the limit does not depend on $Z\backslash\{z_j\}$, it is also unconditional limit. Then, using Slutsky's theorem,

$$\sqrt{n}\, p_{ij}^{(ij)} = \sqrt{\frac{\ell}{n}}\sqrt{\frac{n^2}{\ell} z_i'(Z'Z)^{-2}_{(ij)} z_i}\frac{g_{ij}' z_j}{\sqrt{\ell}} \xrightarrow{d} \sqrt{\alpha}\sqrt{\frac{1}{(1-\alpha)^3}}\mathcal{N}(0,1) \stackrel{d}{=} \mathcal{N}\left(0, \frac{\alpha}{(1-\alpha)^3}\right).$$

Consider now the diagonal elements $p_{ii}^{(ij)}$ and $p_{jj}^{(ij)}$ in the denominator of (2). Under the assumptions made, both converge to a constant limit $\alpha/(1-\alpha)$ (Anatolyev and Yaskov 2017), making the denominator of (2) converge in probability to $(1+\alpha/(1-\alpha))^2 - 0^2 = 1/(1-\alpha)^2$. By Slutsky's theorem, the limiting distribution of $\sqrt{n}\, p_{ij}$ is $\mathcal{N}\left(0, \alpha(1-\alpha)\right)$. $\square$

# References


Anatolyev, S. and P. Yaskov (2017): "Asymptotics of diagonal elements of projection matrices under many instruments/regressors," *Econometric Theory*, 33(3), 717–738.

Athey, S., G.W. Imbens, and S. Wager (2018): "Approximate residual balancing,". *Journal of the Royal Statistical Society (Series B)*, 80(4), 597–623.

Bai, Z. and J. Silverstein (2010): *Spectral analysis of large dimensional random matrices*, $2^{\text{nd}}$ edition, New York, Springer.

Bai, Z. and W. Zhou (2008): "Large sample covariance matrices without independence structures in columns," *Statistica Sinica*, 18(2), 425–442.

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

Bryson, J., R. Vershynin, and H. Zhao (2022): "Marchenko-Pastur law with relaxed independence conditions," *Random Matrices: Theory and Applications*, 10, article 2150040, 1–28.

Cattaneo, M., M. Jansson, and W.K. Newey (2018): "Inference in linear regression models with many covariates and heteroscedasticity," *Journal of the American Statistical Association*, 113(523), 1350-1361.

Chao, J.C., N.R. Swanson, J.A. Hausman, W.K. Newey, and T. Woutersen (2012): "Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments," *Econometric Theory*, 28(1), 42–86.

Mohammadi, M. (2016): "On the bounds for diagonal and off-diagonal elements of the hat matrix in the linear regression model," *REVSTAT – Statistical Journal*, 14(1), 75–87.

Pötscher, B.M. and Prucha, I.R. (2003): "Basic elements of asymptotic theory," in: *A Companion to Theoretical Econometrics*, B.H. Baltagi (ed.), Blackwell: Massachusetts.

Smith, W.L. (2006): "Helly–Bray theorems," in: *Encyclopedia of Statistical Sciences*, S. Kotz, C. Read, N. Balakrishnan, B. Vidakovic and N. Johnson (eds.), John Wiley & Sons.

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

Yaskov, P. (2014): "Lower bounds on the smallest eigenvalue of a sample covariance matrix," *Electronic Communications in Probability*, 19, 1–10.