

ASYMPTOTICS OF DIAGONAL ELEMENTS OF PROJECTION MATRICES UNDER MANY INSTRUMENTS/REGRESSORS

STANISLAV ANATOLYEV

CERGE-EI, Czech Republic and New Economic School, Russia

PAVEL YASKOV

Steklov Mathematical Institute of RAS and NUST 'MISIS'

This article sheds light on the asymptotic behavior of diagonal elements of projection matrices associated with instruments or regressors under many instrument/regressor asymptotics. When the diagonal elements do not exhibit variation asymptotically, certain results in the many instrument/regressor literature lead to elegant solutions and conclusions. We establish conditions when this happens, provide relevant examples, and analyze instrument designs, for which this property does or does not hold.

1. INTRODUCTION

A seminal paper by Bekker (1994) spurred the development of alternative asymptotic tools for models with many instrumental variables and/or regressors. In the conclusions implied by these theories, diagonal elements of projection matrices associated with instruments/regressors play a special role. A critical question is whether the diagonal elements do or do not exhibit variation asymptotically, a situation sometimes referred to as an (asymptotically) ‘balanced design’. When this happens, certain results become attainable from a theoretical viewpoint, and become more convenient and elegant from a practical perspective. In particular, when the diagonal elements are asymptotically constant, asymptotic variances of estimators or test statistics may lose their complexity, as shown in van Hasselt (2010), Hansen, Hausman and Newey (2008), Lee and Okui (2012) and Hausman, Newey, Woutersen, Chao and Swanson (2012, HNWCS henceforth).¹ This means that variance estimates may be simplified, which, in addition to facilitating computations, should improve asymptotic approximations. Under asymptotically balanced design, the LIML estimator in an instrumental variables model can be

We would like to thank the editor Peter Phillips and co-editor Victor Chernozhukov for their quick and professional handling of the manuscript, as well as a diligent referee who provided very useful comments. The second author gratefully acknowledges the financial support of the Russian Science Foundation via grant 14-21-00162. Address correspondence to Stanislav Anatolyev, CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic; e-mail: stanislav.anatolyev@cerge-ei.cz.

shown to be asymptotically efficient in a certain class of IV estimates (Anderson, Kunitomo and Matsushita, 2010; Kolesár, 2015). Moreover, the LIML estimator keeps its consistency and asymptotic efficiency properties even under ‘weak heteroskedasticity’ (Kunitomo, 2012). Wang and Kaffo (2016) use the assumption of asymptotically balanced design to show the validity of their modified bootstrap procedure based on LIML estimation with many instruments. Last, but not least, if the diagonal elements do not vary asymptotically, robust chi-square and F-tests become immediately available (Anatolyev and Gospodinov, 2011; Calhoun, 2011; Anatolyev, 2012). It is important to know the circumstances under which an asymptotically balanced design may or may not occur, on the one hand, and how much distortion a failure of this property may create, on the other.²

One well-known example encountered in previous literature is the case of grouped data (Bekker and van der Ploeg, 2005) where the instruments are indicator variables. If group sizes are equal, it is known that the diagonal of the instrument projection matrix is asymptotically homogeneous (e.g., van Hasselt, 2010), and is asymptotically heterogeneous otherwise.³ Another example where the instruments are jointly normal is presented in Anatolyev (2012). To the best of our knowledge, there are no systematic studies of this phenomenon in a random sampling setting, although the common presumption in the many instrument/regressor literature is that the property under consideration is not likely to hold.

In this article, we formulate sufficient conditions under which the diagonal elements are asymptotically nonstochastic. We provide relevant examples in which this property of diagonal elements holds, and work out setups often referenced in the literature for when it does not. In particular, situations that deal with independent instruments (including gaussian), that deal with instruments drawn from a log-concave distribution (again including gaussian), that deal with instruments following a factor model, as well as some others, belong to the case of diagonal elements displaying asymptotic constancy. On the other hand, several situations that are considered in the many instrument literature, such as those of dummy instruments—both stand-alone and those that interact with other instruments, as in Angrist and Krueger (1991)—are characterized by nontrivial asymptotic variation in diagonal elements. The asymptotic variance turns out to take fancy forms, often involving Poisson distributions or a mixture thereof. We consider each of these examples in some detail and numerically characterize the dependence of asymptotic variation on the ratio of instrument numerosity to the sample size.

In our derivations, we use some results from random matrix theory, which is a widely used tool in high-dimensional statistics, machine learning theory, and mathematical physics (e.g., Bai and Silverstein, 2010). The impediment to immediate application of the random matrix theory to the present setting is the typical assumption that the entries of the design matrix are IID not only across rows but also across columns; the existing relaxations of this condition are not sufficient to cover interesting cases arising in regression contexts. Some of the results we derive here are themselves on the edge of today’s research in random matrix theory.

The article is organized as follows. The setup and objectives are described in Section 2. In Section 3 we give some theoretical results that are useful in designing further examples. Section 4 contains examples with asymptotically balanced design, while Section 5 describes settings under which the diagonal elements do exhibit variation in the limit. Section 6 concludes. Proofs of most substantive results are listed in Section 7, while more technical propositions and auxiliary proofs are relegated to an online Appendix available at [is.gd/diagPz](https://doi.org/10.1017/S0266466616000165).

2. SETUP

Let Z_l be an $n \times l$ random matrix with rows $\{z'_{lk}\}_{k=1}^n$ distributed independently and identically as z'_l , where $l \leq n$. To simplify notation, we will further omit index l if there is no ambiguity. The vectors $\{z_k\}_{k=1}^n$ comprising the matrix Z are associated with n IID observations on instruments or regressors (which we will refer to as *instruments* throughout) in a random sampling framework. The object of our study is the projector $P = \|P_{ij}\|_{i,j=1}^n$ on the linear span of columns of Z , and particularly its diagonal elements $\{P_{kk}\}_{k=1}^n$. If Z has full rank, then $P = Z(Z'Z)^{-1}Z'$ and $P_{kk} = z'_k(Z'Z)^{-1}z_k$ for any k . In general, $(Z'Z)^{-1}$ should be replaced by $(Z'Z)^+$, the Moore–Penrose generalized inverse of $Z'Z$, and then $\text{rk}(P) = \text{rk}(Z) = \sum_{k=1}^n P_{kk}$. Note that for any $l \times l$ nondegenerate matrix C , the linearly transformed instruments $\{Cz_k\}_{k=1}^n$ correspond to the same projection matrix P as the initial instruments $\{z_k\}_{k=1}^n$. Therefore, when necessary, we may assume that the instruments are normalized so that $\mathbb{E}[zz'] = I_l$.

Throughout, the maintained assumption is that of many instruments, first introduced by Bekker (1994): as $n \rightarrow \infty$,

$$\frac{l}{n} = \alpha + o(1) \tag{MI}$$

for some $\alpha \in (0, 1)$. The rate of convergence is sufficient for our purposes, although it has to be faster in order for asymptotic properties of various estimators and tests to take place.

We are interested in the asymptotic behavior of the following two quantities:⁴

$$\alpha_n = \frac{1}{n} \sum_{k=1}^n P_{kk}$$

and

$$\Delta_n = \frac{1}{n} \sum_{k=1}^n (P_{kk} - \alpha_n)^2.$$

Note that if Z has full rank, then $\alpha_n = l/n \stackrel{A}{=} \alpha$, but in general $\alpha_n = \text{rk}(Z)/n \stackrel{A}{\leq} \alpha$. It is the limits $\text{plim} \alpha_n$ and $\text{plim} \Delta_n$ that explicitly appear in asymptotic variances of many estimators and test statistics (see references in the Introduction).

We are mainly interested in whether

$$\Delta_n \xrightarrow{P} 0$$

within the asymptotic framework (MI), i.e., if there is no variation in the diagonal elements of P , asymptotically. In this case, the asymptotic variances of estimators simplify, certain estimators gain attractive properties, robust chi-square tests become available, etc. (see references in the Introduction).

We will use the following notation. Let $\lambda_{\min}(A)$, $\lambda_{\min}^*(A)$, and $\lambda_{\max}(A)$ be the least eigenvalue, the least nonzero eigenvalue, and the largest eigenvalue of a symmetric positive semidefinite matrix A , respectively. Define also the L^2 (Euclidean) norm $\|a\| = \sqrt{a'a}$ for any vector a . Negative indexing of a matrix means deletion of a corresponding row. By $\mathbb{I}_{\{\mathcal{A}\}}$ we denote the indicator function for event \mathcal{A} , by $U[0, 1]$ we denote the uniform distribution on $[0, 1]$, and by $Po(a)$ we denote the Poisson distribution with parameter a .

3. SOME USEFUL NOTIONS AND INTERMEDIATE RESULTS

A matrix relationship useful for many derivations is the Sherman–Morrison formula (Sherman and Morrison, 1950), which states that, for each $a \in \mathbb{R}^l$ and any $l \times l$ nondegenerate matrix A ,

$$(A + aa')^{-1} = A^{-1} - \frac{A^{-1}aa'A^{-1}}{1 + a'A^{-1}a} \tag{1}$$

and, as a result,

$$a'(A + aa')^{-1}a = a'A^{-1}a - \frac{(a'A^{-1}a)^2}{1 + a'A^{-1}a} = \frac{a'A^{-1}a}{1 + a'A^{-1}a}. \tag{2}$$

In particular, this relationship helps us to write out a typical diagonal element of P as a function of a quadratic form whose vector and matrix are independent:

$$P_{kk} = \frac{z'_k(Z'_{-k}Z_{-k})^{-1}z_k}{1 + z'_k(Z'_{-k}Z_{-k})^{-1}z_k}. \tag{3}$$

Recall that Z_{-k} denotes the matrix obtained from Z by removing k -th row from Z .

The critical technical condition we will be constantly using is

Condition A. There is $\delta > 0$ such that $\mathbb{P}(\lambda_{\min}(Z'Z) > \delta n) \rightarrow 1$ as $n \rightarrow \infty$.

Condition A allows one to control expressions like $(Z'Z)^{-1}$. Note that it implies that $Z'Z$ has full rank with probability approaching one, and thus $\alpha_n \xrightarrow{P} \alpha$ as $n \rightarrow \infty$. Heuristically, Condition A allows us to show that

$$\text{tr}(Z'_{-k}Z_{-k})^{-1} \approx \text{tr}(Z'Z)^{-1} \approx \mathbb{E}\text{tr}(Z'Z)^{-1}$$

for n that is large enough. Condition A holds given the following set of conditions.

LEMMA 3.1. Let $\mathbb{E}z_l z_l' = I_l$ and $\mathbb{E}|a'z_l| \geq c > 0$ for some $c > 0$, any $l \geq 1$, and all $a \in \mathbb{R}^l$ with $a'a = 1$. There exists an absolute constant $K > 0$ such that, if $\alpha \leq Kc^2$, then Condition A holds.⁵

A version of Condition A, among other things, guarantees that all diagonal elements of P are bounded away from unity; see Lemma A.1 in the online Appendix. This property is intensively used as an additional assumption in recent works on many instrument asymptotics (e.g., Hansen, Hausman and Newey, 2008; HNWCS, 2012; Bekker and Crudu, 2015).

Let us now introduce a property that plays an important role in random matrix theory (see, for example, chapter 19 in Pastur and Shcherbina, 2011).

Definition. An array of random vectors $\{x_l\}_{l \geq 1}$ with $x_l \in \mathbb{R}^l$ and a non-negative random variable d satisfy *property P* if, for any sequence of $l \times l$ positive semidefinite symmetric matrices $\{A_l\}_{l \geq 1}$ with $\max_l \lambda_{\max}(A_l) < \infty$, we have

$$\frac{x_l' A_l x_l - d \operatorname{tr}(A_l)}{l} \xrightarrow{P} 0$$

as $l \rightarrow \infty$.

Property P is a form of the weak law of large numbers for quadratic forms. Namely, if $d = 1$ and $\mathbb{E}x_l x_l' = I_l$, then $\operatorname{tr}(A_l) = \mathbb{E}x_l' A_l x_l$ and, as a result, Property P states that $(x_l' A_l x_l - \mathbb{E}x_l' A_l x_l)/l \xrightarrow{P} 0$.

Property P is preserved when a finite number of components to random vectors satisfying it is added; see Lemma A.2 in the online Appendix. The conditions under which Property P implies Condition A are stated in Lemma A.3 in the online Appendix.

In important special cases, Property P allows us to obtain explicit analytical formulas for the limit of diagonal elements of P , as the following theorem shows.

THEOREM 3.2. Let $(\{z_l\}_{l \geq 1}, d)$ satisfy Property P and $\{(z_{lk}, d_k)\}_{k=1}^n$ be IID copies of (z_l, d) . If $\alpha \in (0, \mathbb{P}(d > 0))$ and Condition A holds, then

$$P_{kk} \xrightarrow{P} \frac{cd_k}{1 + cd_k}$$

for any fixed k as $n \rightarrow \infty$, where $c > 0$ solves

$$\mathbb{E} \frac{cd}{1 + cd} = \alpha. \tag{4}$$

Heuristically, when Property P holds, we have $z_k'(Z'_{-k} Z_{-k})^{-1} z_k \stackrel{d}{\approx} d_k \operatorname{tr}(Z'_{-k} Z_{-k})^{-1}$, and, using (3) and the implication of Condition A,

$$P_{kk} \stackrel{d}{\approx} \frac{d_k \operatorname{tr}(Z'_{-k} Z_{-k})^{-1}}{1 + d_k \operatorname{tr}(Z'_{-k} Z_{-k})^{-1}} \stackrel{d}{\approx} \frac{d_k \mathbb{E} \operatorname{tr}(Z' Z)^{-1}}{1 + d_k \mathbb{E} \operatorname{tr}(Z' Z)^{-1}}.$$

Further,

$$\alpha \approx \mathbb{E}P_{kk} \approx \mathbb{E} \frac{d_k \mathbb{E}\text{tr}(Z'Z)^{-1}}{1 + d_k \mathbb{E}\text{tr}(Z'Z)^{-1}},$$

hence equation (4) for c . If $d = 1$, then $P_{kk} \xrightarrow{P} \alpha$ for each fixed k , and then $\Delta_n \xrightarrow{P} 0$. If d is a nongenerate random variable, P_{kk} are asymptotically different and $\Delta_n \xrightarrow{P} 0$.

Remark 3.3. If $\alpha > \mathbb{P}(d > 0)$ in Theorem 3.2, then $Z'Z = \sum_{k=1}^n z_{lk} z'_{lk}$ may be degenerate with a large probability. Indeed, if $\alpha > \mathbb{P}(d > 0)$ and $z_{lk} = \sqrt{d_k} u_{lk}$ for IID $l \times 1$ random vectors u_{lk} and IID non-negative random variables d_k , then

$$\text{rank}(Z'Z) \leq \sum_{k=1}^n \mathbb{I}_{\{d_k > 0\}} = n\mathbb{P}(d > 0) + o_p(n) < l$$

for n that is large enough. As a result, $\det(Z'Z) = 0$.

Finally, the following lemma provides certain sufficient conditions for Property P. These conditions will be used in the next section to construct examples of instrument designs for which $\Delta_n \xrightarrow{P} 0$.

LEMMA 3.4. *For each $l \geq 1$, let x_l be a centered random vector in \mathbb{R}^l with $\mathbb{E}x_l x'_l = I_l$, and let d be a random variable. Then $(\{d x_l\}_{l \geq 1}, d^2)$ satisfies Property P if any of the following conditions holds:*

- (a) *each $x_l = (\xi_1, \dots, \xi_l)'$ for some ξ_j that are infinite linear combinations of $\{e_k\}_{k \geq 1}$ converging in mean square, where $\{e_k\}_{k \geq 1}$ are IID random variables with zero mean and unit variance;*
- (b) *each $x_l = (\xi_1, \dots, \xi_l)'$ for some ξ_j as in (a), where $\{e_k\}_{k \geq 1}$ are independent random variables with zero mean, unit variance, and $\mathbb{E}|e_k|^{2+\delta} \leq C$ for some $C, \delta > 0$ not depending on k ;*
- (c) *each x_l is a random vector with a log-concave density;*
- (d) *each $x_l = F_l(v_m)$ for a centered Gaussian random vector v_m in \mathbb{R}^m , where $m = m(l)$, $\lambda_{\max}(\text{var}(v_m)) \leq C$ for some $C > 0$ not depending on m , and $F_l : \mathbb{R}^m \rightarrow \mathbb{R}^l$ is a c -Lipschitz map with $\|F_l(u) - F_l(v)\| \leq c\|u - v\|$ for all $u, v \in \mathbb{R}^m$ and some $c > 0$ not depending on l .*

4. EXAMPLES WHEN $\Delta_n \xrightarrow{P} 0$

4.1. Indicator instruments with equal group sizes

A well-known example in the literature is the case of grouped data (Bekker and van der Ploeg, 2005) where the instruments are indicator variables. The instrument design in this example is deterministic and hence does not fit in our IID framework. However, α_n and Δ_n can be computed by brute force. Assume that all

observations are split into l groups. If group sizes are equal, any size equals n/l . The instrument matrix is comprised of indicator variables: $Z = I_l \otimes \iota_{n/l}$, where ι_m is an m -vector of ones, and the diagonal elements of P are $P_{kk} = l/n$ for all k . Then it is easy to derive that $\alpha_n = l/n$ and $\Delta_n = 0$.

The example in Subsection 5.2 shows how abruptly the situation changes when dummy instruments have random design, even though group sizes stay equal on average.

4.2. Gaussian instruments

Our first example in an IID environment is most straightforward. Let $z_l = (1, v'_{l-1})'$ for a Gaussian random vector v_{l-1} in \mathbb{R}^{l-1} .

THEOREM 4.1. *As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{P} a$, $P_{kk} \xrightarrow{P} a$ for any fixed k , and $\Delta_n \xrightarrow{P} 0$.*

See also a brute force proof in Anatolyev (2012, Appendix A). The conclusion implies, in particular, that Gaussian designs for instruments in simulation studies (e.g., Hahn and Inoue, 2002; Davidson and Mackinnon, 2006; Anderson, Kunitomo and Matsushita, 2011) may have missed important effects of asymptotic heterogeneity of the diagonal of P present in general theoretical results.

4.3. Independent instruments

This example generalizes the case of Gaussian instruments. Let $\{e_k\}_{k \geq 1}$ be independent random variables with zero mean, unit variance, and $\mathbb{E}|e_k|^{2+\delta} \leq C$ for some $C, \delta > 0$ not depending on k . Now, let $z_l = (1, e_1, \dots, e_{l-1})'$.

THEOREM 4.2. *As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{P} a$, $P_{kk} \xrightarrow{P} a$ for any fixed k , and $\Delta_n \xrightarrow{P} 0$.*

This means that the instrument design with many independent instruments drawn from different distributions (barring those with extremely thick tails) does conform to the simplifying implication.

4.4. Instruments drawn from log-concave distribution

This example generalizes the case of gaussian instruments in a different direction. Let $z_l = (1, v'_{l-1})'$ for an \mathbb{R}^{l-1} -valued random vector v_{l-1} having a centered log-concave density.⁶ Recall some useful properties of (probability) distributions with log-concave densities or, more generally, log-concave distributions.⁷

This class of distributions generalizes the multivariate normal distribution. In the one-dimensional case, it includes many common distributions, for example, Weibull, Gamma, and Beta distributions when their parameters are greater or equal to one (see Bagnoli and Bergstrom, 2005). Like the class of normal distributions, it is closed under affine transformations, formation of products of measures,

marginalization, convolution, and weak limits (see Saumard and Wellner, 2014). This is why the class of log-concave distributions is considered a natural nonparametric generalization of the class of normal distributions.

THEOREM 4.3. *As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{P} a$, $P_{kk} \xrightarrow{P} a$ for any fixed k , and $\Delta_n \xrightarrow{P} 0$.*

4.5. Instruments distributed according to Gaussian copula

The results for Gaussian instruments can be extended to certain nonlinear transformations of Gaussian vectors (e.g., Gaussian copulas). Let $z_l = (1, v'_{l-1})'$, where $v_{l-1} = F_{l-1}(u_{l-1})$ for a centered Gaussian random vector u_{l-1} in \mathbb{R}^{l-1} and, for $x = (x_1, \dots, x_{l-1})' \in \mathbb{R}^{l-1}$,

$$F_{l-1}(x) = (f_1(x_1), \dots, f_{l-1}(x_{l-1}))'.$$

Here $f_k : \mathbb{R} \rightarrow \mathbb{R}$ are c -Lipschitz functions for some $c > 0$ (see Lemma 3.4(d)). Each f_k may depend on l , but c depends neither on k nor on l .

Suppose also every v_{l-1} has zero mean and there are constants $\lambda, C > 0$ such that $\lambda_{\min}(\mathbb{E}v_{l-1}v'_{l-1}) \geq \lambda$ and $\lambda_{\max}(\text{var}(u_{l-1})) \leq C$ for each $l > 1$.

THEOREM 4.4. *As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{P} a$, $P_{kk} \xrightarrow{P} a$ for any fixed k , and $\Delta_n \xrightarrow{P} 0$.*

4.6. Instruments that are sums of IID random variables

Let $\{e_k\}_{k \geq 1}$ be IID random variables with zero mean and unit variance. For each $j \geq 1$, let ξ_j be an infinite linear combination of e_k that converges in mean square. Assume that $z_l = (1, \xi_1, \dots, \xi_{l-1})'$ and $\mathbb{E}z_l z'_l$ is nondegenerate for all $l > 1$.

THEOREM 4.5. *As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{P} a$, $P_{kk} \xrightarrow{P} a$ for any fixed k , and $\Delta_n \xrightarrow{P} 0$.*

4.7. Instruments following a factor model

Let $z_l = (1, v'_{l-1})'$ with $v_{l-1} = \Lambda_{l-1,m} f_m + \varepsilon_{l-1}$, where f_m is an $m \times 1$ vector of latent factors, ε_{l-1} is an $(l-1) \times 1$ disturbance term, and $\Lambda_{l-1,m}$ is an $(l-1) \times m$ matrix of factor loadings. Suppose m satisfies $m/l < 1$, $f_m = (\sigma_1 e_1, \dots, \sigma_m e_m)'$, and $\varepsilon_{l-1} = (\sigma_{m+1} e_{m+1}, \dots, \sigma_{m+l-1} e_{m+l-1})'$ for some constants $\{\sigma_k\}_{k=1}^{l+m-1}$ and independent random variables $\{e_k\}_{k=1}^{l+m-1}$ with zero mean, unit variance, and $\mathbb{E}|e_k|^{2+\delta} \leq C$ for some $C, \delta > 0$ not depending on k, l, m .

THEOREM 4.6. *As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{P} a$, $P_{kk} \xrightarrow{P} a$ for any fixed k , and $\Delta_n \xrightarrow{P} 0$.*

5. EXAMPLES WHEN $\Delta_n \xrightarrow{p} 0$

5.1. Indicator instruments with unequal group sizes

In the well known case of grouped data (Bekker and van der Ploeg, 2005), if the group sizes are not equal, the diagonal of the instrument projection matrix is generally asymptotically heterogeneous. Let us assume that all observations are split into l groups and proceed by brute force. Denoting the size of the g^{th} group by n_g , we observe that the instrument matrix is comprised of indicator variables: $Z = \text{dg}\{\iota_{n_g}\}_{g=1}^l$, where ι_m is an m -vector of ones. Then the diagonal elements of P are $P_{kk} = n_g^{-1}$ such that the k^{th} observation belongs to g^{th} group, and

$$\alpha_n = \frac{1}{n} \sum_{g=1}^l n_g \frac{1}{n_g} = \frac{l}{n}$$

and

$$\Delta_n = \frac{1}{n} \sum_{g=1}^l n_g \left(\frac{1}{n_g} - \frac{l}{n} \right)^2 = \frac{1}{n} \sum_{g=1}^l \frac{1}{n_g} - \left(\frac{l}{n} \right)^2,$$

which need not equal zero. See Abutaliev and Anatolyev (2013) for a numerical evaluation of the effects of asymptotic heterogeneity of the diagonal of P on the asymptotic variance of LIML estimates.

5.2. Dummy instruments with disjoint supports

Let $z_l = (d_1, \dots, d_l)'$, where d_1, \dots, d_l are Bernoulli random variables subject to the constraint $\sum_{j=1}^l d_j = 1$. Each d_j implicitly depends on l . We first consider the design when the underlying factors have the same frequency of occurrence: $\mathbb{P}(d_j = 1) = 1/l$ for any fixed j . Let $\zeta \sim \text{Po}(1/\alpha)$, and let

$$\zeta = \frac{1}{\zeta + 1}.$$

THEOREM 5.1. *As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{p} \mathbb{E}\zeta$, $P_{kk} \xrightarrow{d} \zeta$ for any fixed k , and $\Delta_n \xrightarrow{p} \text{var}(\zeta)$.*

Figure 1(i) illustrates the dependence between $a = \text{plim} \alpha_n = \mathbb{E}\zeta$ and $\Delta = \text{plim} \Delta_n = \text{var}(\zeta)$ when α changes from 0 to 1.

Generally, when $\mathbb{P}(d_j = 1)$ nontrivially depends on j and asymptotically behaves like $O(1/l)$, the distribution of ζ will be a mixture of Poisson distributions. To see this, consider the following design. Let u be a random variable with values in $[0, 1]$ and a bounded density $f = f(u)$, $u \in [0, 1]$. Set $d_j = 1$ if $(j-1)/l \leq u < j/l$, $1 \leq j \leq l$, and $d_j = 0$ otherwise.

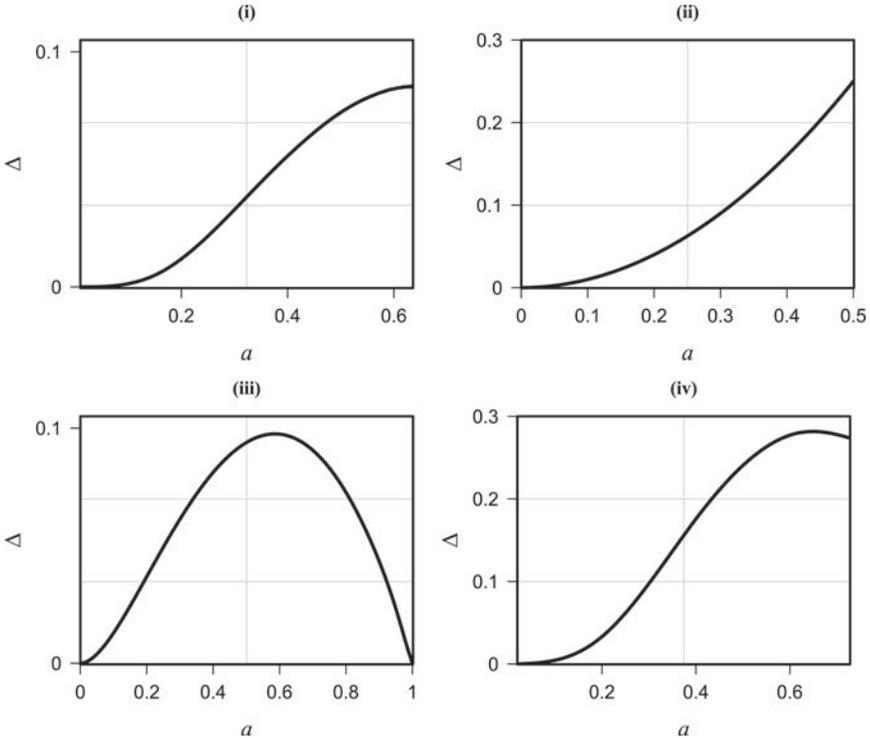


FIGURE 1. Illustrations for examples of Section 5.

Notes: each figure illustrates one of examples of Section 5: (i) Theorem 5.1 with $a = \text{plim } a_n$ and $\Delta = \text{plim } \Delta_n$. (ii) Theorem 5.3 with $a = \text{plim } a_n$ and $\Delta = \text{plim } \Delta_n$. (iii) Theorem 5.6 with $a = \text{plim } a_n$ and $\Delta = \text{plim } \Delta_n$. (iv) Theorem 5.7 for $z \sim U[0, 1]$ with $a = \mathbb{E}v'V(\xi)v$ and $\Delta = \text{var}(v'V(\xi)v)$.

COROLLARY 5.2. Let ξ satisfy

$$\mathbb{P}(\xi = m) = \int_0^1 \frac{e^{-\lambda(u)} \lambda(u)^m}{m!} f(u) du, \quad m = 0, 1, \dots,$$

for $\lambda(u) = f(u)/\alpha$. Then the conclusion of Theorem 5.1 holds.

The case under consideration is related to the famous setup of Angrist and Krueger (1991) where dummy variables with disjoint supports are used. One group of such instruments (30 units) is formed as the products of year-of-birth and quarter-of-birth dummies. Another group of instruments (150 units) is formed as the products of state-of-birth and quarter-of-birth dummies. The diagonal elements of projection matrices for each group can be analyzed along the lines of Corollary 5.2. The case when the instrument set includes both groups is less amenable to analysis because of the complicated structure of $(Z'Z)^{-1}$.

5.3. Many instruments interacted with dummy variables

We consider the case of one dummy variable first. Let u_l be one of a set of random vectors z_l from Subsection 4.2 through 4.7 with nondegenerate $\mathbb{E}u_l u_l'$. Construct z_l as $z_l = u_l d$, where d is a Bernoulli random variable. Let also $z_{lk} = u_{lk} d_k$ for each k , where $\{(d_k, u_{lk})\}_{k=1}^n$ are IID copies of (d, u_l) .

By Remark 3.3, $Z'Z$ is degenerate with large probability when $l/n = \alpha + o(1)$ and $\alpha > \mathbb{P}(d = 1)$. We will consider the case with $\alpha < \mathbb{P}(d = 1)$.

THEOREM 5.3. *Let $\alpha < \mathbb{P}(d = 1)$, and let u_l , $l \geq 1$, be independent of d . As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{p} \alpha$,*

$$P_{kk} \xrightarrow{p} \alpha \frac{d_k}{\mathbb{P}(d = 1)}$$

for any fixed k , and

$$\Delta_n \xrightarrow{p} \alpha^2 \frac{\mathbb{P}(d = 0)}{\mathbb{P}(d = 1)}.$$

Figure 1(ii) illustrates the dependence between $a = \text{plim} \alpha_n = \alpha$ and $\Delta = \text{plim} \Delta_n$ in Theorem 5.3 when $\mathbb{P}(d = 1) = 1/2$ and α changes from 0 to $1/2$.

We now consider the case of several dummy variables. Let

$$z_l = d \otimes u_l = (d_1 u_l', \dots, d_m u_l')',$$

where $d = (d_1, \dots, d_m)'$ is a vector of m Bernoulli random variables subject to the constraint $d_1 + \dots + d_m \leq 1$, and m is a fixed number. Write $z_{lk} = (d_{jk})_{j=1}^m \otimes u_{lk}$ for each k , $1 \leq k \leq n$, where $\{u_{lk}, (d_{jk})_{j=1}^m\}_{k=1}^n$ are IID copies of (u_l, d) .

Again, it can be shown (cf. Remark 3.3) that $Z'Z$ is degenerate with large probability when $\alpha > \min_j \mathbb{P}(d_j = 1)$.

COROLLARY 5.4. *Let $\alpha < \min_j \mathbb{P}(d_j = 1)$, and let u_l , $l \geq 1$, be independent of d . As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{p} \alpha m$,*

$$P_{kk} \xrightarrow{p} \sum_{j=1}^m \frac{\alpha d_{jk}}{\mathbb{P}(d_j = 1)}$$

for any fixed k , and

$$\Delta_n \xrightarrow{p} \sum_{j=1}^m \frac{\alpha^2}{\mathbb{P}(d_j = 1)} - (\alpha m)^2.$$

Remark 5.5. The independence of d and u_l in Theorem 5.3 and Corollary 5.4 can be relaxed as long as Condition A holds. In general, Condition A may not hold. For example, let $u_l = (1, e_1, \dots, e_{l-1})'$ for IID $\{e_k\}_{k \geq 1}$ and $e_1 = (1 - d)\zeta / \text{var}(d)$, where ζ is independent of d and has $\mathbb{E}\zeta = 0$ and $\mathbb{E}\zeta^2 = 1$. Then $d e_1 = 0$ and $\lambda_{\min}(Z'Z) = 0$.

5.4. Instrument interactions with many dummy variables

This artificial example is encountered in the simulation sections of many recent theoretical studies, in particular, HNWCS (2012), Anatolyev (2013), and Bekker and CruDu (2015). Let

$$z_l = \left(1, v, v^2, v^3, v^4, vu'_{l-5}\right)',$$

where $u_{l-5} = (d_1, \dots, d_{l-5})'$ contains IID Bernoulli random variables d_j with $\mathbb{P}(d_j = 0) = \mathbb{P}(d_j = 1) = 1/2$, and $v \sim \mathcal{N}(0, 1)$ is independent of u_{l-5} . Write $z_{lk} = (1, v_k, \dots, v_k^4, v_k u'_{l-5,k})'$ for each k , $1 \leq k \leq n$, where $\{(v_k, u_{l-5,k})\}_{k=1}^n$ are IID copies of (v, u_{l-5}) .

Let $c > 0$ solve the equation

$$\mathbb{E} \frac{cv^2}{1 + cv^2} = \alpha.$$

THEOREM 5.6. As $n \rightarrow \infty$, we have $\alpha_n \xrightarrow{P} \alpha$,

$$P_{kk} \xrightarrow{P} \frac{cv_k^2}{1 + cv_k^2}$$

for any fixed k , and

$$\Delta_n \xrightarrow{P} \text{var} \left(\frac{cv^2}{1 + cv^2} \right).$$

Figure 1(iii) illustrates the dependence between $\alpha = \text{plim} \alpha_n = \alpha$ and $\Delta = \text{plim} \Delta_n$ when α changes from 0 to 1.

5.5. Instruments that are splines

Hansen, Hausman and Newey (2008), HNWCS (2012) and other recent treatments propose to approximate an unknown nonlinear reduced form by a linear combination of many approximating functions. A common approach is to use B-splines with a sufficient degree of smoothness and uniformly spaced knots. Below we will study (a simple version of) this case.

Let z be a $[0, 1]$ -valued ‘basic’ instrumental variable. Let

$$z_l = (B_1(z), \dots, B_{l+1}(z))',$$

where $B_j = B_j(z)$, $j = 1, \dots, l+1$, are B-splines in their natural order. To produce results that are as clear as possible, we consider only linear B-splines with uniformly spaced knots $\{s_j\}_{j=-1}^{l+1}$ with $s_j = j/l$. All results can be rewritten to accommodate higher-order B-splines (e.g., cubic B-splines).

Note first that $B_k(z)B_j(z) = 0$ for all $z \in [0, 1]$ when $|k - j| > 1$. Hence, $Z'Z$ is a tridiagonal matrix for linear B-splines (it would be 9-diagonal for cubic B-splines, for example). The inverse $(Z'Z)^{-1}$ is a matrix with a complicated structure (e.g., see Meurant, 1992). To proceed, we simplify by setting $z_l = (\zeta_1, \zeta_2, \dots, \zeta_{2l})'$, where $\zeta_{2j-1} = B_j(z)\mathbb{I}_{[s_{j-1} \leq z < s_j]}$ and $\zeta_{2j} = B_{j+1}(z)\mathbb{I}_{[s_{j-1} \leq z < s_j]}$ for $1 \leq j \leq l$. For linear B-splines, by definition, if $z \in [0, 1]$, then $B_1(z) = \zeta_1(z)$, $B_j(z) = \zeta_{2j-2}(z) + \zeta_{2j-1}(z)$ for $j = 2, \dots, l + 1$, and $B_{l+1}(z) = \zeta_{2l}(z)$.

In this case, $Z'Z$ is a block diagonal matrix with 2×2 diagonal blocks, and its inverse $(Z'Z)^{-1}$ has a similar structure. Let $\{u_i\}_{i=0}^\infty$ be IID random variables with $u_i \sim U[0, 1]$. Define $v_i = (u_i, 1 - u_i)'$, $i \geq 0$, $v = v_0$, and

$$V(0) = \frac{1}{v'v}, \quad V(m) = \left(\sum_{i=0}^m v_i v_i' \right)^{-1}, \quad m \geq 1.$$

THEOREM 5.7. *Let $z \sim U[0, 1]$. As $n \rightarrow \infty$, we have $P_{kk} \xrightarrow{d} v'V(\xi)v$ for any fixed k , where $\xi \sim \text{Po}(1/\alpha)$ is independent of all other random variables involved.*

Under the conditions of Theorem 5.7, $\eta = v'V(\xi)v$ has the following cumulative distribution function:

$$\mathbb{P}(\eta \leq x) = \sum_{m=0}^\infty \exp\{-1/\alpha\} \frac{\mathbb{P}(v'V(m)v \leq x)}{\alpha^m m!}.$$

As in Corollary 5.2, one can show that the result of Theorem 5.7 still holds if z has a bounded density $f = f(z)$, $z \in [0, 1]$, and that the distribution of ξ is a mixture of Poisson distributions from Corollary 5.2. The proof is available on request.

We also conjecture that $\alpha_n \xrightarrow{P} \mathbb{E}v'V(\xi)v$ and $\Delta_n \xrightarrow{P} \text{var}(v'V(\xi)v)$, though we do not have a formal proof. Figure 1(iv) illustrates the dependence between $a = \text{plim } \alpha_n = \mathbb{E}v'V(\xi)v$ and $\Delta = \text{plim } \Delta_n = \text{var}(v'V(\xi)v)$ when α changes from 0 to $1/2$ (recall that the number of instruments is $2l \approx 2\alpha n$).

6. CONCLUSION

We have derived sufficient conditions for asymptotic homogeneity of a diagonal of the projection matrix under many instrument/regressor asymptotics. We have analyzed several examples when asymptotic homogeneity holds, and several examples when the diagonal is asymptotically heterogeneous. Intuitively, whether a given instrument design leads to the former or the latter case depends on whether there is enough mixing inside the utilized instruments, or whether there are instead common factors in a non-negligible fraction of them. This circumstance validates or, conversely, precludes the validity of a law of large numbers for certain quadratic forms.

We have also calibrated the amount of variation in the diagonal elements, $\text{plim } \Delta_n$, which sometimes attains values up to 0.25 and higher. Because this

quantity explicitly appears in asymptotic variances of some estimators and test statistics, it would be interesting to know how much distortion can be caused by ignoring asymptotic heterogeneity and setting $\text{plim } \Delta_n$ to zero. Let us look at the asymptotic variance of the modified J-statistic (Lee and Okui, 2012; Anatolyev, 2013). The relative difference between this asymptotic variance and the same quantity when the condition $\text{plim } \Delta_n = 0$ is imposed equals

$$\frac{\kappa}{2\alpha(1-\alpha)} \text{plim } \Delta_n,$$

where κ denotes excess kurtosis of the structural error. If the structural error is even moderately leptokurtic, this formula and our calibration imply that sufficiently high relative distortions are possible (e.g., up to 50%, if excess kurtosis equals unity).

7. MAIN PROOFS

Proof of Theorem 4.1. We can normalize z_l so that $z_l = (1, e_1, \dots, e_{l-1})'$ for $e_k \sim \text{IID } \mathcal{N}(0, 1)$ and apply Theorem 4.2 to obtain the desired result. ■

Proof of Theorem 4.2. By Lemma 3.4(b) and Lemma A.2, $(\{z_l\}_{l \geq 1}, 1)$ satisfies Property P. By note 5 and Lemma A.7, there is $c > 0$ such that $\mathbb{E}|z_l' a| \geq c$ for all $l \geq 1$ and $a \in \mathbb{R}^l$ with $a' a = 1$. By Lemma A.3, Condition A holds and $\alpha_n \xrightarrow{P} \alpha$.

Moreover, by Theorem 3.2, $P_{kk} \xrightarrow{P} \alpha$ for any fixed k . Recall a version of the Lebesgue dominated convergence theorem:

if $\zeta_n \xrightarrow{d} \zeta$ and $|\zeta_n| \leq 1$ a.s. for all n , then $\mathbb{E}\zeta_n \rightarrow \mathbb{E}\zeta$. (5)

By this theorem and the exchangeability of $\{(P_{kk}, \alpha_n)\}_{k=1}^n$,

$$\mathbb{E}\Delta_n = \mathbb{E} \frac{1}{n} \sum_{k=1}^n (P_{kk} - \alpha_n)^2 = \mathbb{E}(P_{11} - \alpha_n)^2 \rightarrow 0. \quad \blacksquare$$

Proof of Theorem 4.3. We can normalize v_{l-1} in z_l by some affine transformation so that $\mathbb{E}v_{l-1}v_{l-1}' = I_{l-1}$. As a result, $\mathbb{E}z_l z_l' = I_l$ for $z_l = (1, v_{l-1}')'$, since $\mathbb{E}v_{l-1}$ has zero entries. Affine transformations preserve log-concavity of distributions (see Proposition 3.1 in Saumard and Wellner, 2014). Therefore, the normalized v_{l-1} still has a log-concave density.

Moreover, for any nonzero vector $b \in \mathbb{R}^{l-1}$, $u = v_{l-1}' b / \sqrt{b' b}$ is a random variable with zero mean, unit variance, and a log-concave density. Therefore, by Theorem 1 in Adamczak, Latała, Litvak, Oleszkiewicz, Pajor, and Tomczak-Jaegermann (2014), there is an absolute constant $\kappa > 0$ such that $\mathbb{P}(|u| > \kappa t) \leq e^{-t}$ for all $t \geq 1$. This implies that $\mathbb{E}u^4 \leq C$ for some absolute constant $C > 0$. Hence, $\mathbb{E}|a' z_l|^4 \leq 8a_0^4 + 8(b' b)^2 C \leq 8 + 8C$ for any $a = (a_0, b)'$ with $a_0 \in \mathbb{R}$, $b \in \mathbb{R}^{l-1}$,

and $a_0^2 + b'b = 1$. By note 5, there is $c > 0$ such that $\mathbb{E}|z'_l a| \geq c$ for all $l \geq 1$ and $a \in \mathbb{R}^l$ with $a'a = 1$.

By Lemma 3.4(c), Lemma A.2, and Lemma A.3, $(\{z_l\}_{l \geq 1}, 1)$ satisfies Property P and Condition A holds. The rest of the proof follows the same lines as the proof of Theorem 4.2. ■

Proof of Theorem 4.4. We can normalize z_l to $z_l = (1, w'_{l-1})'$ for $w_{l-1} = (\mathbb{E}v_{l-1}v'_{l-1})^{-1/2}v_{l-1}$. Since $\lambda_{\max}((\mathbb{E}v_{l-1}v'_{l-1})^{-1/2}) \leq 1/\sqrt{\lambda}$, $v_{l-1} = F_{l-1}(u_{l-1})$, and F_{l-1} consists of c -Lipschitz functions, we can write $w_{l-1} = G_{l-1}(u_{l-1})$ for some $(c/\sqrt{\lambda})$ -Lipschitz map $G_{l-1} : \mathbb{R}^{l-1} \rightarrow \mathbb{R}^{l-1}$. Hence, by Lemma 3.4(d) and Lemma A.2, $(\{z_l\}_{l \geq 1}, 1)$ satisfies Property P.

Now, let us show that Condition A holds. Let $\varphi_b : \mathbb{R}^{l-1} \rightarrow \mathbb{R}$ be defined by $\varphi_b(x) = x'b$ for some $b \in \mathbb{R}^{l-1}$ with $b'b = 1$. Then φ_b is a 1-Lipschitz function. Arguing as in the proof of Lemma 3.4(d) and using Theorem 2.7 and Proposition 1.3 in Ledoux (2001), we arrive at the inequality $\mathbb{P}(|w'_{l-1}b - \text{med}(w'_{l-1}b)| > t) \leq 2 \exp\{-C_1 t^2\}$, $t > 0$, for some $C_1 = C_1(c/\sqrt{\lambda}, C) > 0$, where $\text{med}(\xi)$ is a median of a random variable ξ , and $C > 0$ satisfies $\lambda_{\max}(\text{var}(u_{l-1})) \leq C$. By Proposition 1.8 in Ledoux (2001), there are $C_k = C_k(C_1) > 0$, $k = 2, 3$, such that

$$\mathbb{P}(|w'_{l-1}b - \mathbb{E}(w'_{l-1}b)| > t) = \mathbb{P}(|w'_{l-1}b| > t) \leq C_3 \exp\{-C_2 t^2\}, \quad t > 0.$$

Hence, $\mathbb{E}(w'_{l-1}b)^4 \leq C_4(b'b)^2$ for some $C_4 > 0$ and all $b \in \mathbb{R}^{l-1}$. As a result,

$$\mathbb{E}(z'_l a)^4 \leq 8a_0^2 + 8\mathbb{E}(w'_{l-1}b)^4 \leq 8 + 8C_4$$

for all $a = (a_0, b)'$ with $a_0 \in \mathbb{R}$, $b \in \mathbb{R}^{l-1}$, and $a_0^2 + b'b = 1$. By note 5, $\mathbb{E}|z'_l a| \geq c$ for some $c > 0$ and all $a \in \mathbb{R}^l$ with $a'a = 1$. By Lemma A.3, Condition A holds.

The rest of the proof follows the same lines as the proof of Theorem 4.2. ■

Proof of Theorem 4.5. We can normalize z_l so that $z_l = (1, v'_{l-1})'$ satisfies $\mathbb{E}v_{l-1}v'_{l-1} = I_{l-1}$ (and $\mathbb{E}z_l z'_l = I_l$, since $\mathbb{E}v_{l-1}$ has zero entries), where $v_{l-1} = (\xi_1, \dots, \xi_{l-1})'$ and each ξ_j is an infinite linear combination (in L_2) of $\{e_k\}_{k \geq 1}$. By construction, there is an $(l-1) \times \infty$ matrix Γ_{l-1} such that $v_{l-1} = \Gamma_{l-1}e$ (in L_2) and $I_{l-1} = \mathbb{E}v_{l-1}v'_{l-1} = \mathbb{E}\Gamma_{l-1}ee'\Gamma'_{l-1} = \Gamma_{l-1}\Gamma'_{l-1}$, where $e = (e_1, e_2, \dots)'$.

Hence, for any $a = (a_0, \dots, a_{l-1})' \in \mathbb{R}^l$ with $a'a = 1$, there is $b = (b_1, b_2, \dots)'$ with $a_0^2 + b'b = 1$ such that $z'_l a = a_0 + e'b$. One can take $b = \Gamma'_{l-1}(a_1, \dots, a_{l-1})'$. The rest of the proof follows the same lines as the proof of Theorem 4.2. ■

Proof of Theorem 4.6. The proof is similar to the proof of Theorem 4.5. ■

Proof of Theorem 5.1. The instrument matrix Z is degenerate with positive probability. If $z_{lk} = (d_{k1}, \dots, d_{kl})'$, $1 \leq k \leq n$, then $d_{kq}d_{kr} = 0$ for all k, q, r with

$q \neq r$. As a result, $(Z'Z)^+$ is a diagonal matrix with diagonal entries

$$s_q = \begin{cases} 1/\sum_{k=1}^n d_{kq} & \text{if } \sum_{k=1}^n d_{kq} > 0, \\ 0 & \text{otherwise,} \end{cases} \quad q = 1, \dots, l.$$

Consider the case $k = 1$ without loss of generality. We have

$$P_{11} = z'_{11}(Z'Z)^+z_{11} = \sum_{q=1}^l \frac{d_{1q}}{\sum_{k=2}^n d_{kq} + d_{1q}},$$

where $0/0 = 0$. Let us show that $P_{11} \xrightarrow{d} 1/(\xi + 1)$, where $\xi \sim \text{Po}(1/a)$. By Poisson's limit theorem,

$$\begin{aligned} \mathbb{P}(P_{11} \leq x) &= \sum_{q=1}^l \mathbb{P}\left(\frac{1}{\sum_{k=2}^n d_{kq} + 1} \leq x, d_{1q} = 1\right) \\ &= \mathbb{P}\left(\frac{1}{\sum_{k=2}^n d_{k1} + 1} \leq x\right) \rightarrow \mathbb{P}\left(\frac{1}{\xi + 1} \leq x\right) \end{aligned}$$

for almost all $x \geq 0$, where we used that $n\mathbb{P}(d_{11} = 1) = n/l \rightarrow 1/a$. Hence, $P_{11} \xrightarrow{d} \zeta$.

Consider normalized $z_l = \sqrt{l}(d_1, \dots, d_l)'$. Then $\mathbb{E}z_l z'_l = I_l$, $\lambda_{\max}(\mathbb{E}z_l z'_l) = 1$, and $\lambda_{\min}^*(Z'Z) \geq l$. By Lemma A.5 and (5), $\alpha_n \xrightarrow{P} \lim_{n \rightarrow \infty} \mathbb{E}P_{11} = \mathbb{E}\zeta$ and

$$\Delta_n = \frac{1}{n} \sum_{k=1}^n P_{kk}^2 - \alpha_n^2 \xrightarrow{P} \lim_{n \rightarrow \infty} \text{var}(P_{11}) = \text{var}(\zeta). \quad \blacksquare$$

Proof of Corollary 5.2. Note that P_{11} is same as in the proof of Theorem 5.1. Let us demonstrate that $P_{11} \xrightarrow{d} 1/(\xi + 1)$, where ξ is given in Corollary 5.2. Note that P_{11} takes values $1/(m + 1)$ with $m = 0, 1, \dots$. By the independence of d_{1q} and $\{d_{kq}\}_{k=2}^n$, we have

$$\mathbb{P}\left(P_{11} = \frac{1}{m + 1}\right) = \sum_{q=1}^l \mathbb{P}\left(\frac{1}{\sum_{k=2}^n d_{kq} + 1} = \frac{1}{m + 1}, d_{1q} = 1\right) = \sum_{q=1}^l \mathbb{P}(\xi_q = m)p_q,$$

where ξ_q follows the Binomial distribution with parameters $(n - 1, p_q)$ and

$$p_q = \int_{(q-1)/l}^{q/l} f(u) du.$$

As f is bounded by say $C > 0$, we conclude that $p_q \leq C/l$ for all q . Therefore, by the Le Cam inequality,

$$\sum_{m=0}^{\infty} \left| \mathbb{P}(\xi_q = m) - e^{-(n-1)p_q} \frac{((n-1)p_q)^m}{m!} \right| \leq 2np_q^2 \leq 2n \left(\frac{C}{l}\right)^2 = o(1). \quad (6)$$

Thus,

$$\mathbb{P}\left(P_{11} = \frac{1}{m+1}\right) = \sum_{q=1}^l p_q e^{-(n-1)p_q} \frac{((n-1)p_q)^m}{m!} + o(1).$$

Fix m and make $g_m(u) = \exp\{-u\}u^m/m!$, $u \geq 0$. Also, define

$$f_n(u) = l \int_{(q-1)/l}^{q/l} f(v) dv, \quad (q-1)/l \leq u < q/l, \tag{7}$$

for $q = 1, \dots, l$ and $f_n(1) = 0$. Write

$$\sum_{q=1}^l p_q e^{-(n-1)p_q} \frac{((n-1)p_q)^m}{m!} = \int_0^1 g_m((n-1)f_n(u)/l) f(u) du.$$

By the Lebesgue differentiation theorem, $(n-1)f_n(u)/l \rightarrow f(u)/\alpha$ for almost all $u \in [0, 1]$. Hence, by the Lebesgue dominated convergence theorem,

$$\begin{aligned} \mathbb{P}\left(P_{11} = \frac{1}{m+1}\right) &= \int_0^1 g_m((n-1)f_n(u)/l) f(u) du + o(1) \\ &= \int_0^1 g_m(f(u)/\alpha) f(u) du + o(1). \end{aligned}$$

The latter is equivalent to $P_{11} \xrightarrow{d} 1/(\xi + 1)$.

The rest of the proof follows the same lines as the proof of Theorem 5.1. We only need to note that, for normalized $z_l = \sqrt{l}(d_1, \dots, d_l)'$, $\lambda_{\min}^*(Z'Z) \geq l$ and $\lambda_{\max}(\mathbb{E}z_l z_l') \leq C$, where C gives an upper bound for $f = f(u)$. ■

Proof of Theorem 5.3. We further assume that $z_l = du_l$ is normalized so that $\mathbb{E}u_l u_l' = I_l$. By Lemma 3.4 and Lemma A.2, $(\{z_l\}_{l \geq 1}, d)$ satisfies Property P (recall that $d^2 = d$). It was verified in the proofs of Theorems 4.2 through 4.6 that $\mathbb{E}|u_l' a| \geq c_0$ for some $c_0 > 0$, all $l \geq 1$, and any $a \in \mathbb{R}^l$ with $a' a = 1$. Then, there is $c > 0$ such that $\mathbb{E}|z_l' a| \mathbb{I}_{\{d > 0\}} = \mathbb{P}(d = 1) \mathbb{E}|u_l' a| \geq c$. Therefore, by Lemma A.3, Condition A holds.

Set $\pi = \mathbb{P}(d = 1)$. By Theorem 3.2,

$$P_{kk} \xrightarrow{p} \frac{cd_k}{1 + cd_k} = \frac{ad_k}{\pi}$$

for any fixed k , where $c > 0$ solves $\mathbb{E}(cd)/(1 + cd) = \pi c/(1 + c) = \alpha$. In addition, by Lemma A.5 and (5), $\alpha_n \xrightarrow{p} \lim_{n \rightarrow \infty} \mathbb{E}P_{11} = \alpha$ and

$$\Delta_n = \frac{1}{n} \sum_{k=1}^n P_{kk}^2 - \alpha_n^2 \xrightarrow{p} \lim_{n \rightarrow \infty} \text{var}(P_{11}) = \alpha^2 \frac{1 - \pi}{\pi}. \quad \blacksquare$$

Proof of Corollary 5.4. Inequalities $\sum_{j=1}^m d_{jk} \leq 1$ that hold for any fixed k imply that $d_{ik}d_{jk} = 0$ for all i, j, k with $i \neq j$. As a result, $Z'Z$ is a block-diagonal matrix with diagonal blocks Z'_jZ_j , $j = 1, \dots, m$, having the same structure as $Z'Z$ in Theorem 5.3, with d_k replaced by d_{jk} , respectively. Hence, by $d_{jk}^2 = d_{jk}$,

$$P_{kk} = z'_{lk}(Z'Z)^{-1}z_{lk} = \sum_{j=1}^m d_{jk}u'_{lk} \left(Z'_jZ_j \right)^{-1} u_{lk}.$$

By Theorem 5.3, $d_{jk}u'_{lk}(Z'_jZ_j)^{-1}u_{lk} \xrightarrow{P} \alpha d_{jk}/\mathbb{P}(d_j = 1)$ for any $j = 1, \dots, m$. Since m is fixed, we get

$$P_{kk} \xrightarrow{P} \sum_{j=1}^m \frac{\alpha d_{jk}}{\mathbb{P}(d_j = 1)}.$$

As shown in the proof of Theorem 5.3, Condition A holds for each block Z'_jZ_j . Since the number of blocks is finite, Condition A holds for $Z'Z$. By Lemma A.5 and (5), $\alpha_n \xrightarrow{P} \lim_{n \rightarrow \infty} \mathbb{E}P_{11} = \alpha m$ and

$$\Delta_n = \frac{1}{n} \sum_{k=1}^n P_{kk}^2 - \alpha_n^2 \xrightarrow{P} \lim_{n \rightarrow \infty} \text{var}(P_{11}) = \sum_{j=1}^m \frac{\alpha^2}{\mathbb{P}(d_j = 1)} - (\alpha m)^2. \quad \blacksquare$$

Proof of Theorem 5.6. Write $z_l = (w'_5, v u'_{l-5})'$ for $w_5 = (1, v, v^2, v^3, v^4)'$ and $u_{l-5} = (d_1, \dots, d_{l-5})'$. We may further assume that w_5 and u_{l-5} are normalized to $\mathbb{E}z_l z'_l = I_l$ by the linear transformation $(w_5, v u_{l-5}) \rightarrow (\bar{w}_5, v \bar{u}_{l-5})$, where $\bar{w}_5 = (\mathbb{E}w_5 w'_5)^{-1/2} w_5$ and $\bar{u}_{l-5} = (2d_1 - 1, \dots, 2d_{l-5} - 1)'$.

By Lemmas A.2 and 3.4, $(\{z_l\}_{l \geq 1}, v^2)$ satisfies Property P. By the Khintchin inequality, for any $a \in \mathbb{R}^5$ and $b = (b_1, \dots, b_{l-5})' \in \mathbb{R}^{l-5}$ with $a'a + b'b = 1$,

$$\mathbb{E}(a'w_5 + vb'u_{l-5})^4 \leq 8\mathbb{E}(a'w_5)^4 + 8\mathbb{E}v^4 \mathbb{E} \left(\sum_{q=1}^{l-5} b_q (2d_q - 1) \right)^4 \leq 8\mathbb{E}(w'_5 w_5)^2 + C$$

for some constant $C > 0$. This and note 5 imply that $\mathbb{E}|z'_l a| = \mathbb{E}|z'_l a| \mathbb{I}_{\{|v|>0\}} \geq c$ for some $c > 0$, all $l \geq 1$, and $a \in \mathbb{R}^l$ with $a'a = 1$. Hence, by Lemma A.3, Condition A holds.

Now, by Theorem 3.2, $P_{kk} \xrightarrow{P} cv_k^2/(1 + cv_k^2)$ for any fixed k , where $c > 0$ solves the equation $\mathbb{E}(cv^2)/(1 + cv^2) = \alpha$. Finally, by Lemma A.2 and (5),

$$\alpha_n \xrightarrow{P} \lim_{n \rightarrow \infty} \mathbb{E}P_{11} = \mathbb{E} \frac{cv^2}{1 + cv^2} = \alpha$$

and

$$\Delta_n = \frac{1}{n} \sum_{k=1}^n P_{kk}^2 - \alpha_n^2 \xrightarrow{P} \lim_{n \rightarrow \infty} \text{var}(P_{11}) = \text{var} \left(\frac{cv^2}{1 + cv^2} \right). \quad \blacksquare$$

Proof of Theorem 5.7. We need to introduce a few facts that derive from the definition of B-splines. First, setting $d_q = 1$ if $(q - 1)/l \leq z < q/l$ and $d_q = 0$ otherwise, we see that $d_q d_r = 0$ for all $q \neq r$. Second, let $\{\tilde{u}_q\}_{q=1}^l$ be independent random variables such that each \tilde{u}_q is independent of z and distributed as $q - lz$ conditionally on $d_q = 1$. That is, the density of \tilde{u}_q is equal to $f((q - u)/l)/l$, $u \in [0, 1]$, where $f(u)$ is the density of z . Then each variable

$$\bar{u}_q = \begin{cases} q - lz & \text{if } d_q = 1, \\ \tilde{u}_q & \text{if } d_q = 0, \end{cases}$$

has the same distribution as \tilde{u}_q , and is independent of d_q . Moreover, $z_l = (\bar{u}_1 d_1, (1 - \bar{u}_1) d_1, \bar{u}_2 d_2, (1 - \bar{u}_2) d_2, \dots, (1 - \bar{u}_l) d_l)'$ for linear B-splines.

Suppose now that $((\bar{u}_{kq}, d_{kq}))_{q=1}^l$, $k = 1, \dots, n$, are IID copies of $((\bar{u}_q, d_q))_{q=1}^l$. Then

$$z_{lk} = (\bar{u}_{k1} d_{k1}, (1 - \bar{u}_{k1}) d_{k1}, \dots, (1 - \bar{u}_{kl}) d_{kl})',$$

$k = 1, \dots, n$, are IID copies of z_l .

As the instrument matrix Z is degenerate with positive probability, set $P = Z(Z'Z)^+ Z'$. Since $d_{kq} d_{kr} = 0$ for all k, q, r with $q \neq r$, $(Z'Z)^+$ is a block diagonal matrix with 2×2 diagonal blocks $(S_q)^+$, where

$$S_q = \sum_{k=1}^n d_{kq} v_{kq} v'_{kq}$$

and $v_{kq} = (\bar{u}_{kq}, 1 - \bar{u}_{kq})'$, $q = 1, \dots, l$. Consider, without loss of generality, P_{kk} for $k = 1$:

$$P_{11} = z'_{11} (Z'Z)^+ z_{11} = \sum_{q=1}^l d_{1q} v'_{1q} \left(\sum_{k=1}^n d_{kq} v_{kq} v'_{kq} \right)^+ v_{1q}.$$

Fix $x \geq 0$. Using the independence of v_{1q} and d_{1q} , we deduce that $\mathbb{P}(P_{11} \leq x) = I$, where $I = \sum_{q=1}^l \mathbb{P}(A_q) \mathbb{P}(d_{1q} = 1)$, and

$$A_q = \left\{ v'_{1q} \left(\sum_{k=2}^n d_{kq} v_{kq} v'_{kq} + v_{1q} v'_{1q} \right)^+ v_{1q} \leq x \right\}. \tag{8}$$

If $z \sim U[0, 1]$, then $\{(d_{kq}, v_{kq})\}_{k=1}^n$ are identically distributed over q . In particular, $\mathbb{P}(A_q)$ is constant over q , and $I = \mathbb{P}(A_1)$ since $\mathbb{P}(d_{1q} = 1) = 1/l$. Because (d_{11}, v_{11}) is independent of $\sum_{k=2}^n d_{k1} v_{k1} v'_{k1}$ and because $(d_{k1})_{k=1}^n$ and $(v_{k1})_{k=1}^n$ are independent collections of IID random elements, we have

$$I = \mathbb{P} \left(v' \left(\sum_{j=1}^{\xi_n} \tilde{v}_j \tilde{v}'_j + v v' \right)^+ v \leq x \right),$$

where $\xi_n = \sum_{j=2}^n d_{j1}$ is independent of all other random variables involved, and $(v, \tilde{v}_1, \tilde{v}_2, \dots)$ are IID copies of v_{11} . Next, if $z \sim U[0, 1]$, then v has the same distribution as $(z, 1 - z)'$. In addition, by the Poisson limit theorem, $\xi_n \xrightarrow{d} \xi \sim \text{Po}(1/\alpha)$ since $n\mathbb{P}(d_{11} = 1) \rightarrow 1/\alpha$. Hence,

$$\begin{aligned} I &= \sum_{m=1}^{\infty} \mathbb{P} \left(v' \left(\sum_{j=1}^m \tilde{v}_j \tilde{v}'_j + vv' \right)^+ v \leq x \right) \mathbb{P}(\xi_n = m) + \mathbb{P}(v'V(0)v \leq x) \mathbb{P}(\xi_n = 0) \\ &= \sum_{m=1}^{\infty} \mathbb{P} \left(v' \left(\sum_{j=1}^m \tilde{v}_j \tilde{v}'_j + vv' \right)^+ v \leq x \right) \mathbb{P}(\xi = m) \\ &\quad + \mathbb{P}(v'V(0)v \leq x) \mathbb{P}(\xi = 0) + o(1) \\ &= \mathbb{P}(v'V(\xi)v \leq x) + o(1). \end{aligned}$$

This finishes the proof. ■

NOTES

1. Notably, the example of efficiency ranking between LIML and HLIM provided on p.224 does not attain its point when the sample variance of diagonal elements is zero in the limit.
2. Anderson, Kunitomo and Matsushita (2010) find that “the effects of non-normality of disturbance terms on the cdf of the LIML estimator are often very small”. However, this may have occurred because their instrument design is asymptotically balanced due to joint normality of instruments, see Subsection 4.2.
3. Abutaliev and Anatolyev (2013) numerically evaluate the effects of asymptotic heterogeneity of the diagonal on the asymptotic variance of LIML estimates in a setting with indicator instruments and Skewed Student errors.
4. Recall that in a linear regression of y on z , the k^{th} diagonal element P_{kk} of P is the leverage score of the k^{th} observation, for the k^{th} fitted value of y is

$$\widehat{y}_k = P_{kk}y_k + \sum_{j \neq k} P_{kj}y_j.$$

If P_{kk} is too large, then y_k is badly overfit. In particular, if P_{kk} is close to one, then \widehat{y}_k is determined solely by one observation. So, for the collection $\{P_{kk}\}_{k=1}^n$, the sample mean α_n can be interpreted as the average leverage score, and the sample variance Δ_n measures how far the leverage scores are spread apart.

5. One can take $K = 1/16$ in Lemma 3.1 (see the proof of Corollary 3.4 in Yaskov, 2014; see also Yaskov, 2016), but it is far from optimal. In addition, since $\mathbb{E}(a'z_l)^2 = 1$ when $a'a = 1$, condition $\mathbb{E}|a'z_l| \geq c$ means that random variables $a'z_l$ are not too heavy-tailed. Indeed, if, for example, $\mathbb{E}u^2 = 1$ and $\mathbb{E}|u|^{2+\delta} \leq C$ for some $C, \delta > 0$, then $1 = \mathbb{E}|u|^{\frac{\delta}{1+\delta}} |u|^{\frac{2+\delta}{1+\delta}} \leq (\mathbb{E}|u|)^{\frac{\delta}{1+\delta}} (\mathbb{E}|u|^{2+\delta})^{\frac{1}{1+\delta}}$ and $\mathbb{E}|u| \geq c = C^{-1/\delta}$.

6. That is, a density whose logarithm is a concave function (setting $\log 0 = -\infty$).

7. A log-concave distribution in \mathbb{R}^d can be defined as a distribution supported on an affine subspace $H \subset \mathbb{R}^d$ of dimension $k \leq d$, where this distribution has a log-concave density w.r.t. the Lebesgue measure on this subspace (see Theorem 3.2 in Borell, 1975).

REFERENCES

- Abutaliev, A. & S. Anatolyev (2013) Asymptotic variance under many instruments: numerical computations. *Economics Letters* 118(2), 272–274.
- Adamczak, R., R. Latała, A.E. Litvak, K. Oleszkiewicz, A. Pajor, & N. Tomczak-Jaegermann (2014) A short proof of Paouris' inequality. *Canadian Mathematical Bulletin* 57, 3–8.
- Anatolyev, S. (2012) Inference in regression models with many regressors. *Journal of Econometrics* 170(2), 368–382.
- Anatolyev, S. (2013) Instrumental variables estimation and inference in the presence of many exogenous regressors. *Econometrics Journal* 16(1), 27–72.
- Anatolyev, S. & N. Gospodinov (2011) Specification testing in models with many instruments. *Econometric Theory* 27, 427–441.
- Angrist, J. & A. Krueger (1991) Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 106, 979–1014.
- Anderson, T.W., N. Kunitomo, & Y. Matsushita (2010) On the asymptotic optimality of the LIML estimator with possibly many instruments. *Journal of Econometrics* 157, 191–204.
- Anderson, T.W., N. Kunitomo, & Y. Matsushita (2011) On finite sample properties of alternative estimators of coefficients in a structural equation with many instruments. *Journal of Econometrics* 165, 58–69.
- Bagnoli, M. & T. Bergstrom (2005) Log-concave probability and its applications. *Economic Theory* 26, 445–469.
- Bai, Z. & J. Silverstein (2010) *Spectral analysis of large dimensional random matrices*, 2nd edition. Springer.
- Bekker, P.A. (1994) Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681.
- Bekker, P.A. & F. Crudu (2015) Jackknife instrumental variable estimation with heteroskedasticity. *Journal of Econometrics* 185(2), 332–342.
- Bekker, P.A. & J. van der Ploeg (2005) Instrumental variable estimation based on grouped data. *Statistica Neerlandica* 59, 239–267.
- Borell, C. (1975) Convex set functions in d-space. *Periodica Mathematica Hungarica* 6(2), 111–136.
- Calhoun, G. (2012) Hypothesis testing in linear regression when k/n is large. *Journal of Econometrics* 165(2), 368–382.
- Davidson, R. & J.G. MacKinnon (2006) The case against JIVE. *Journal of Applied Econometrics* 21(6), 827–833.
- Hahn, J. & A. Inoue (2002) A Monte Carlo comparison of various asymptotic approximations to the distribution of instrumental variables estimators. *Econometric Reviews* 21, 309–336.
- Hansen, C., J. 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.C. Chao, & N.R. Swanson (2012) Instrumental variable estimation with heteroskedasticity and many instruments. *Quantitative Economics* 3, 211–255.
- Kolesár, M. (2015) Minimum distance approach to inference with many instruments. manuscript. Princeton University.
- Kunitomo, N. (2012) An optimal modification of the LIML estimation for many instruments and persistent heteroscedasticity. *Annals of the Institute of Statistical Mathematics* 64(5), 881–910.
- Ledoux, M. (2001) The concentration of measure phenomenon. *Mathematical Surveys and Monographs*, 89. American Mathematical Society.
- Lee, Y. & R. Okui (2012) Hahn–Hausman test as a specification test. *Journal of Econometrics* 167(1), 133–139.
- Meurant, G. (1992) A review on the inverse of symmetric tridiagonal and block tridiagonal matrices. *SIAM Journal on Matrix Analysis and Applications* 13(3), 707–728.
- Pastur, L. & M. Shcherbina (2011) *Eigenvalue Distribution of Large Random Matrices*. Mathematical Surveys and Monographs, 171. American Mathematical Society.

- Saumard, A. & J.A. Wellner (2014) Log-concavity and strong log-concavity: A review. *Statistics Surveys* 8, 45–114.
- Sherman, J. & W.J. Morrison (1950) Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *Annals of Mathematical Statistics* 21, 124–127.
- van Hasselt, M. (2010) Many instruments asymptotic approximations under nonnormal error distributions. *Econometric Theory* 26, 633–645.
- Wang, W. & M. Kaffo (2016) Bootstrap inference for instrumental variable models with many weak instruments. *Journal of Econometrics* 192(1), 231–268.
- Yaskov, P.A. (2014) Lower bounds on the smallest eigenvalue of a sample covariance matrix. *Electronic Communications in Probability* 19, article 83.
- Yaskov, P.A. (2016) Controlling the least eigenvalue of a random Gram matrix. *Linear Algebra and its Applications* 504, 108–123.