CERGE UK

Univerzita Karlova
Centrum pro ekonomický výzkum a doktorské studium

P.O.Box 882, Politických vězňů 7, CZ 111 21 Praha 1
Tel.: (+420) 224 005 161, Fax: (+420) 224 005 333
study@cerge-ei.cz, http://www.cerge-ei.cz

V Praze, 4.3.2024

POTVRZENÍ O STUDIU

Potvrzujeme tímto, že Maksim Smirnov, narozen 18.7.1998, je v akademickém roce 2023/2024 (1.10.2023-30.9.2024) studentem 2. ročníku prezenčního doktorského studia na Univerzitě Karlově – Centru pro ekonomický výzkum a doktorské studium (CERGE UK).

In Prague, 4.3.2024

CONFIRMATION OF STUDY

This is to confirm that Maksim Smirnov, born on 18.7.1998, is a full-time 2nd-year student of the PhD program at Charles University – Center for Economic Research and Graduate Education (CERGE UK) in the academic year 2023/2024 (1.10.2023-30.9.2024).

UNIVERZITA KARLOVA
CERGE
Politických vězňů 7
111 21 Praha 1
IČO: 00216208, DIČ: CZ00216208
(2)

Tereza Kulhánková
Referentka studijního oddělení
Study Affairs Office Assistant

Upozornění: Tento dokument je platný pouze s úředním podpisem a razítkem UK.
Notice: This document is valid only with a signature and an official stamp of the Charles University.

Many Instruments Estimation and Inference under Clustered Dependence

Stanislav Anatolyev[1] Maksim Smirnov[2]

[1]CERGE-EI and NES
[2]CERGE-EI

This version is very preliminary, please do not cite

Abstract

The literature on many weak instruments in a heteroskedastic environment under data independence is largely developed. When data dependence, in particular clustering, is present, it poses difficulties in making correct and convenient inferences. We show that clustering either deems the jackknife instrumental variables estimation inconsistent, or makes its inferences hugely distorted. We suggest, instead of following the "save the Jackknife" approach, an alternative approach, which is computationally attractive and allows general structures of intra-cluster correlations. We use the natural extension of jackknifing, the leave-cluster-out methodology, applied to the instrument projection matrix, which allows one to dispose of the cross-cluster dependencies in the influence function of the structural parameter estimator. We set out a formal asymptotic framework to analyze the proposed cluster-jackknife instrumental variables (CJIV) estimator, with an increasing number of clusters, possibly increasing heterogeneous cluster sizes, and possible presence of many weak instruments. We prove a central limit theorem for the influence function embedded in the CJIV estimator, and show consistency of the associated CJIV variance estimator. We study the importance of instrument design on the properties of CJIV, run a simulation study revealing its finite sample properties, and compare with other estimators in relevant empirical contexts.

Keywords: instrumental variables, many instruments, clustered dependence

1 Introduction

Two-stage least squares estimator is not suitable for many instruments settings (see e.g. Bekker, 1994), while jackknife instrumental variables estimator (JIVE) is consistent and asymptotically normal under many instruments and independent data (Angrist et al., 1999, Chao et al., 2012). In clustered environments under reasonable assumptions JIVE is still consistent; however, asymptotic inferences are distorted. Furthermore, in certain relevant settings clustered dependence can even render JIVE inconsistent. In this paper we show that for JIVE the asymptotic bias and distortions in inferences may be huge, especially when clusters are big (and unboundedly increase) and the instrumental design has a certain degree of cluster dependence. Further, we propose an alternative estimator based on leave-cluster-out (LCO) methodology that is robust to many weak instruments under clustering and show its asymptotic properties.

A number of approaches have been proposed in order to mitigate the negative consequences of the instrumental variables multiplicity. Donald and Newey (2001) suggested a corrected version of the two-stage least squares estimator, which is robust to many instruments bias. Bekker (1994) showed that limited information maximum likelihood estimator (LIML) is consistent under random sampling. C. Hansen et al. (2008) and van Hasselt (2010) extended the asymptotic theory by showing asymptotic normality of LIML under arbitrary distributions of errors. Fuller estimator (Fuller, 1977), which is a corrected version of LIML estimator, is also often used under many instruments.

In this paper, we focus on another approach, which is to employ the leave-out jackknife idea to deal with the negative impact of many instruments. Angrist et al. (1999) and Blomquist and Dahlberg (1999) proposed versions of a jackknife IV estimator that are robust to many instruments. Later Chao et al. (2012) provided a concrete asymptotic analysis of JIVE under many instruments and random sampling. Our proposed cluster-jackknife instrumental variables (CJIV) estimator generalizes the idea of JIVE and the analysis of Chao et al. (2012) to contexts with clustering.

There are, to the best of our knowledge, only three papers that tried to tackle the clustered dependence under many instruments, namely Hausman et al. (2011), Chao et al. (2023) and Ligtenberg (2023). Hausman et al. (2011) suggest computationally intensive penalization methods, the main goal being to achieve the existence of moments. Later, Carrasco (2012) noted that the existence of moments and consistency of their procedure are not simultaneously possible. The other stab at the problem in Chao et al. (2023) proposes filtering out the individual effects in a within-type transformation and applying the jackknife methodology on top of that. However, this proposal is tied to the classical one-way error component model (ECM) structure of regression errors, and is not expected to work for more general within-cluster dependencies. We suggest another approach to tackle the problem, which is computationally attractive and is not tied to a rigid ECM structure. Instead, it allows a more general setup, where the structural and reduced form errors may be conditionally heteroskedastic and heterocorrelated in an arbitrary unknown form. Finally, Ligtenberg (2023) proposes an AR-type test that is suitable for many instruments and clustering. Conversely, we propose not a testing procedure but the whole framework for both estimation and inference that is robust to many weak instruments and clustering, and can be used in other contexts (e.g. few instruments with cluster dependence). Similar to Ligtenberg

(2023), instead of jackknifing machinery, we use its natural extension – LCO methodology, applied to the matrix of projection on the space of instruments. However, we push it further by developing both inference and estimation in a broad class of problems. In the present instrumental variables setup, LCO allows one to dispose of the very source of problems – intra-cluster dependencies in the influence function of the structural parameter's estimator.

Both our proposed CJIV and the associated CJIV variance estimators are computationally simple and, importantly, allow general structures of intra-cluster correlations. We also work towards the vectorization of the formula for the CJIV variance estimator to reduce the computational burden. We set out a formal asymptotic framework with an increasing number of clusters, possibly increasing average cluster size, and possible presence of many weak instruments. Within this framework, we prove a central limit theorem for the influence function embedded in the CJIV estimator, thus establishing its normality, and prove the consistency of both CJIV and the associated CJIV variance estimators. Moreover, we argue that CJIV might be useful in setups even with few instruments that are correlated within clusters. In addition, we study the importance of instrument design on the properties of CJIV, run an extensive simulation study revealing finite sample properties and advantages of the proposed approach, apply CJIV methodology to setups from relevant empirical papers (Angrist and Krueger, 1991 and Autor et al., 2013), and evaluate its computational intensity.

In addition, we propose and analyze a modification of CJIV estimator, WCJIV, in which pairs of observations are weighted by sizes of corresponding clusters. This modification is more robust to cluster size heterogeneity and large clusters.

The remainder of this paper is structured as follows. In Section 2, we set up the model, discuss limitations of JIVE in terms of asymptotic bias and validity of inference, and propose our CJIV estimator. In Section 3, we present formal asymptotic theory of CJIV. In Section 4, we describe the variance estimator and provide some simulation evidence of its validity. In Section 5, we discuss possible generalizations of our LCO approach outside of many instruments clustering setup. In Section 6, we apply CJIV to empirical studies and compare its performance with 2SLS estimator and JIVE. Section 7 concludes.

2 Problems with JIVE, and the solution

2.1 Setup

We consider the following linear IV model:

$$\begin{aligned} y &= X\beta + \varepsilon, \\ X &= \Upsilon + u, \end{aligned}$$

where X is an $n \times L$ vector of the endogenous variables, β is the coefficient vector of interest. U is an $n \times L$ matrix of reduced-form disturbances, and Υ is an $n \times L$ matrix of reduced-form values.

While Υ is not observed, we assume that we observe an $n \times K$ matrix of instruments Z that approximate Υ well asymptotically. Given that instruments are many, and endogenous regressors are few, it is not a restrictive assumption.

We can conduct most of the analysis conditionally on $\mathcal{Z} = (Z, \Upsilon)$ in the spirit of Chao et al. (2012).

We introduce clustering by considering a known partition $\mathcal{G}$ of population into G mutually exclusive clusters with heterogeneous and asymptotically increasing sizes n_g, $g \in \mathcal{G}$ (in Section 5 we discuss possible generalization to multiway clustering), and allowing $\mathbb{E}[x_i y_j] \neq 0$ for $x \in \{\varepsilon, u\}$, $y \in \{\varepsilon, u\}$ if i and j are in the same cluster (i.e. $i \in g$ and $j \in g$ for some $g \in \mathcal{G}$). Throughout the paper we assume w.l.o.g. that observations are ordered so that they are grouped by cluster.

2.2 The tale of two examples

Let $P = Z\left(Z'Z\right)^{-1} Z'$, with p_{ij} denoting the (i,j)th element of P. $A_{[g]} = (A_i)_{i \in g}$ for some A.

Denote, by $\dot{P}$ the de-diagonalized P, that is, P with the main diagonal removed:

$$\dot{P} = \begin{bmatrix} 0 & p_{1,2} & \cdots & p_{1,n-1} & p_{1,n} \\ p_{1,2} & 0 & \cdots & p_{2,n-1} & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{1,n-1} & p_{2,n-1} & \cdots & 0 & p_{n-1,n} \\ p_{1,n} & p_{2,n} & \cdots & p_{n-1,n} & 0 \end{bmatrix}.$$

$\dot{P}$ is the key component of JIVE estimator. Also, denote by $\mathring{P}$ the LCO version of P, that is, P with diagonal blocks corresponding to clusters removed. For example, when $n_g = 2$ for all clusters,

$$\mathring{P} = \begin{bmatrix} 0 & 0 & p_{1,3} & p_{1,4} & \cdots & p_{1,n-1} & p_{1,n} \\ 0 & 0 & p_{2,3} & p_{2,4} & \cdots & p_{2,n-1} & p_{2,n} \\ p_{1,3} & p_{2,3} & 0 & 0 & \cdots & p_{3,n-1} & p_{3,n} \\ p_{1,4} & p_{2,4} & 0 & 0 & \cdots & p_{4,n-1} & p_{4,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{1,n-1} & p_{2,n-1} & p_{3,n-1} & p_{4,n-1} & \cdots & 0 & 0 \\ p_{1,n} & p_{2,n} & p_{3,n} & p_{4,n} & \cdots & 0 & 0 \end{bmatrix}.$$

We consider two running examples, in which different instrument designs induce different properties of estimators. In both, the clusters are equally sized, so n_g is constant across g.

Example DI: dummy supergroup instruments. Suppose the number of groups G is proportional to S, number of supergroups (for example, industries and firms, respectively, with equally-sized industries), so that one supergroup covers G/S clusters. Let $n(s)$ be the size of supergroup s, and let all supergroups and all clusters be equally-sized for the simplicity of exposition. Then, $Z = I_S \otimes \iota_{n(s)}$, leading to $Z'Z = n(s) I_S$, $(Z'Z)^{-1} = I_S/n(s)$ and hence $P = (1/n(s)) I_s \otimes (\iota_{n(s)} \iota'_{n(s)})$. This matrix is block-diagonal, but blocks are $n(s) \times n(s)$, so LCO yields non-null but block-sparse $\mathring{P}$, which contains $G/S\,(G/S - 1)$ off-diagonal blocks of $(n(s) - n_g) \times (n(s) - n_g)$ squares of $1/n(s)$. By construction, $\ell = S$; instruments are moderately many if S is big. When $S = G/2$ (only two groups belong to a supergroup), $P = 1/(2n_g) I_{G/2} \otimes (\iota_{2n_g} \iota'_{2n_g})$, and $\mathring{P}$ contains half of non-zero $2n_g \times 2n_g$ blocks of P; when $G/S > 2$ (i.e. supergroups contain more clusters), a larger proportion of blocks are retained in $\mathring{P}$ (see Supplementary appendix for details).

Example RI: random independent instruments. Here, the elements of Z except the first one are i.i.d. standard normal, and the first one equals 1. Then, non-diagonal elements of P, by symmetry, possess the same marginal distribution.

2.3 Three cases against JIVE

Akin to Davidson and MacKinnon (2006)'s title, we point at futility of relying on JIVE in the clustered data setup. The JIVE estimator reads

$$\hat{\beta}_{JIVE} = \left(X'\dot{P}X\right)^{-1} X'\dot{P}Y = \left(\sum_{i\neq j} p_{ij}X_iX_j'\right)^{-1} \sum_{i\neq j} p_{ij}X_iY_j.$$

Then,

$$\sqrt{n}\left(\hat{\beta}_{JIVE} - \beta\right) = \left(H_{JIVE} + o\left(1\right)\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i\neq j} p_{ij}X_i\varepsilon_j,$$

where

$$H_{JIVE} = \text{plim}\frac{1}{n}\sum_{i\neq j} p_{ij}X_iX_j'$$

is assumed to be non-singular.

The JIVE, by dropping the terms related to correlations of endogenous regressors with the structural error for the same unit, removes the adverse effects of endogeneity when many instruments are present, but gives up on full utilization of information corresponding to that unit. While this information is possible to utilize under strong restrictions like conditional homoskedasticity, which is what bias-corrected 2SLS and LIML do, it appears impossible to accomplish in a non-restricted environment. Thus, in return JIVE buys the robustness to conditional heteroskedasticity under many instruments.

We consider expansions of the JIVE influence function into linear and quadratic components

$$\frac{1}{\sqrt{n}}\sum_{i\neq j} p_{ij}X_i\varepsilon_j = \frac{1}{\sqrt{n}}\sum_{i}\left(1-p_{ii}\right)\Upsilon_i\varepsilon_i + \frac{1}{\sqrt{n}}\sum_{i\neq j} p_{ij}u_i\varepsilon_j \equiv A_{1n}^{JIVE} + A_{2n}^{JIVE}.$$

The Supplementary Appendix presents this expansion in detail.

To simplify the further discussion, let the cluster sizes be equal, with $n = n_gG$, and all the errors are conditionally homoskedastic, within-cluster equicorrelated, i.e. $E\left[u_i\varepsilon_j\right] = \sigma_{u\varepsilon}\mathbb{I}_{\{C_i=C_j\}}$ for constant $\sigma_{u\varepsilon}$. We consider the two examples of instrument design that lead to different orders and hence asymptotic properties of the JIVE.

Example DI. For the dummy supergroup instruments example, the properties for the components in the expansions for the JIVE are[1]

$$E\left[A_{1n}^{JIVE}\right] = 0, \quad E\left[A_{2n}^{JIVE}\right] = O\left(\frac{sn_g}{\sqrt{n}}\right),$$

[1]see S1 in Appendix S

$$var\left(A_{1n}^{JIVE}\right)=O\left(n_g\right),\quad var\left(A_{2n}^{JIVE}\right)=O\left(\frac{sn_g^2}{n}\right).$$

It follows that

$$\sqrt{n}\left(\hat{\beta}_{JIVE}-\beta\right)=\left(H_{JIVE}+o\left(1\right)\right)^{-1}O_p\left(\frac{Sn_g}{\sqrt{n}}\right),$$

and

$$\hat{\beta}_{JIVE}=\beta+O_p\left(\frac{S}{G}\right),$$

so the JIVE is asymptotically biased in case a supergroup covers an asymptotically fixed number of clusters so that $S/G=O(1)$, be n_g asymptotically increasing or fixed. This asymptotic bias originates from the biasedness of the quadratic part of the JIVE influence function (A_{2n}^{JIVE}) under this particular instrument design, which arises from a lot of cross-cluster correlations among the structural and reduced-form errors falling into the same cluster, and corresponding weights p_{ij} not vanishing asymptotically because of the instrument design.

Example RI. For the random independent instruments example, the properties for the linear component in the expansion for the JIVE are the same:

$$E\left[A_{1n}^{JIVE}\right]=0,\quad var\left(A_{1n}^{JIVE}\right)=O\left(n_g\right).$$

As for the quadratic component, the variance has the order[2]

$$var\left(A_{2n}^{JIVE}\right)=O\left(n_g^2\right).$$

Furthermore, the linear component A_{1n}^{JIVE} and the quadratic component A_{2n}^{JIVE} are correlated.[3] Yet another feature of JIVE is that clustering induces a (higher-order) bias into quadratic component:[4]

$$E\left[A_{2n}^{JIVE}\right]=O\left(\frac{n_g}{\sqrt{n}}\right).$$

leading to an additional component of (higher-order) bias for the JIVE estimator of order $n^{-1/2}O\left(n_g/\sqrt{n}\right)=O\left(1/G\right)$. Thus, in this example, the JIVE is consistent, under some

[2]see S2 in Appendix S

[3]Indeed,

$$cov_{\mathcal{Z}}\left(A_{1n}^{JIVE},A_{2n}^{JIVE}\right)=\frac{1}{n}\sum_i\sum_{\substack{k\\C_k=C_i}}\sum_{\substack{\ell\neq k\\C_\ell=C_i}}E\left[\varepsilon_i u_k\varepsilon_\ell\right]\left(1-p_{ii}\right)\Upsilon_i p_{k\ell}\neq 0.$$

[4]see S3 in Appendix S

restriction on the rate of growth of n_g in case it is asymptotically increasing, but its influence function is not clean and contaminated by various factors that complicate modification and/or pivotization of the JIVE estimator.

In the two examples above, the orders of the bias term are different so much that they resulted in JIVE inconsistency in one case and its consistency in the other. The JIVE (in)consistency depends largely on the instrument design. Consider a more general conditionally heteroskedastic environment, and denote $\sigma_{ij}^{eu} = E\left[e_i u_j\right]$ for i and j within the same cluster. Then, a necessary condition for consistency of JIVE is[5]

$$\sum_{\substack{i \neq j \\ C_j = C_i}} p_{ij} \sigma_{ij}^{u\varepsilon} = o(n),$$

This does not hold in Example DI when a supergroup covers an asymptotically fixed number of clusters and $\sigma_{ij}^{u\varepsilon}$ does not depend on (i,j), so that $\sum_{i \neq j, C_j = C_i} p_{ij} = O(n)$. However, this condition does hold in Example RI. Heuristically, the instrument design in Example DI results in all non-zero elements in $\dot{P}$, although individually small, being of the same sign, so en mass they weigh much higher than an analogous sum of elements of $\dot{P}$ in Example RI, as these tend to be distributed around an asymptotically zero mean.[6]

Even when the JIVE is consistent, the use of the standard JIVE variance estimator not accounting for clustering can lead to big biases and inferential distortions when the clusters are asymptotically increasing, beyond those that could be expected when a wrong variance estimator is used. The JIVE pivotization is unable to keep up with asymptotically increasing cluster size.

To summarize, the three cases against JIVE are: (1) in some setups, JIVE is inconsistent; (2) when JIVE is consistent, under-independence JIVE inference is invalid; (3) even when JIVE is consistent, the linear and quadratic terms in the JIVE influence function are correlated and the quadratic term has a non-zero (higher-order) bias, which greatly complicates asymptotic derivations and pivotization. These properties of JIVE motivate one to deviate from such a construct instead of attempting to "save" it, which is arguably impossible if one does not impose strong restrictions on the within-cluster correlation structure.

2.4 Solution: CJIV

The CJIV estimator accommodates the LCO idea and excludes the terms that correspond to observations within the same cluster:

$$\hat{\beta}_{CJIV} = \left(X' \mathring{P} X\right)^{-1} X' \mathring{P} Y = \left(\sum_{\substack{i,j \\ C_i \neq C_j}} p_{ij} X_i X_j' \right)^{-1} \sum_{\substack{i,j \\ C_i \neq C_j}} p_{ij} X_i y_j.$$

[5]see S4 in Appendix S

[6]The sum of elements of $\dot{P}$ equals $\sum_{j \neq i} p_{ij} = \sum_{i,j} p_{ij} - \sum_i p_{ii} = n - \ell$, so their average is $(n-\ell) / \left(n^2 - n\right)$, which is asymptotically $O\left(1/n\right)$.

Then,

$$\sqrt{n}\left(\hat{\beta}_{CJIV}-\beta\right)=\left(H_{CJIV}+o\left(1\right)\right)^{-1}\frac{1}{\sqrt{n}}\sum_{\substack{i,j\\C_i\neq C_j}}p_{ij}X_i\varepsilon_j,$$

where

$$H_{CJIV}=\text{plim}\frac{1}{n}\sum_{\substack{i,j\\C_i\neq C_j}}p_{ij}X_iX_j'$$

is assumed to be non-singular.

The CJIV estimator goes farther than the JIVE and drops the terms related to correlations of endogenous regressors with the structural errors for the same cluster, gives up on full utilization of information corresponding to that cluster. As a result, estimation and inference is robust to arbitrary patterns of within-cluster conditional heteroskedasticity and heterocorrelation. More generally, if the dependence structure is known to a researcher, one can remove arbitrary elements of P, so that the resulting matrix gives zero weight to pairs of observations that might be correlated (see Section 5).

Example DI. For the dummy supergroup instruments example, the JIVE removes n nonzero diagonal elements from P, while the CJIV, out of $S\left(n(s)\right)^2-n=n\left(n(s)-1\right)$ off-diagonal elements, removes $G\left(n_g\right)^2-n=n\left(n_g-1\right)$ more elements. The proportion of elements removed by the CJIV on top of elements removed by the JIVE is $\left(n_g-1\right)/\left(n(s)-1\right),$ which is asymptotically proportional to S/G, and is increasing with S.

Example RI. For the random independent instruments example, the CJIV, out of $n^2-n=n\left(n-1\right)$ off-diagonal elements, removes $G\left(n/G\right)^2-n=n\left(n/G-1\right)$ more elements, the proportion of $\left(n/G-1\right)/\left(n-1\right)$, which is asymptotically inversely proportional to G.

The CJIV influence function contains two terms:

$$\frac{1}{\sqrt{n}}\sum_{\substack{i,j\\C_i\neq C_j}}p_{ij}X_i\varepsilon_j=\frac{1}{\sqrt{n}}\sum_{\substack{i,j\\C_i\neq C_j}}p_{ij}\Upsilon_i\varepsilon_j+\frac{1}{\sqrt{n}}\sum_{\substack{i,j\\C_i\neq C_j}}p_{ij}u_i\varepsilon_j\equiv A_{1n}^{CJIV}+A_{2n}^{CJIV}.$$

The orders of these two terms sharply differ from those for JIVE, hence resulting in different asymptotic properties of the CJIV. In particular, LCO removes any bias in A_{2n}^{CJIV}:

$$E\left[A_{2n}^{CJIV}\right]=\frac{1}{\sqrt{n}}\sum_{\substack{i,j\\C_j\neq C_i}}p_{ij}E\left[u_i\varepsilon_j\right]=0,$$

because $E\left[u_ie_j\right]=0$ whenever i and j belong to different clusters. Also, LCO makes the two components A_{1n}^{CJIV} and A_{2n}^{CJIV} uncorrelated:

$$cov\left(A_{1n}^{CJIV},A_{2n}^{CJIV}\right)=\frac{1}{n}\sum_{\substack{i,j\\C_j\neq C_i}}p_{ij}\Upsilon_i\sum_{\substack{k,\ell\\C_\ell\neq C_k}}p_{k\ell}E\left[\varepsilon_ju_k\varepsilon_\ell\right]=0,$$

because $E[\varepsilon_j u_k \varepsilon_\ell] = 0$ whenever k and ℓ belong to different clusters, no mater which cluster j belongs to. We consider the two examples of instrument design, across which, similarly to JIVE, the variance of the quadratic component differs.

Example DI. For the dummy supergroup instruments example, the variance properties for the components in the expansions for the CJIV are[7]

$$var\left(A_{1n}^{CJIV}\right) = O\left(n_g\right), \quad var\left(A_{2n}^{CJIV}\right) = O\left(\frac{sn_g}{G}\right) = O\left(n_g\right).$$

It follows that

$$\hat{\beta}_{CJIV} = \beta + O_p\left(\frac{1}{\sqrt{G}}\right),$$

so the CJIV is $\sqrt{G}$-consistent, provided that G is asymptotically increasing. Recall that JIVE is inconsistent because of the bias in A_{2n}^{JIVE} is of a rather high order.

Example RI. For the random independent instruments example, the variance properties for the components in the expansions for the CJIV are[8]

$$var\left(A_{1n}^{CJIV}\right) = O\left(n_g\right), \quad var\left(A_{2n}^{CJIV}\right) = O\left(n_g^2\right).$$

It follows that

$$\hat{\beta}_{CJIV} = \beta + O_p\left(\frac{n_g}{\sqrt{n}}\right).$$

2.5 Solution improved: weighted CJIV

We also consider a modification of our CJIV estimator, which is weighted by cluster sizes :

$$\hat{\beta}_{WCJIV} = \left(\sum_{\substack{i,j \\ C_i \neq C_j}} p_{ij} X_i X_j' / \sqrt{|C_i| \cdot |C_j|}\right)^{-1} \sum_{\substack{i,j \\ C_i \neq C_j}} p_{ij} X_i y_j / \sqrt{|C_i| \cdot |C_j|}.$$

Weighted CJIV by construction normalizes terms both in the numerator and the denominator, so that we analyze self-normalized cluster-level averages. It makes our approach more robust to cluster size heterogeneity and to large clusters.

[7]Indeed,

$$var\left(A_{2n}^{CJIV}\right) = \frac{1}{n} \sum_i \sum_{\substack{j \\ C_j \neq C_i}} p_{ij} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \\ C_\ell = C_j}} p_{k\ell} \left(E[u_i u_k] E[\varepsilon_j \varepsilon_\ell] + E[u_i \varepsilon_k] E[u_\ell \varepsilon_j]\right) = O\left(\frac{sn_g}{G}\right)$$

(see S1 in Appendix S)

[8]see S2 in Appendix S

3 Asymptotic theory (preliminary)

We combine several asymptotic frameworks, including many instruments asymptotics (see Bekker, 1994), many weak instrument asymptotics (see Chao and Swanson, 2005), and many clusters asymptotics (see e.g. B. E. Hansen and Lee, 2019). We require the number of clusters to grow with the sample size and we allow for heterogeneous unbounded cluster sizes, individually weak instruments, and the number of instruments growing with sample size.

Assumption 1. $K = K_n \to \infty$, Z includes among its columns a vector of ones, $\text{rank}(Z) = K$, and for some, $C < 1$ $p_{ii} \leq C \; \forall \; i$ a.s.n. As $n \to \infty$, $\max_{g \in \mathcal{G}} \frac{n_g^3}{n} \to 0$.

Assumption 1 formally sets up the many instruments framework and many clusters framework. At the same time, the restrict the growth rate of maximum cluster size, which is common in the literature.

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

Assumption 2 is similar to the one made by Chao et al. (2012). It links the structure of the reduced form to the latent underlying instrument z_i and to the "scaling" matrix $\tilde{S}_n$. Also Assumption 2 restricts the concentration parameter, which is a measure of instrument strength, to the number of instruments and maximum cluster size. The restriction $\sqrt{K} n_{\max}/r_n \to 0$ is stronger than assumed for consistency of JIVE (Chao et al., 2012) under random sampling by a factor of $n_{\max}$. Though it is still less restrictive than $K/r_n \to 0$ required for consistency of 2SLS under random sampling (Chao and Swanson, 2005).

Assumption 3. There is a constant, C, such that conditional on $\mathcal{Z} = (\Upsilon, Z)$, the observations $(\varepsilon_{[g]}, U_{[g]})(g = 1, ..., G)$ are independent, with $\mathbb{E}[\varepsilon_i|\mathcal{Z}] = \mathbb{E}[U_i|\mathcal{Z}] = 0 \; \forall \; i$, $\sup_i \mathbb{E}[\varepsilon_i^2|\mathcal{Z}] < C$, and $\sup_i \mathbb{E}[||U_i||^2|\mathcal{Z}] \leq C$ a.s.

Theorem 1. Suppose Assumptions 1-3 and T are satisfied. Then, $r_n^{-1/2} S_n'(\hat{\beta}_{CJIV} - \beta) \xrightarrow{p} 0$, and $\hat{\beta} \xrightarrow{p} \beta$.

Conjecture. Weighted CJIV estimator is consistent under non-stronger assumptions.

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

Assumption 4 ensures that the reduced form is well approximated by a linear combination of instruments.

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

Theorem 2. Suppose that Assumptions 1-5 are satisfied and $\max_{g\in\mathcal{G}} \frac{n_{\max}^5}{n} \to 0$, $\sigma_i^2 \geq C > 0$ a.s., and $Kn_{\max}^2/r_n$ is bounded. Then V_n is nonsingular a.s.n, and

$$V_n^{-1/2} S_n'(\hat{\beta}_{CJIV} - \beta) \xrightarrow{d} \mathcal{N}(0, I_L).$$

Theorem 3. Suppose that Assumptions 1-5 are satisfied, $\sigma_i^2 \geq C > 0$ a.s., and $Kn_{\max}^2/r_n$ is bounded. Then $V_{W,n}$ is nonsingular a.s.n, and

$$V_{W,n}^{-1/2} S_n'(\hat{\beta}_{WCJIV} - \beta) \xrightarrow{d} \mathcal{N}(0, I_L).$$

4 Variance estimation

The population variance of (non-normalized) CJIV influence function is by definition

$$\mathbb{E}\left[\left(\sum_{\substack{i,j\\ C_i\neq C_j}} p_{ij}x_ie_j\right)^2 |\mathcal{Z}\right] = \mathbb{E}\left[\sum_i \sum_{\substack{j\\ C_j\neq C_i}} \sum_k \sum_{\substack{\ell\\ C_\ell\neq C_k}} x_i x_k' e_j e_\ell p_{ij} p_{k\ell} |\mathcal{Z}\right]$$

$$= \mathbb{E}\left[\sum_i \sum_{\substack{j\\ C_j\neq C_i}} \sum_{\substack{k\\ C_k=C_j}} \sum_{\substack{\ell\\ C_\ell=C_i}} x_i x_k e_j e_\ell p_{ij} p_{k\ell} + \sum_i \sum_k \sum_{\substack{j\\ C_j\notin\{C_i,C_k\}}} \sum_{\substack{\ell\\ C_\ell=C_j}} x_i x_k e_j e_\ell p_{ij} p_{k\ell}\right].$$

It consists of two parts, the first one representing association between first-stage residuals u and second-stage errors ε, while the second one – between second-stage errors ε. The population variance above can be straightforwardly estimated by its sample analog

$$\hat{\Sigma} = \sum_i \sum_{\substack{j\\ C_j\neq C_i}} \sum_{\substack{k\\ C_k=C_j}} \sum_{\substack{\ell\\ C_\ell=C_i}} x_i x_k \hat{e}_j \hat{e}_\ell p_{ij} p_{k\ell} + \sum_i \sum_k \sum_{\substack{j\\ C_j\notin\{C_i,C_k\}}} \sum_{\substack{\ell\\ C_\ell=C_j}} x_i x_k \hat{e}_j \hat{e}_\ell p_{ij} p_{k\ell},$$

where $\hat{e}_i$ is a residual $\hat{e}_i = y_i - x_i'\hat{\delta}_{CJIV}$. Thus, the variance estimator of $\hat{\delta}_{CJIV}$ is

$$\hat{V}_n = \hat{H}_{CJIV}^{-1} \hat{\Sigma} \hat{H}_{CJIV}^{-1},$$

and $\hat{H}_{CJIV}$ is as usual estimated by $\sum_{\substack{i,j\\ C_i\neq C_j}} p_{ij}x_ix_j'$.

For the efficiency of computation, we propose a vectorized variance estimator formula. We use $\sum_{\substack{i\in g\\ j\in h}} p_{ij}x_ix_j' = X_{[g]}'P_{[gh]}X_{[h]}$. Then

$$\hat{H}_{CJIV} = X'PX - \sum_g X_{[g]}'P_{[gh]}X_{[h]},$$

which is more computationally efficient. Similarly, the influence function variance can be represented by

$$\hat{\Sigma} = \sum_{\substack{g,h\\ g\neq h}} X_{[g]}'P_{[gh]}e_{[h]}e_{[g]}'P_{[gh]}X_{[h]} + \sum_{\substack{g,h,g'\\ g'\notin\{g,h\}}} X_{[g]}'P_{[gg']}e_{[g']}e_{[g']}'P_{[g'h]}X_{[h]}.$$

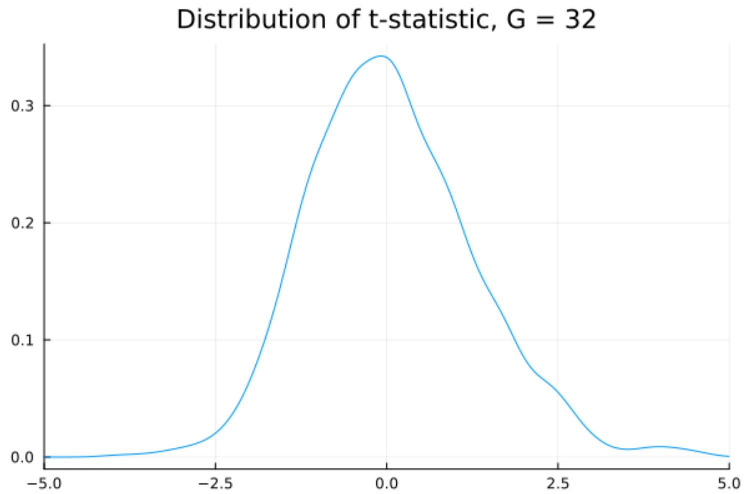

Figure 1: N = 512, G = 32, homogeneous cluster sizes

These vectorized formulae substitute multiple sums across all observations by sums across all clusters, which usually have considerably fewer summands. As a result, we can substantially reduce the computation time.

Next we show some suggestive evidence that our variance estimator is asymptotically valid, and inferences that use it are correct.

4.1 Simulations

We set up a simulation study that shows that ... The design is heteroskedastic, with cluster dependence, in the spirit of Example DI from Section 2. All variables are indexed by supergroup s, (nested in it) cluster g, and generic i. Endogenous variable is a scalar, instruments are supergroup dummies ("supergroup-specific intercept"), a continuous instrument with its powers up to 4, and interactions of this continuous instrument with supergroup dummies ("supergroup-specific slope"):

$$y_{sgi} = X_{sgi}\beta + \varepsilon_{sgi},$$

$$X_{sgi} = \alpha_s + \theta_s Z_{sgi} + \sum_{k=1}^{4} \omega_k Z_{sgi}^k + u_{sgi}.$$

For more details on DGP, see S5 in Appendix. Figure

Figure 1 shows the distribution of the t-statistic (which uses estimated variance). This distribution already has a bell-shaped curve even for 32 clusters ($n_{\max} = 16$, $N = 532$). First, Figure 1 suggests that variance estimator that we propose in Section 4 produces normally distributed t-statistics as desired. Second, the quality of approximation seems to be pretty good already with a modest number of clusters.

5 Discussion and extensions

The leave-cluster-out idea that we employ can be extended to deal with other known structures of dependence. Here we focus on clustered data with clusters being non-overlapping, thus the population variance-covariance matrix of errors is nicely represented by a block-diagonal matrix with blocks corresponding to clusters. More generally than in our setup, if the data contains multiway clustering, a leave-out procedure similar to ours can be carried out. While we essentially assign zero weight (in the estimator) to pairs of observations from the same cluster because they might be correlated, in multiway clustering setups it might be also reasonable to give zero weight to pairs of observations that are associated. Further, similar leave-out ideas can be applied to known network structures with geometrically decaying correlation. In this case, we can give zero weight to observations that are highly associated with one another.

Though we focus on many instruments setup, the leave-cluster-out procedure might be also useful in few instruments environments. Section 6.2 illustrates how CJIV works in the setup of Autor et al. (2013) with one instrument only. Heuristically, the reason for leave-cluster-out in such setups is that the fitted values from the first stage are weighted averages of endogenous xs:

$$\hat{x}_i = \sum_{j=1}^{n} p_{ij} x_j = p_{ii} x_i + \sum_{\substack{i \neq j \\ C_i = C_j}} p_{ij} x_j + \sum_{C_i \neq C_j} p_{ij} x_j,$$

and, while the first term is always endogenous (and non-negligible with many instruments), the second term might also be endogenous if there is cluster dependence and $\mathbb{E}[\varepsilon_i x_j] \neq 0$ for some (i, j) in the same cluster. Together with the instrument design, this cluster dependence of unobservables might even make fitted values endogenous and bias 2SLS estimator towards the probability limit of OLS estimator. Therefore, CJIV is more robust to these cases of strong cluster-dependence also with few instruments.

6 Applications

In this section we illustrate the performance of CJIV with two classical empirical examples and compare it to 2SLS and JIVE. The first one is a study of returns to schooling by Angrist and Krueger (1991) and employs many instruments. The second one is a study of Chinese import penetration in the US by Autor et al. (2013) and employs only one instrument.

6.1 Returns to schooling

We apply CJIV to two setups from Angrist and Krueger (1991). In one of them we use quarter of birth dummies as instruments for education (30 instruments in total). In another one, we augment the set of instruments by state of birth dummies and their interactions with quarter of birth (180 instruments in total). We treat states as clusters. Table 1 presents the results. CJIV provides a larger effect than 2SLS, the effect is quite precise, and it is more stable than JIVE, that differs dramatically in these two setups.

Instruments set	OLS	2SLS	JIVE	CJIV
30 instruments	0.0632	0.0600	0.0860	0.1311
		(0.0290)		(0.0211)
180 instruments	0.0624	0.0635	0.1329	0.1147
		(0.0122)		(0.0187)

Table 1: Angrist and Krueger (1991) re-estimation of $\hat{\beta}$; s.e. in parentheses

6.2 Effects of import penetration

Autor et al. (2013) instrument for US local exposure to Chinese imports using Chinese import exposure of other countries, so they have a single instrument in their specifications. We run the annual change in manufacturing employment per working-age population on the annual change in imports from China per working-age population and a bunch of controls (like in Table 3, column 6 in Autor et al., 2013, but unweighted). Similarly to Section 6.1, we treat states as clusters. Standard errors are clustered at the state level, like in the paper. The results are given in Table 2. Similar to Section 6.1, CJIV indicates a stronger effect than 2SLS, and quite precisely. It suggests that CJIV is capable of providing point estimates that are not worse than alternative estimators, regarding neither bias, nor precision, even in few instruments setups.

2SLS	JIVE	CJIV
-0.3028	-0.3447	-0.3586
(0.1005)		(0.0994)

Table 2: Autor et al. (2013) re-estimation of $\hat{\beta}$; s.e. in parentheses

7 Conclusion

In this paper we propose two novel leave-cluster-out estimators and derive their asymptotic properties under many weak instruments and many clusters. We find that CJIV estimator is consistent and asymptotically normal and argue that the inferences based on it are valid. We further discuss that similar leave-out ideas can be employed in other contexts with known dependence structure. In progress is also the proof of variance estimator consistency and inference validity.

Appendix

Appendix S.

S1. Indeed,

$$E\left[A_{2n}^{JIVE}\right] = \frac{1}{\sqrt{n}} \sum_{i} \sum_{\substack{j \neq i \\ C_j = C_i}} p_{ij} E\left[u_i \varepsilon_j\right] = O\left(\frac{s n_g}{\sqrt{n}}\right),$$

as

$$\sum_{i} \sum_{\substack{j \neq i \\ C_j = C_i}} p_{ij} = \frac{s}{n} G\left(\frac{n}{G} - 1\right) \frac{n}{G} = s\left(n_g - 1\right).$$

As for $var\left(A_{2n}^{JIVE}\right)$, it rests on the behaviour of the following sums (see SA):

$$\begin{aligned}
\sum_{i} \sum_{\substack{j \\ C_j \neq C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \\ C_\ell = C_j}} p_{ij} p_{k\ell} &= \left(\frac{s}{n}\right)^2 \sum_{g_1} \sum_{g_2 \neq g_1} \sum_{i \in C_{g_1}} \sum_{j \in C_{g_2}} n_g^2 = \left(\frac{s}{n}\right)^2 s \frac{G}{s}\left(\frac{G}{s} - 1\right) n_g^2 n_g^2 = s\left(1 - \frac{s}{G}\right) n_g^2, \\
\sum_{i} \sum_{\substack{j \neq i \\ C_j = C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \neq k \\ C_\ell = C_j}} p_{ij} p_{k\ell} &= \left(\frac{s}{n}\right)^2 \sum_{g} \sum_{\substack{j \neq i \\ i,j \in C_g}} n_g\left(n_g - 1\right) = \left(\frac{s}{n}\right)^2 G n_g^2 \left(n_g - 1\right)^2 < s\left(1 - \frac{s}{G}\right) n_g^2.
\end{aligned}$$

S2. Indeed, the order rests on the following two terms (see SA) both of which can be bounded from above:

$$\begin{aligned}
\left|\sum_{i} \sum_{\substack{j \\ C_j \neq C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \\ C_\ell = C_j}} p_{ij} p_{k\ell}\right| &\leq \left(\sum_{i} \sum_{\substack{j \\ C_j \neq C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \\ C_\ell = C_j}} p_{ij}^2\right)^{1/2} \left(\sum_{i} \sum_{\substack{j \\ C_j \neq C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \\ C_\ell = C_j}} p_{k\ell}^2\right)^{1/2} \\
&\leq \left(n_g^2 \sum_{i} \sum_{\substack{j \\ C_j \neq C_i}} p_{ij}^2\right)^{1/2} \left(\sum_{i} \sum_{\substack{k \\ C_k = C_i}} \sum_{j} \sum_{\substack{\ell \\ C_\ell = C_j}} p_{k\ell}^2\right)^{1/2} \\
&\leq \left(n_g^2 \sum_{i} \sum_{j} p_{ij}^2\right)^{1/2} \left(n_g \sum_{k} n_g \sum_{\ell} p_{k\ell}^2\right)^{1/2} = \left(n_g^2 \ell\right)^{1/2} \left(n_g^2 \ell\right)^{1/2} = O\left(n n_g^2\right).
\end{aligned}$$

and in a similar vein, though less tightly:

$$\begin{aligned}
\left|\sum_{i} \sum_{\substack{j \neq i \\ C_j = C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \neq k \\ C_\ell = C_j}} p_{ij} p_{k\ell}\right| &\leq \left(\sum_{i} \sum_{\substack{j \neq i \\ C_j = C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \neq k \\ C_\ell = C_j}} p_{ij}^2\right)^{1/2} \left(\sum_{i} \sum_{\substack{j \neq i \\ C_j = C_i}} \sum_{\substack{k \\ C_k = C_i}} \sum_{\substack{\ell \neq k \\ C_\ell = C_j}} p_{k\ell}^2\right)^{1/2} \\
&\leq \left(n_g^2 \sum_{i} \sum_{\substack{j \neq i \\ C_j = C_i}} p_{ij}^2\right)^{1/2} \left(\sum_{i} \sum_{\substack{k \\ C_k = C_i}} \sum_{j} \sum_{\substack{\ell \\ C_\ell = C_j}} p_{k\ell}^2\right)^{1/2} = O\left(n n_g^2\right).
\end{aligned}$$

S3. First, because of the presence of a constant among the instruments,

$$\sum_i \sum_{j \neq i} p_{ij} = \sum_i \sum_j p_{ij} - \sum_i p_{ii} = n - \ell = O(n).$$

Hence, due to symmetry across off-diagonal elements,

$$\sum_i \sum_{\substack{j \neq i \\ C_j = C_i}} p_{ij} = \frac{G\left(n_g^2 - n_g\right)}{n^2 - n} O\left(\sum_i \sum_{j \neq i} p_{ij}\right) = O(n_g).$$

where $G\left(n_g^2 - n_g\right)$ is the number of off-diagonal elements in the double sum of interest, and $n^2 - n$ is the total number of off-diagonal elements in P. Therefore,

$$E\left[A_{2n}^{JIVE}\right] = \frac{\sigma_{u\varepsilon}}{\sqrt{n}} \sum_i \sum_{\substack{j \neq i \\ C_j = C_i}} p_{ij} = O\left(\frac{n_g}{\sqrt{n}}\right).$$

S4. Presuming that $\sum_{i \neq j} x_i p_{ij} x_j = O_p(n)$, the difference $\hat{\beta} - \beta$ will be $o_p(1)$ if $\sum_{j \neq i} p_{ij} x_i \varepsilon_j = O_p(n)$. Considering the order of the bias of this quantity, we evaluate the expectation

$$E\left[\sum_{j \neq i} p_{ij} x_i \varepsilon_j\right] = \sum_{\substack{i \neq j \\ C_j = C_i}} p_{ij} \sigma_{ij}^{u\varepsilon}.$$

S5. Simulation setup details:

$$\begin{aligned}
& Z_{gi} = \zeta_g + \zeta_{gi}, && \\
& \zeta_g \overset{\text{iid}}{\sim} \mathcal{N}(0,1), && \zeta_{gi} \overset{\text{iid}}{\sim} \mathcal{N}(0,1). \\
& u_{gi} = v_{gi} + v_{gi}^u + \xi_g, && \varepsilon_{gi} = v_{gi} + v_{gi}^\varepsilon + \xi_g, \\
& \xi_g \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \zeta_g^2\right), && v_{gi} \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \zeta_{gi}^2\right), \\
& v_{gi}^u \overset{\text{iid}}{\sim} \mathcal{N}(0,1), && v_{gi}^\varepsilon \overset{\text{iid}}{\sim} \mathcal{N}(0,1).
\end{aligned}$$

Appendix A.

Proofs are in progress; so normalizations in certain places are subject to further work.

Lemma A1. If, conditional on $\mathcal{Z} = (\Upsilon, Z)$, $(W_{[g]}, Y_{[g]})(g = 1, ..., G)$ are independent a.s., W_i and Y_i are scalars, and P is a symmetric, idempotent matrix of rank K, then for $\bar{w} = \mathbb{E}[(W_1, ..., W_n)'|\mathcal{Z}]$ and $\bar{y} = \mathbb{E}[(Y_1, ..., Y_n)'|\mathcal{Z}]$, $\bar{\sigma}_{Y_n} = \max_i \sqrt{\mathrm{Var}(Y_i|\mathcal{Z})}$, $\bar{\sigma}_{W_n} = \max_i \sqrt{\mathrm{Var}(w_i|\mathcal{Z})}$, and $D_n = \bar{\sigma}^2_{W_n}\bar{\sigma}^2_{Y_n} n^2_{\max} K + \bar{y}'\bar{y} n_{\max} \bar{\sigma}^2_{W_n} + \bar{w}'\bar{w} n_{\max} \bar{\sigma}^2_{Y_n}$, there exists $C > 0$ such that

$$\mathbb{E}\left[\left(\sum_{C_i \neq C_j} p_{ij} W_i Y_j - \sum_{C_i \neq C_j} p_{ij} \bar{w}_i \bar{y}_j\right)^2\right] \leq C D_n \quad a.s.$$

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

$$\sum_{C_i \neq C_j} p_{ij} W_i Y_j - \sum_{C_i \neq C_j} p_{ij} \bar{w}_i \bar{y}_j = \sum_{C_i \neq C_j} p_{ij} \tilde{W}_i \tilde{Y}_j + \sum_{C_i \neq C_j} p_{ij} \tilde{W}_i \bar{y}_j + \sum_{C_i \neq C_j} p_{ij} \bar{w}_i \tilde{Y}_j.$$

By CS and $\sum_j p_{ij}^2 = p_{ii}$

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{C_i \neq C_j} p_{ij} \tilde{W}_i \tilde{Y}_j\right)^2 |\mathcal{Z}\right] &= \sum_{\mathcal{I}} p_{ij} p_{k\ell} \left(\mathbb{E}\left[\tilde{W}_i \tilde{W}_k |\mathcal{Z}\right] \mathbb{E}\left[\tilde{Y}_j \tilde{Y}_\ell |\mathcal{Z}\right] + \mathbb{E}\left[\tilde{W}_i \tilde{Y}_k |\mathcal{Z}\right] \mathbb{E}\left[\tilde{Y}_j \tilde{W}_\ell |\mathcal{Z}\right]\right) \\
&\leq 2\bar{\sigma}^2_{W_n} \bar{\sigma}^2_{Y_n} \sum_{\mathcal{I}} |p_{ij} p_{k\ell}| \leq 2\bar{\sigma}^2_{W_n} \bar{\sigma}^2_{Y_n} \sqrt{\sum_{C_i \neq C_j} \sum_{\mathcal{I}_{ij}} p_{ij}^2 \sum_{C_k \neq C_\ell} \sum_{\mathcal{I}_{k\ell}} p_{k\ell}^2} \\
&\leq 2\bar{\sigma}^2_{W_n} \bar{\sigma}^2_{Y_n} n^2_{\max} K,
\end{aligned}$$

where $\mathcal{I} = \{(i, j, k, \ell) | C_i = C_k \neq C_j = C_\ell\}$, and $\mathcal{I}_{ij} = \{(k, \ell) | C_i = C_k \neq C_j = C_\ell\}$.

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{C_i \neq C_j} p_{ij} \tilde{W}_i \bar{y}_j\right)^2 |\mathcal{Z}\right] &= \mathbb{E}\left[\left(\sum_{i,j} p_{ij} \tilde{W}_i \bar{y}_j\right)^2 |\mathcal{Z}\right] + \mathbb{E}\left[\left(\sum_{C_i = C_j} p_{ij} \tilde{W}_i \bar{y}_j\right)^2 |\mathcal{Z}\right] \\
&\quad - 2\mathbb{E}\left[\sum_{i,j} \sum_{C_k = C_\ell} p_{ij} \tilde{W}_i \bar{y}_j p_{k\ell} \tilde{W}_k \bar{y}_\ell |\mathcal{Z}\right].
\end{aligned}$$

Now let $\tilde{W} = (\tilde{W}_1, ..., \tilde{W}_n)'$, $\bar{x}(g) = (\bar{x}_i)_{i \in g}$ for $x \in \{w, y\}$, $\mathring{P}$ be a block-diagonal version of P with G blocks corresponding to clusters, $\Sigma_{X_n} = \mathbb{E}\left[\tilde{X}\tilde{X}'|\mathcal{Z}\right]$ for $X \in \{W, Y\}$, $\Sigma_{X_n(g)} =$

$\mathbb{E}\left[\tilde{X}(g)\tilde{X}(g)'|\mathcal{Z}\right]$ for $X \in \{W, Y\}$. Consider three terms above separately:

$$
\begin{aligned}
\mathbb{E}\left[\left(\sum_{i,j} p_{ij}\tilde{W}_i\bar{y}_j\right)^2|\mathcal{Z}\right] &= \mathbb{E}\left[\left(\bar{y}'P\tilde{W}\right)^2|\mathcal{Z}\right] = \bar{y}'P\mathbb{E}\left[\tilde{W}\tilde{W}'|\mathcal{Z}\right]P\bar{y} \\
&\leq \bar{y}'P\bar{y}\left|\left|\Sigma_{\bar{W}_n}\right|\right| = \bar{y}'\bar{y}\max_g\left|\left|\Sigma_{\bar{W}_n(g)}\right|\right| \\
&\leq \bar{y}'\bar{y}\max_g\left|\left|\Sigma_{\bar{W}_n(g)}\right|\right|_F = \bar{y}'\bar{y}\max_g\sqrt{\sum_{i,j\in g}\Sigma^2_{\bar{W}_n(g),ij}} \\
&\leq \bar{y}'\bar{y}\max_g\sqrt{n_g^2\left(\max_{i\in g}\operatorname{Var}(W_i|\mathcal{Z})\right)^2} \\
&\leq \bar{y}'\bar{y}n_{\max}\bar{\sigma}^2_{W_n}, \\
\mathbb{E}\left[\left(\sum_{C_i=C_j} p_{ij}\tilde{W}_i\bar{y}_j\right)^2|\mathcal{Z}\right] &= \mathbb{E}\left[\left(\sum_{i,j\in g} p_{ij}\tilde{W}_i\bar{y}_j\right)^2|\mathcal{Z}\right] = \mathbb{E}\left[\left(\sum_g \bar{y}(g)'P_{gg}\tilde{W}(g)\right)^2|\mathcal{Z}\right] \\
&\leq \sum_g \bar{y}(g)'P_{gg}^2\bar{y}(g)\left|\left|\mathbb{E}\left[\tilde{W}(g)\tilde{W}(g)'|\mathcal{Z}\right]\right|\right| \\
&\leq \sum_g \bar{y}(g)'\bar{y}(g)\left|\left|\Sigma_{\bar{W}_n(g)}\right|\right|\cdot\left|\left|P_{gg}\right|\right|^2 \\
&\leq n_{\max}\bar{\sigma}^2_{W_n}\sum_g \bar{y}(g)'\bar{y}(g) \leq \bar{y}'\bar{y}n_{\max}\bar{\sigma}^2_{W_n}, \\
\left|\mathbb{E}\left[\sum_{i,j}\sum_{C_k=C_\ell} p_{ij}\tilde{W}_i\bar{y}_j p_{k\ell}\tilde{W}_k\bar{y}_\ell|\mathcal{Z}\right]\right| &= \left|\mathbb{E}\left[\bar{y}'P\tilde{W}\tilde{W}'\mathring{P}\bar{y}|\mathcal{Z}\right]\right| = \left|\bar{y}'P\mathbb{E}\left[\tilde{W}\tilde{W}'|\mathcal{Z}\right]\mathring{P}\bar{y}\right| \\
&\leq \sqrt{\bar{y}'P\bar{y}}\sqrt{\bar{y}'\mathring{P}'\mathring{P}\bar{y}}\left|\left|\Sigma_{\bar{W}_n}\right|\right| \leq \bar{y}'\bar{y}\sqrt{\max_g\left|\left|P_{gg}\right|\right|}\left|\left|\Sigma_{\bar{W}_n}\right|\right| \\
&\leq \bar{y}'\bar{y}n_{\max}\bar{\sigma}^2_{W_n}.
\end{aligned}
$$

Hence,

$$
\mathbb{E}\left[\left(\sum_{C_i\neq C_j} p_{ij}\tilde{W}_i\bar{y}_j\right)^2|\mathcal{Z}\right] \leq C\bar{y}'\bar{y}n_{\max}\bar{\sigma}^2_{W_n}.
$$

Similarly, interchanging roles of Y_i and W_j,

$$
\mathbb{E}\left[\left(\sum_{C_i\neq C_j} p_{ij}\bar{w}_i\tilde{Y}_j\right)^2|\mathcal{Z}\right] \leq C\bar{w}'\bar{w}n_{\max}\bar{\sigma}^2_{Y_n}.
$$

Then by c_r-inequality

$$
\begin{aligned}
\mathbb{E}\left[\left(\sum_{C_i\neq C_j} p_{ij}W_iY_j - \sum_{C_i\neq C_j} p_{ij}\bar{w}_i\bar{y}_j\right)^2\right] &\leq C\left(\mathbb{E}\left[\left(\sum_{C_i\neq C_j} p_{ij}\tilde{W}_i\tilde{Y}_j\right)^2 |\mathcal{Z}\right]\right. \\
&\quad + \mathbb{E}\left[\left(\sum_{C_i\neq C_j} p_{ij}\tilde{W}_i\bar{y}_j\right)^2 |\mathcal{Z}\right] \\
&\quad \left. + \mathbb{E}\left[\left(\sum_{C_i\neq C_j} p_{ij}\bar{w}_i\tilde{Y}_j\right)^2 |\mathcal{Z}\right]\right) \\
&\leq C\left(\bar{\sigma}^2_{W_n}\bar{\sigma}^2_{Y_n}n^2_{\max}K + \bar{y}'\bar{y}n_{\max}\bar{\sigma}^2_{W_n} + \bar{w}'\bar{w}n_{\max}\bar{\sigma}^2_{Y_n}\right) \\
&\leq CD_n.
\end{aligned}
$$

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

(i) $P = P(\mathcal{Z})$ is a symmetric, idempotent matrix with $\text{rank}(P) = K$ and $p_{ii} \leq C < 1$;

(ii) $\{C_g\}_{g=1}^G$ is a partition of $\{1, ..., n\}$;

(iii) $(W_{gG}, U_g, \varepsilon_g)_g$ are independent, and $D_G = \sum_{g=1}^G \mathbb{E}[W_{gG}W'_{gG}|\mathcal{Z}]$ satisfies $||D_G|| \leq C$ a.s.n;

(iv) $\mathbb{E}[W_{gG}|\mathcal{Z}] = 0$, $\mathbb{E}[U_g|\mathcal{Z}] = 0$, $\mathbb{E}[\varepsilon_g|\mathcal{Z}] = 0$, and there exists a constant C such that $\mathbb{E}[||U_i||^4|\mathcal{Z}] \leq C$, $\mathbb{E}[||\varepsilon_i||^4|\mathcal{Z}] \leq C$;

(v) $\sum_{g=1}^G \mathbb{E}\left[||W_{gG}||^4|\mathcal{Z}\right] \xrightarrow{a.s.} 0$;

(vi) $K \to \infty$ as $G \to \infty$.

Then

1. for
$$\bar{\Sigma}_G \equiv \sum_{\substack{C_i\neq C_j\\ C_k=C_i\\ C_\ell=C_j}} p_{ij}p_{k\ell}\left(\mathbb{E}[U_iU'_k|\mathcal{Z}]\mathbb{E}[\varepsilon_j\varepsilon_\ell|\mathcal{Z}] + \mathbb{E}[U_i\varepsilon_k|\mathcal{Z}]\mathbb{E}[\varepsilon_jU'_\ell|\mathcal{Z}]\right)/K$$
and any sequences c_{1G} and c_{2G} depending on $\mathcal{Z}$ of conformable vectors with $||c_{1G}|| \leq C$, $||c_{2G}|| \leq C$, and $\Xi_G = c'_{1G}D_Gc_{1G} + c'_{2G}\bar{\Sigma}_Gc_{2G} > 1/C$ a.s.n, it follows that
$$Y_G = \Xi_G^{-1/2}\left(c'_{1G}\sum_g W_{gG} + c'_{2G}\sum_{C_i\neq C_j} U_ip_{ij}\varepsilon_j/\sqrt{K}\right) \xrightarrow{d} N(0,1), \quad a.s.;$$
i.e., $\Pr(Y_G \leq y|\mathcal{Z}) \xrightarrow{a.s.} \Phi(y)$ for all y.

2. for

$$\bar{\Sigma}_G \equiv \sum_{\substack{C_i \neq C_j \\ C_k = C_i \\ C_\ell = C_j}} p_{ij} p_{k\ell} \left(\mathbb{E}[U_i U_k' | \mathcal{Z}] \mathbb{E}[\varepsilon_j \varepsilon_\ell | \mathcal{Z}] + \mathbb{E}[U_i \varepsilon_k | \mathcal{Z}] \mathbb{E}[\varepsilon_j U_\ell' | \mathcal{Z}] \right) / \left(K |C_i| \cdot |C_j| \right)$$

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

$$Y_G = \Xi_G^{-1/2} \left(c_{1G}' \sum_g W_{gG} + c_{2G}' \sum_{C_i \neq C_j} U_i p_{ij} \varepsilon_j / \sqrt{K |C_i| \cdot |C_j|} \right) \xrightarrow{d} N(0,1), \quad a.s.;$$

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

Proof. Let $b_{1G} = c_{1G} \Xi_G^{-1/2}$ and $b_{2G} = c_{2G} \Xi_G^{-1/2}$ and note that these are bounded in G because Ξ_G is bounded away from 0 by hypothesis. Let $w_{gG} = b_{1G}' W_{gG}$ and $u_i = b_{2G}' U_i$. Then, $Y_G = w_{1G} + \sum_{g=2}^G y_{gG}$, $y_{gG} = w_{gG} + \bar{y}_{gG}$, $\bar{y}_{gG} = \sum_{g'<g} \sum_{i \in g} \sum_{j \in g'} (u_i p_{ij} \varepsilon_j + u_j p_{ji} \varepsilon_i)/\sqrt{K}$.

Also, $\mathbb{E}[||w_{1G}||^4 | \mathcal{Z}] \leq \sum_g \mathbb{E}[||w_{gG}||^4 | \mathcal{Z}] \leq C \mathbb{E}[||W_{gG}||^4 | \mathcal{Z}] \to 0$ a.s., so by a conditional version of M, we deduce that for any $v > 0$, $P(|w_{1G}| \geq v | \mathcal{Z}) \to 0$. Moreover, note that $\sup_n \mathbb{E}[|P(|w_{1G}| \geq v | \mathcal{Z})|^2] < \infty$. It follows that, by the Corollary to Theorem 25.12 of Billingsley (1986), $P(|w_{1G}| \geq v) = \mathbb{E}[P(|w_{1G}| \geq v | \mathcal{Z})] \to 0$; i.e. $w_{1G} \xrightarrow{p} 0$ unconditionally. Hence, $Y_G = \sum_{i=2}^G y_{gG} + o_p(1)$.

Let $\mathcal{X}_g = (\{W_{iG}', U_i', \varepsilon_i\}_{i \in g})'$ for $g = 1, ..., G$. Define the σ-fields $F_{g,G} = \sigma(\mathcal{X}_1, ..., \mathcal{X}_g)$ for $g = 1, ..., G$. By construction, $F_{g-1,G} \subseteq F_{g,G}$. Conditional on $\mathcal{Z}$, $\{y_{gG}, \mathcal{F}_{g,G}, 1 \leq g \leq G, G \geq 2\}$ is a martingale difference array.

Now we need to verify that $\{y_{gG}, \mathcal{F}_{g,G}, 1 \leq g \leq G, G \geq 2\}$ is square integrable, conditional on $\mathcal{Z}$, i.e. $\mathbb{E}[y_{gG}^2 | \mathcal{Z}] < \infty$. First note that $\mathbb{E}[w_{gG} \bar{y}_{gG} | \mathcal{Z}] = 0$ a.s. Then $\mathbb{E}[y_{gG}^2 | Z] = \mathbb{E}[w_{gG}^2 | Z] + \mathbb{E}[\bar{y}_{gG}^2 | Z]$.

$$\begin{aligned}
\mathbb{E}[\bar{y}_{gG}^2 | \mathcal{Z}] &\leq \frac{1}{K} \mathbb{E}\left[\sum_{g'<g} (u_{[g]}' P_{[g,g']} e_{[g']})^2 + 2 |u_{[g']} P_{[g,g']} e_{[g]}| \cdot |u_{[g]} P_{[g,g']} e_{[g']}| + (u_{[g']}' P_{[g,g']} e_{[g]})^2 \right] \\
&\leq \frac{C}{K} n_{\max}^2 \sum_{g'<g} \lambda_{\max}(P_{[g,g']} P_{[g,g']}') \\
&\leq \frac{C}{K} n_{\max}^2 \sum_{g'=1}^G \operatorname{tr}(P_{[g,g']} P_{[g,g']}') \\
&= \frac{C}{K} n_{\max}^2 \sum_{i \in g} p_{ii} \\
&\leq \frac{C}{K} n_{\max}^3, \\
\mathbb{E}[w_{gG}^2 | \mathcal{Z}] &\leq \sqrt{\mathbb{E}[w_{gG}^4 | \mathcal{Z}]} \to 0 \text{ a.s.}
\end{aligned}$$

Hence, $\mathbb{E}[y_{gG}^2 | Z] < \infty$.

Then

$$
\begin{aligned}
s_G^2(\mathcal{Z}) &= \mathbb{E}\left[\left(\sum_{g=2}^G y_{gG}\right)^2 |\mathcal{Z}\right] = \sum_{g=2}^G \mathbb{E}\left[\bar{w}_{gG}^2\right] + \sum_{g=2}^G \mathbb{E}\left[\bar{y}_{gG}^2|\mathcal{Z}\right] \\
&= b_{1G}' D_G b_{1G} - \mathbb{E}[w_{1G}] + b_{2G}'\bar{\Sigma}_G b_{2G} + o_{a.s.}(1) \\
&= \Xi_G^{-1/2}\left(c_{1G}' D_G c_{1G} + c_{2G}'\bar{\Sigma}_G c_{2G}\right)\Xi_G^{-1/2} + o_{a.s.}(1) \\
&= \Xi_G^{-1/2}\Xi_G\Xi_G^{-1/2} + o_{a.s.}(1) = 1 + o_{a.s.}(1) \xrightarrow{a.s.} 1,
\end{aligned}
$$

Thus, $s_G^2(\mathcal{Z})$ is is bounded and bounded away from 0 a.s.

Next, we check if Lindeberg's condition holds. $\sum_{g=2}^G \mathbb{E}[|y_{gG}|^4|\mathcal{Z}] \leq \sum_{g=2}^G \mathbb{E}\left[w_{gG}^4|\mathcal{Z}\right] + \sum_{g=2}^G \mathbb{E}\left[\bar{y}_{gG}^4|\mathcal{Z}\right]$. The first term is

$$
\sum_{g=2}^G \mathbb{E}\left[w_{gG}^4|\mathcal{Z}\right] = \sum_{g=2}^G \mathbb{E}\left[||b_{1G}' W_{gG}||^4|\mathcal{Z}\right] \leq C \sum_{g=2}^G \mathbb{E}\left[||W_{gG}||^4|\mathcal{Z}\right] \to 0.
$$

The second term is

$$K^2 \sum_{g=2}^{G} \mathbb{E}\left[\bar{y}_{gG}^4 | \mathcal{Z}\right] = \mathbb{E}\left[\sum_{g=2}^{G} \sum_{g_1<g} \left(u_g' P_{gg_1} \varepsilon_{g_1}\right)^4\right] + 3\mathbb{E}\left[\sum_{g=2}^{G} \sum_{\substack{g_1,g_2<g \\ g_1 \neq g_2}} \left(u_g' P_{gg_1} \varepsilon_{g_1}\right)^2 \left(u_g' P_{g}\right.\right.$$

$$+ \mathbb{E}\left[\sum_{g=2}^{G} \sum_{g_1<g} \left(\varepsilon_g' P_{gg_1} u_{g_1}\right)^4\right] + 3\mathbb{E}\left[\sum_{g=2}^{G} \sum_{\substack{g_1,g_2<g \\ g_1 \neq g_2}} \left(\varepsilon_g' P_{gg_1} u_{g_1}\right)^2 \left(\varepsilon_g'\right.\right.$$

$$K^2 \mathbb{E}\left[\sum_{g=2}^{G} \sum_{g_1<g} \left(u_g' P_{gg_1} \varepsilon_{g_1}\right)^4\right] \leq \mathbb{E}\left[\sum_{g=2}^{G} \sum_{g_1<g} \left(u_g' u_g\right)^2 \left(\varepsilon_{g_1}' \varepsilon_{g_1}\right)^2 \lambda_{\max}^2 \left(P_{gg_1} P_{g_1 g}\right)\right]$$

$$\leq C n_{\max}^4 \sum_{g=2}^{G} \sum_{g_1<g} \mathbb{E}\left[\left(\max_{i \in g}\{u_i\}\right)^2\right] \mathbb{E}\left[\left(\max_{j \in g_1}\{\varepsilon_j\}\right)^2\right] \lambda_{\max}\left(P_{gg}\right.$$

$$\leq C n_{\max}^4 \sum_{g=2}^{G} \operatorname{tr}\left(\sum_{g_1} P_{gg_1} P_{g_1 g}\right) = C n_{\max}^4 \sum_{g=2}^{G} \operatorname{tr}\left(P_{gg}\right) \leq C n_{\max}^4 K$$

$$K^2 \mathbb{E}\left[\sum_{g=2}^{G} \sum_{\substack{g_1,g_2<g \\ g_1 \neq g_2}} \left(u_g' P_{gg_1} \varepsilon_{g_1}\right)^2 \left(u_g' P_{gg_2} \varepsilon_{g_2}\right)^2\right] \leq \mathbb{E}\left[\sum_{g=2}^{G} \sum_{\substack{g_1,g_2<g \\ g_1 \neq g_2}} \left(u_g' u_g\right)^2 \left(\varepsilon_{g_1}' \varepsilon_{g_1}\right) \left(\varepsilon_{g_2}' \varepsilon_{g_2}\right) ||P_{gg_1}||^2 ||P_{gg_2}||^2\right]$$

$$\leq C n_{\max}^4 \sum_{g=2}^{G} \sum_{g_1<g} ||P_{gg_1}||^2 \sum_{g_2<g} ||P_{gg_2}||^2$$

$$\leq C n_{\max}^4 \sum_{g=2}^{G} \left(\sum_{g_1<g} ||P_{gg_1}||_F^2\right)^2$$

$$= C n_{\max}^4 \sum_{g=2}^{G} \left(\sum_{g_1} \sum_{\substack{i \in g \\ j \in g_1}} p_{ij}^2\right)^2$$

$$\leq C n_{\max}^4 \sum_{g=2}^{G} \left(\sum_{i \in g} p_{ii}\right)^2$$

$$\leq C n_{\max}^5 \sum_{g=2}^{G} \sum_{i \in g} p_{ii}^2 \leq C n_{\max}^5 K.$$

Then, a.s. $\sum_{g=2}^{G} \mathbb{E}\left[\bar{y}_{gG}^4 | \mathcal{Z}\right] \to 0$, and $\sum_{g=2}^{G} \mathbb{E}[|y_{gG}|^4 | \mathcal{Z}] \to 0$.

To apply the martingale central limit theorem, it suffices to show that for any $v > 0$

$$P\left(\left|\sum_{g=2}^{G}\mathbb{E}\left[y_{gG}^2|\mathcal{X}_1,...,\mathcal{X}_{g-1},\mathcal{Z}\right] - s_G^2(\mathcal{Z})\right| \geq v|\mathcal{Z}\right) \to 0.$$

As noted above, $\mathbb{E}[w_{gG}\bar{y}_{gG}|\mathcal{Z}] = 0$ a.s., thus we can write

$$\sum_{g=2}^{G}\mathbb{E}\left[y_{gG}^2|\mathcal{X}_1,...,\mathcal{X}_{g-1},\mathcal{Z}\right] - s_G^2(\mathcal{Z}) = \sum_{g=2}^{G}\left(\mathbb{E}\left[w_{gG}^2|\mathcal{X}_1,...,\mathcal{X}_{g-1},\mathcal{Z}\right] - \mathbb{E}\left[w_{gG}^2|\mathcal{Z}\right]\right)$$
$$+\sum_{g=2}^{G}\mathbb{E}\left[w_{gG}\bar{y}_{gG}|\mathcal{X}_1,...,\mathcal{X}_{g-1},\mathcal{Z}\right] + \sum_{g=2}^{G}\left(\mathbb{E}\left[\bar{y}_{gG}^2|\mathcal{X}_1,...,\mathcal{X}_{g-1},\mathcal{Z}\right] - \mathbb{E}\left[\bar{y}_{gG}^2|\mathcal{Z}\right]\right).$$

In progress an argument why first two terms are 0 asymptotically.

It remains only to show that, for any $v > 0$,

$$P\left(\left|\sum_{g=2}^{G}\left(\mathbb{E}\left[\bar{y}_{gG}^2|\mathcal{X}_1,...,\mathcal{X}_{g-1},\mathcal{Z}\right] - \mathbb{E}\left[\bar{y}_{gG}^2|\mathcal{Z}\right]\right)\right| \geq v|\mathcal{Z}\right) \to 0 \quad \text{a.s.}$$

Now write

$$
\begin{aligned}
&\sum_{g=2}^{G}\left(\mathbb{E}\left[\bar{y}_{gG}^2|\mathcal{X}_1,...,\mathcal{X}_{g-1},\mathcal{Z}\right]-\mathbb{E}\left[\bar{y}_{gG}^2|\mathcal{Z}\right]\right)\\
&=\frac{1}{K}\left(\sum_{g=2}^{G}\sum_{\substack{g'<g\\h<g}}\sum_{\substack{i,k\in g\\j\in g'\\\ell\in h}}p_{ij}p_{k\ell}\left(E[u_iu_k|\mathcal{Z}]e_je_\ell+E[e_iu_k|\mathcal{Z}]u_je_\ell+E[u_ie_k|\mathcal{Z}]e_ju_\ell+E[e_ie_k|\mathcal{Z}]u_ju_\ell\right)\right.\\
&\quad\left.-\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\j,\ell\in g'}}p_{ij}p_{k\ell}\left(E[u_iu_k|\mathcal{Z}]E[e_je_\ell|\mathcal{Z}]+2E[e_iu_k|\mathcal{Z}]E[u_je_\ell|\mathcal{Z}]+E[e_ie_k|\mathcal{Z}]E[u_ju_\ell|\mathcal{Z}]\right)\right)\\
&=\frac{1}{K}\left(\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\j,\ell\in g'}}p_{ij}p_{k\ell}\left(E[u_iu_k|\mathcal{Z}]e_je_\ell+E[e_iu_k|\mathcal{Z}]u_je_\ell+E[u_ie_k|\mathcal{Z}]e_ju_\ell+E[e_ie_k|\mathcal{Z}]u_ju_\ell\right)\right.\\
&\quad+\ 2\sum_{g=3}^{G}\sum_{h<g'<g}\sum_{\substack{i,k\in g\\j\in g'\\\ell\in h}}p_{ij}p_{k\ell}\left(E[u_iu_k|\mathcal{Z}]e_je_\ell+E[e_iu_k|\mathcal{Z}]u_je_\ell+E[u_ie_k|\mathcal{Z}]e_ju_\ell+E[e_ie_k|\mathcal{Z}]u_ju_\ell\right)\\
&\quad\left.-\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\j,\ell\in g'}}p_{ij}p_{k\ell}\left(E[u_iu_k|\mathcal{Z}]E[e_je_\ell|\mathcal{Z}]+2E[e_iu_k|\mathcal{Z}]E[u_je_\ell|\mathcal{Z}]+E[e_ie_k|\mathcal{Z}]E[u_ju_\ell|\mathcal{Z}]\right)\right)\\
&=\frac{1}{K}\left(2\sum_{g=3}^{G}\sum_{h<g'<g}\sum_{\substack{i,k\in g\\j\in g'\\\ell\in h}}p_{ij}p_{k\ell}\left(E[u_iu_k|\mathcal{Z}]e_je_\ell+E[e_iu_k|\mathcal{Z}]u_je_\ell+E[u_ie_k|\mathcal{Z}]e_ju_\ell+E[e_ie_k|\mathcal{Z}]u_ju_\ell\right)\right.\\
&\quad+\ \sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\j,\ell\in g'}}p_{ij}p_{k\ell}E[u_iu_k|\mathcal{Z}]\left(e_je_\ell-E[e_je_\ell|\mathcal{Z}]\right)\\
&\quad+\ \sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\j,\ell\in g'}}p_{ij}p_{k\ell}E[e_iu_k|\mathcal{Z}]\left(u_je_\ell-E[u_je_\ell|\mathcal{Z}]\right)\\
&\quad+\ \sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\j,\ell\in g'}}p_{ij}p_{k\ell}E[u_ie_k|\mathcal{Z}]\left(e_ju_\ell-E[e_ju_\ell|\mathcal{Z}]\right)\\
&\quad\left.+\ \sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\j,\ell\in g'}}p_{ij}p_{k\ell}E[e_ie_k|\mathcal{Z}]\left(u_ju_\ell-E[u_ju_\ell|\mathcal{Z}]\right)\right).
\end{aligned}
$$

By applying Lemma A7, we show that each term converges to 0 in probability.

The preceding argument shows that as $G \to \infty$, $P\left(Y_G \leq y|\mathcal{Z}\right) \to \Phi(y)$ a.s. $\mathbb{P}_{\mathcal{Z}}$, for every real number y, where $\Phi(y)$ denotes denotes the c.d.f. of a standard normal distribution. Moreover, it is clear that, for some $v > 0$, $\sup_G \mathbb{E}[|P(Y_G \leq y|\mathcal{Z})|^{1+v}] < \infty$. Hence, by a version of the dominated convergence theorem, as given by Theorem 25.12 of Billingsley (1986), we deduce that $P(Y_G \leq y) = \mathbb{E}[P(Y_G \leq y|\mathcal{Z})] \to \mathbb{E}[\Phi(y)] = \Phi(y)$, which gives the first conclusion.

The second conclusion follows similarly; instead of Lemma A7 we use Lemma A8.

Lemma A5. If Assumptions 1-3 are satisfied, then

$$(i)\ S_n^{-1}\hat{H}S_n^{-1} = \sum_{C_i \neq C_j} p_{ij} z_i z_j'/n + o_p(1),$$

$$(ii)\ S_n^{-1}\sum_{C_i \neq C_j} p_{ij} X_i \varepsilon_j = O_p\left(\sqrt{n_{\max}} + \sqrt{K n_{\max}^2/r_n}\right).$$

Proof. Let e_k be kth unit vector and let $Y_i = e_k' S_n^{-1} X_i = e_k' z_i/\sqrt{n} + e_k' S_n^{-1} U_i$ and $W_i = e_\ell' S_n^{-1} X_i$ for some (k, ℓ). By Assumption 2, $\lambda_{\min}(S_n) \geq C\sqrt{r_n}$, then $||S_n^{-1}|| \leq C/\sqrt{r_n}$. It follows that a.s.

$$\begin{aligned}
\mathbb{E}[Y_i|\mathcal{Z}] &= e_k' z_i/\sqrt{n}, \quad \mathrm{Var}(Y_i|\mathcal{Z}) = e_k' S_n^{-1}\mathbb{E}[U_i U_i'] S_n^{-1\prime} e_k \leq C||S_n^{-1}||^2 \leq C/r_n, \\
\mathbb{E}[W_i|\mathcal{Z}] &= e_\ell' z_i/\sqrt{n}, \quad \mathrm{Var}(W_i|\mathcal{Z}) = e_\ell' S_n^{-1}\mathbb{E}[U_i U_i'] S_n^{-1\prime} e_\ell \leq C||S_n^{-1}||^2 \leq C/r_n.
\end{aligned}$$

Note that a.s.

$$\begin{aligned}
&\bar{\sigma}_{W_n}\bar{\sigma}_{Y_n} n_{\max}\sqrt{K} \leq C n_{\max}\sqrt{K}/r_n \to 0, \\
&\sqrt{\bar{y}'\bar{y} n_{\max}\bar{\sigma}_{W_n}^2} \leq C\sqrt{\left|\left|\sum_i z_i z_i'/n\right|\right|}\sqrt{n_{\max}/r_n} \to 0, \\
&\sqrt{\bar{w}'\bar{w} n_{\max}\bar{\sigma}_{Y_n}^2} \leq C\sqrt{\left|\left|\sum_i z_i z_i'/n\right|\right|}\sqrt{n_{\max}/r_n} \to 0.
\end{aligned}$$

Because $e_k' S_n^{-1}\hat{H}S_n^{-1} e_\ell = e_k' S_n^{-1}\sum_{C_i \neq C_j} p_{ij} X_i X_j' S_n^{-1} e_\ell = \sum_{C_i \neq C_j} p_{ij} Y_i W_j$ and $p_{ij}\bar{y}_i\bar{w}_j = p_{ij} e_k' z_i z_j' e_\ell/n$, applying Lemma A1 and the conditional version of M, we deduce that for any $v > 0$ and $A_n = \{|e_k' S_n^{-1}\hat{H}S_n^{-1\prime} e_\ell - \sum_{C_i \neq C_j} p_{ij} e_k' z_i z_j' e_\ell/n| \geq v\}$, $P(A_n|\mathcal{Z}) \xrightarrow{a.s.} 0$. By the dominated convergence theorem, $P(A_n) = \mathbb{E}[P(A_n|\mathcal{Z})] \to 0$.

The preceding argument establishes the first conclusion for the (k, ℓ)th element. Doing so for every element of $S_n^{-1}\hat{H}S_n^{-1}$ completes the proof of the first conclusion.

For the second conclusion, let $Y_i = e_k' S_n^{-1} X_i = e_k' z_i/\sqrt{n} + e_k' S_n^{-1} U_i$ for some k as before and $W_i = \varepsilon_i$. Note that $\mathbb{E}[W_i|\mathcal{Z}] = 0$ and $\mathrm{Var}(W_i|\mathcal{Z}) \leq C$. Then by Lemma A1,

$$\mathbb{E}\left[\left(e_k' S_n^{-1}\sum_{C_i \neq C_j} p_{ij} X_i \varepsilon_j\right)^2 |\mathcal{Z}\right] \leq C K n_{\max}^2/r_n + C n_{\max}.$$

The conclusion then follows from the fact that $\mathbb{E}[A_n^2|\mathcal{Z}] \leq C c_n$ implies $A_n = O_p(\sqrt{c_n})$.

Lemma A6. If Assumptions 1–4 are satisfied, then

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

Proof. Denote $I_{[g]}$ to be $n_g \times n_g$ identity matrix. Using Lemma A5,

$$\begin{aligned}
S_n^{-1}\hat{H}S_n^{-1\prime} &= \sum_{C_i \neq C_j} p_{ij} z_i z_j'/n + o_p(1) \\
&= \sum_{g,h} z_{[g]}' P_{[g,h]} z_{[h]}/n - \sum_{g} z_{[g]}' P_{[g,g]} z_{[g]}/n + o_p(1) \\
&= \sum_{g} z_{[g]}' \left(\sum_{h} P_{[g,h]} z_{[h]} - z_{[g]} \right)/n + \sum_{g} z_{[g]}' (I_{[g]} - P_{[g,g]}) z_{[g]}/n + o_p(1).
\end{aligned}$$

It suffices to show that $\sum_g z_{[g]}' \left(\sum_h P_{[g,h]} z_{[h]} - z_{[g]}\right)/n = o_{\text{a.s.}}(1)$. Let $\bar{z}_{[g]} = \sum_h P_{[g,h]} z_{[h]}$ and $\bar{z}_i = \sum_j p_{ij} z_j$. Note that

$$||z_{[g]} - \bar{z}_{[g]}||^2/n = \sum_{i \in g} ||z_i - \bar{z}_i||^2$$

and

$$\sum_g ||z_{[g]} - \bar{z}_{[g]}||^2/n = \sum_i ||z_i - \bar{z}_i||^2 \to 0 \quad \text{a.s.}$$

by Lemma A6 in Chao et al. (2012). It follows that

$$\begin{aligned}
\left|\left| \sum_g z_{[g]}' \left(\bar{z}_{[g]} - z_{[g]}\right)/n \right|\right| &\leq \sum_g ||z_{[g]}|| \times || \left(\bar{z}_{[g]} - z_{[g]}\right) ||/n \\
&\leq \sqrt{\sum_g ||z_{[g]}||^2/n} \sqrt{\sum_g || \left(\bar{z}_{[g]} - z_{[g]}\right) ||^2/n} \to 0 \quad \text{a.s.}
\end{aligned}$$

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

$$\left|\left|S_n'(\hat{\delta}-\delta)/\sqrt{r_n}\right|\right| \geq \sqrt{\lambda_{\min}(S_n S_n'/r_n)}\left|\left|\hat{\delta}-\delta\right|\right| \geq C\left|\left|\hat{\delta}-\delta\right|\right|.$$

Therefore, $S_n'(\hat{\delta}-\delta)/\sqrt{r_n} \xrightarrow{p} 0$ implies $\hat{\delta} \xrightarrow{p} \delta$. Lemma A6 implies that $\left(S_n^{-1}\hat{H}S_n^{-1\prime}\right)^{-1} = O_p(1)$. By Lemma A5,

$$r_n^{-1/2}S_n'(\hat{\delta}-\delta) = \left(S_n^{-1}\hat{H}S_n^{-1\prime}\right)^{-1} S_n^{-1}\sum_{C_i \neq C_j} X_i p_{ij}\xi_j/\sqrt{r_n} = O_p(1)o_p(1) \xrightarrow{p} 0.$$

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

Assumption A1. $||M_{gg}|| \leq 1$.

Assumption A2 $\lambda_{\min}\left(\sum_g z'_{[g]}(I_{[g]} - P_{[g,g]})\mathbb{E}\left[\varepsilon_{[g]}\varepsilon'_{[g]}|\mathcal{Z}\right](I_{[g]} - P_{[g,g]})z_{[g]}\right)/n \geq C > 0$.

Proof of Theorem 2. Define

$$Y_G = \sum_{g=1}^{G} z'_{[g]}(I_{[g]} - P_{[g,g]})\varepsilon_{[g]}/\sqrt{n} + S_n^{-1}\sum_{C_i \neq C_j} U_i\varepsilon_j p_{ij}.$$

By Assumptions 2-4,

$$\begin{aligned}\mathbb{E}\left[\left|\left|\sum_{g=1}^{G}(z_{[g]} - \bar{z}_{[g]})\varepsilon_{[g]}/\sqrt{n}\right|\right|^2|\mathcal{Z}\right] &= \sum_{g=1}^{G}\left|\left|z_{[g]} - \bar{z}_{[g]}\right|\right|^2\mathbb{E}\left[\left|\left|\varepsilon_{[g]}\right|\right|^2|\mathcal{Z}\right]/n \\ &\leq C\sum_{g=1}^{G}\left|\left|z_{[g]} - \bar{z}_{[g]}\right|\right|^2 n_g/n \to 0 \text{ a.s.}\end{aligned}$$

Then by M,

$$S_n^{-1}\sum_{C_i \neq C_j} X_i\varepsilon_j p_{ij} - Y_G = \sum_{g=1}^{G}(z_{[g]} - \bar{z}_{[g]})\varepsilon_{[g]}/\sqrt{n} \xrightarrow{p} 0.$$

Let $\Gamma_G = \mathrm{Var}(Y_G|\mathcal{Z})$, so

$$\begin{aligned}\Gamma_G = &\sum_{g=1}^{G} z'_{[g]}(I_{[g]} - P_{[g,g]})\mathbb{E}\left[\varepsilon_{[g]}\varepsilon'_{[g]}\right](I_{[g]} - P_{[g,g]})z_{[g]}/n \\ &+ S_n^{-1}\sum_{g,h}\sum_{\substack{i,k\in g\\ j,\ell\in h}}(\mathbb{E}[U_iU'_k|\mathcal{Z}]\mathbb{E}[\varepsilon_j\varepsilon_\ell|\mathcal{Z}] + \mathbb{E}[U_i\varepsilon_k|\mathcal{Z}]\mathbb{E}[\varepsilon_jU'_\ell|\mathcal{Z}])p_{ij}p_{k\ell}S_n^{-1\prime}.\end{aligned}$$

$S_G^{-1} \leq C/\sqrt{r_n}$ by Assumption 2 , and $\sum_{g,h}\sum_{\substack{i,k\in g\\ j,\ell\in h}}|p_{ij}p_{k\ell}| \leq n_{\max}^2 K$. By Assumption A1 and Assumption 3, $\mathbb{E}\left[\sum_{g=1}^{G} z'_{[g]}(I_{[g]} - P_{[g,g]})\varepsilon_{[g]}\varepsilon'_{[g]}(I_{[g]} - P_{[g,g]})z_{[g]}/n\right] \leq C\sum_{g=1}^{G}||\sum_{i\in g} z_iz'_i/n|| \leq C$. Then by boundedness of $Kn_{\max}^2/r_n$, $||\Gamma_G|| \leq C$ a.s.n.

By Assumption A2, $\lambda_{\min}(\Gamma_G) \geq C > 0$ a.s.n, and $||\Gamma_G^{-1}|| \leq C$ a.s.n.

Let α be a $L \times 1$ nonzero vector. Let $W_{gG} = \sum_{i,j\in g} z_i\varepsilon_j p_{ij}/\sqrt{n}$, $c_{1G} = \Gamma_G^{-1/2}\alpha$, and $c_{2G} = \sqrt{K}S_n^{-1}\Gamma_G^{-1/2}\alpha$. Condition (i) of Lemma A2 is satisfied.

Condition (ii) of Lemma A2 is satisfied by Assumption A2.

Condition (iii) of Lemma A2 is satisfied by Assumptions 3 and 5.

Condition (iv) of Lemma A2 is satisfied by Assumptions 3 and 5:

$$\begin{aligned}\sum_g \mathbb{E}[||W_{gG}||^4|\mathcal{Z}] &\leq \sum_g ||z_{[g]}||^4||P_{[g,g]}||^4\left(\sum_{i\in g}\mathbb{E}[\varepsilon_i^2|\mathcal{Z}]\right)^2/n^2 \\ &\leq Cn_{\max}^2\sum_g ||z_{[g]}||^4/n^2 \to 0 \text{ a.s.}\end{aligned}$$

Condition (v) of Lemma A2 is satisfied by Assumption 1.

Further note that $||c_{1G}|| \leq C$ and $||c_{2G}|| \leq C$ a.s.n. Also by construction

$$\Xi_G = \text{Var}\left(c'_{1G}\sum_g W_{gG} + c'_{2G}\sum_{C_i \neq C_j} U_i \varepsilon_j p_{ij}/\sqrt{K}|\mathcal{Z}\right) = \text{Var}\left(\alpha'\Gamma_G^{-1/2}{}'Y_G|\mathcal{Z}\right) = \alpha'\alpha.$$

By Lemma A2, it follows that

$$(\alpha'\alpha)^{-1/2}\alpha'\Gamma_G^{-1/2}Y_G \xrightarrow{d} \mathcal{N}(0,1) \quad \text{a.s..}$$

By the Cramér-Wold device, $\Gamma_G^{-1/2}Y_G \xrightarrow{d} \mathcal{N}(0, I_L)$ a.s.

It follows that $\Gamma_G^{-1/2}Y_G = O_p(1)$. $||\Gamma_G^{1/2}|| \leq C$ a.s.n.; $V_n = H_n^{-1}\Gamma_G H_n^{-1}$, $\lambda_{\min}(H_n^{-1}\Gamma_G H_n^{-1}) \geq C > 0$ a.s.n., then $||V_n^{-1/2}|| \leq C$ a.s.n. Then $\Gamma_G^{1/2} = O_p(1)$ and $V_n^{-1/2} = O_p(1)$. Let $B_G = \bar{V}_n^{-1/2}H_n^{-1}\Gamma_G^{1/2}$. It follows that $B_G\Gamma_G^{-1/2} = V_n^{-1/2}H_n^{-1} = O_p(1)$ and $V_n^{-1/2}Y_G = V_n^{-1/2}\Gamma_G^{1/2}\Gamma_G^{-1/2}Y_G = O_p(1)$. Combining it with Lemma A6 and the definition of $\hat{\delta}$,

$$\begin{aligned}
\bar{V}_n^{-1/2}S'_n(\hat{\delta}-\delta) &= \bar{V}_n^{-1/2}(S_n^{-1}\hat{H}S_n^{-1}{}')^{-1}S_n^{-1}\sum_{C_i \neq C_j} p_{ij}X_i\varepsilon_j \\
&= \bar{V}_n^{-1/2}(H_n^{-1} + o_p(1))\Gamma_G^{1/2}\Gamma_G^{-1/2}(Y_G + o_p(1)) \\
&= B_G\Gamma_G^{-1/2}Y_G + o_p(1).
\end{aligned}$$

Note that B_G is orthogonal and is a function of $\mathcal{Z}$ only. Then using Slutsky theorem,

$$\bar{V}_n^{-1/2}S'_n(\hat{\delta}-\delta) = B_G\Gamma_G^{-1/2}Y_G + o_p(1) \xrightarrow{d} \mathcal{N}(0, I_L).$$

Lemma A7. Suppose that

(a) $P = P(\mathcal{Z})$ is a symmetric, idempotent matrix with $\mathrm{rank}(P) = K$ and $p_{ii} \leq C < 1$;

(b) $(\varepsilon_g, u_g)_{g=1}^G$ are independent conditional on $\mathcal{Z}$; $n_{\max}^5/n \to 0$;

(c) there exists a constant C such that a.s., $\sup_i \mathbb{E}[u_i^4|\mathcal{Z}] \leq C$, $\sup_i \mathbb{E}[\varepsilon_i^4|\mathcal{Z}] \leq C$, and $\sup_i |\phi_i(\mathcal{Z})| \leq C$. Then a.s.

Then, a.s.,

(i) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=2}^G \sum_{g'<g} \sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(\varepsilon_j\varepsilon_\ell - \mathbb{E}[\varepsilon_j\varepsilon_\ell|\mathcal{Z}]\right)\right)^2 |\mathcal{Z}\right] \to 0,$

(ii) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=2}^G \sum_{g'<g} \sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(u_j\varepsilon_\ell - \mathbb{E}[u_j\varepsilon_\ell|Z]\right)\right)^2 |\mathcal{Z}\right] \to 0,$

(iii) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=2}^G \sum_{g'<g} \sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(\varepsilon_j u_\ell - \mathbb{E}[\varepsilon_j u_\ell|Z]\right)\right)^2 |\mathcal{Z}\right] \to 0,$

(iv) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=2}^G \sum_{g'<g} \sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(u_j u_\ell - \mathbb{E}[u_j u_\ell|Z]\right)\right)^2 |\mathcal{Z}\right] \to 0,$

(v) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=3}^G \sum_{h<g'<g} \sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}\varepsilon_j\varepsilon_\ell\right)^2 |\mathcal{Z}\right] \to 0,$

(vi) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=3}^G \sum_{h<g'<g} \sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}u_j\varepsilon_\ell\right)^2 |\mathcal{Z}\right] \to 0,$

(vii) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=3}^G \sum_{h<g'<g} \sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}\varepsilon_j u_\ell\right)^2 |\mathcal{Z}\right] \to 0,$

(viii) $\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=3}^G \sum_{h<g'<g} \sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}u_j u_\ell\right)^2 |\mathcal{Z}\right] \to 0.$

Proof. We show part (i) first:

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(\varepsilon_j\varepsilon_\ell-\mathbb{E}[\varepsilon_j\varepsilon_\ell|\mathcal{Z}]\right)\right)^2|\mathcal{Z}\right]\\
&=\frac{1}{K^2}\left|\sum_{\substack{g,h\\ g'<\min\{g,h\}}}\sum_{\substack{i_1,k_1\in g\\ i_2,k_2\in h\\ j_1,\ell_1,j_2,\ell_2\in g'}} p_{i_1j_1}p_{k_1\ell_1}p_{i_2j_2}p_{k_2\ell_2}\phi_{i_1k_1}\phi_{i_2k_2}\left(\mathbb{E}[\varepsilon_{j_1}\varepsilon_{\ell_1}\varepsilon_{j_2}\varepsilon_{\ell_2}|\mathcal{Z}]-\mathbb{E}[\varepsilon_{j_1}\varepsilon_{\ell_1}|\mathcal{Z}]\mathbb{E}[\varepsilon_{j_2}\varepsilon_{\ell_2}|\mathcal{Z}]\right)\right|\\
&\leq\frac{C}{K^2}\sum_{g'}\sum_{g>g'}\sum_{\substack{i_1,k_1\in g\\ j_1,\ell_1\in g'}}|p_{i_1j_1}p_{k_1\ell_1}|\sum_{h>g'}\sum_{\substack{i_2,k_2\in h\\ j_2,\ell_2\in g'}}|p_{i_2j_2}p_{k_2\ell_2}|=\frac{C}{K^2}\sum_{g'}\left(\sum_{g>g'}\sum_{\substack{i_1\in g\\ j_1\in g'}}|p_{i_1j_1}|\sum_{\substack{k_1\in g\\ \ell_1\in g'}}|p_{k_1\ell_1}|\right)^2\\
&=\frac{C}{K^2}\sum_{g'}\left(\sum_{g>g'}\left(\sum_{\substack{i_1\in g\\ j_1\in g'}}|p_{i_1j_1}|\right)^2\right)^2\leq\frac{C}{K^2}n_{\max}^2\sum_{g'}\left(\sum_{g}\sum_{\substack{i_1\in g\\ j_1\in g'}}p_{i_1j_1}^2\right)^2\\
&=\frac{C}{K^2}n_{\max}^2\sum_{g}\left(\sum_{j\in g}p_{jj}\right)^2\leq\frac{C}{K^2}n_{\max}^3\sum_{g}\sum_{j\in g}p_{jj}^2\leq\frac{C}{K}n_{\max}^3\to 0.
\end{aligned}
$$

Parts (ii), (iii), and (iv) follow similarly by interchanging the roles of ε and u.

Now we show part (v):

$$\mathbb{E}\left[\left(\frac{1}{K}\sum_{g=3}^{G}\sum_{h<g'<g}\sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}\varepsilon_j\varepsilon_\ell\right)^2 |\mathcal{Z}\right]$$

$$= \frac{1}{K^2}\left|\sum_{\substack{g,h\\ g_2<g_1<g,h}}\sum_{\substack{i_1,k_1\in g\\ i_2,k_2\in h\\ j_1,j_2\in g_1\\ \ell_1,\ell_2\in g_2}} p_{i_1j_1}p_{k_1\ell_1}p_{i_2j_2}p_{k_2\ell_2}\phi_{i_1k_1}\phi_{i_2k_2}\mathbb{E}\left[\varepsilon_{j_1}\varepsilon_{\ell_1}\varepsilon_{j_2}\varepsilon_{\ell_2}|\mathcal{Z}\right]\right|$$

$$\leq \frac{C}{K^2}\sum_{\substack{g_2\\ g_1>g_2}}\left(\sum_{g>g_1}\sum_{\substack{i_1,k_1\in g\\ j_1\in g_1\\ \ell_1\in g_2}}|p_{i_1j_1}p_{k_1\ell_1}|\right)\left(\sum_{h>g_1}\sum_{\substack{i_2,k_2\in h\\ j_2\in g_1\\ \ell_2\in g_2}}|p_{i_2j_2}p_{k_2\ell_2}|\right)$$

$$\leq \frac{C}{K^2}n_{\max}^4\sum_{\substack{g_2\\ g_1>g_2}}\sum_{g>g_1}\sum_{\substack{i_1\in g\\ j_1\in g_1}} p_{i_1j_1}^2\sum_{g>g_1}\sum_{\substack{k_1\in g\\ \ell_1\in g_2}} p_{k_1\ell_1}^2$$

$$\leq \frac{C}{K^2}n_{\max}^4\sum_{g}\left(\sum_{j\in g}p_{jj}\right)^2 \leq C\frac{n_{\max}^5}{K}\to 0.$$

Parts (vi), (vii), and (viii) follow similarly by interchanging the roles of ε and u.

Lemma A8. Suppose that

(a) $P = P(\mathcal{Z})$ is a symmetric, idempotent matrix with $\text{rank}(P) = K$ and $p_{ii} \leq C < 1$;

(b) $(\varepsilon_g, u_g)_{g=1}^{G}$ are independent conditional on $\mathcal{Z}$; $n_{\max}^3/n \to 0$;

(c) there exists a constant C such that a.s., $\sup_i \mathbb{E}[u_i^4|\mathcal{Z}] \leq C$, $\sup_i \mathbb{E}[\varepsilon_i^4|\mathcal{Z}] \leq C$, and $\sup_i |\phi_i(\mathcal{Z})| \leq C$. Then a.s.

Then, a.s.,

(i) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(\varepsilon_j\varepsilon_\ell - \mathbb{E}[\varepsilon_j\varepsilon_\ell|\mathcal{Z}]\right)/(|g|\cdot|g'|)\right)^2|\mathcal{Z}\right] \to 0,$$

(ii) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(u_j\varepsilon_\ell - \mathbb{E}[u_j\varepsilon_\ell|Z]\right)/(|g|\cdot|g'|)\right)^2|\mathcal{Z}\right] \to 0,$$

(iii) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(\varepsilon_j u_\ell - \mathbb{E}[\varepsilon_j u_\ell|Z]\right)/(|g|\cdot|g'|)\right)^2|\mathcal{Z}\right] \to 0,$$

(iv) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=2}^{G}\sum_{g'<g}\sum_{\substack{i,k\in g\\ j,\ell\in g'}} p_{ij}p_{k\ell}\phi_{ik}\left(u_j u_\ell - \mathbb{E}[u_j u_\ell|Z]\right)/(|g|\cdot|g'|)\right)^2|\mathcal{Z}\right] \to 0,$$

(v) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=3}^{G}\sum_{h<g'<g}\sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}\varepsilon_j\varepsilon_\ell/(|g|\sqrt{|g'|\cdot|h|})\right)^2|\mathcal{Z}\right] \to 0,$$

(vi) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=3}^{G}\sum_{h<g'<g}\sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}u_j\varepsilon_\ell/(|g|\sqrt{|g'|\cdot|h|})\right)^2|\mathcal{Z}\right] \to 0,$$

(vii) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=3}^{G}\sum_{h<g'<g}\sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}\varepsilon_j u_\ell/(|g|\sqrt{|g'|\cdot|h|})\right)^2|\mathcal{Z}\right] \to 0,$$

(viii) $$\mathbb{E}\left[\left(\tfrac{1}{K}\sum_{g=3}^{G}\sum_{h<g'<g}\sum_{\substack{i,k\in g\\ j\in g'\\ \ell\in h}} p_{ij}p_{k\ell}\phi_{ik}u_j u_\ell/(|g|\sqrt{|g'|\cdot|h|})\right)^2|\mathcal{Z}\right] \to 0.$$

Proof. Almost identical to proof of Lemma A7.

References

Angrist, J. D., Imbens, G. W., & Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, *14*(1), 57–67.

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.

Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The geography of trade and technology shocks in the united states. *American Economic Review*, *103*(3), 220–225.

Bekker, P. A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica: Journal of the Econometric Society*, 657–681.

Billingsley, P. (1986). *Probability and measure*. John Wiley & Sons.

Blomquist, S., & Dahlberg, M. (1999). Small sample properties of liml and jackknife iv estimators: Experiments with weak instruments. *Journal of Applied Econometrics*, *14*(1), 69–88.

Chao, J. C., & Swanson, N. R. (2005). Consistent estimation with a large number of weak instruments. *Econometrica*, *73*(5), 1673–1692.

Chao, J. C., Swanson, N. R., Hausman, J. A., Newey, W. K., & Woutersen, T. (2012). Asymptotic distribution of jive in a heteroskedastic iv regression with many instruments. *Econometric Theory*, *28*(1), 42–86. https://doi.org/10.1017/S0266466611000120

Chao, J. C., Swanson, N. R., & Woutersen, T. (2023). Jackknife estimation of a cluster-sample iv regression model with many weak instruments. *Journal of Econometrics*, *235*(2), 1747–1769. https://doi.org/https://doi.org/10.1016/j.jeconom.2022.12.011

Donald, S. G., & Newey, W. K. (2001). Choosing the number of instruments. *Econometrica*, *69*(5), 1161–1191.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica: Journal of the Econometric Society*, 939–953.

Hansen, B. E., & Lee, S. (2019). Asymptotic theory for clustered samples. *Journal of econometrics*, *210*(2), 268–290.

Hansen, C., Hausman, J., & Newey, W. (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, *26*(4), 398–422.

Hausman, J., Lewis, R., Menzel, K., & Newey, W. (2011). Properties of the cue estimator and a modification with moments. *Journal of Econometrics*, *165*(1), 45–57.

Ligtenberg, J. W. (2023). Inference in iv models with clustered dependence, many instruments and weak identification.

van Hasselt, M. (2010). Many instruments asymptotic approximations under nonnormal error distributions. *Econometric Theory*, *26*(2), 633–645.