Cluster Jackknife Instrumental Variables Estimation

Brigham Frandsen Emily Leslie Samuel McIntyre*

August 7, 2023

Abstract

Applied researchers commonly use jackknife-based instrumental variables estimators to eliminate the bias of two-stage least squares due to many instruments. In settings where inference must be clustered, however, the jackknife fails to eliminate the many-instruments bias. We propose a cluster-jackknife approach in which first-stage predicted values for each observation are constructed from a regression that leaves out the observation's entire cluster, not just the observation itself. The cluster-jackknife instrumental variables estimator (CJIVE) eliminates many-instruments bias, and consistently estimates causal effects in the traditional linear model and local average treatment effects in the heterogeneous treatment effects framework. In addition to establishing CJIVE's theoretical properties, we illustrate the bias of existing approaches and the consistency of CJIVE in Monte Carlo simulations, and apply CJIVE to estimate the effects of pre-trial detention in Miami-Dade County.
JEL Classification: C36

*Frandsen: Department of Economics, Brigham Young University, Email: frandsen@byu.edu. **Leslie:** Department of Economics, Brigham Young University, Email: emily.leslie@byu.edu. **McIntyre:** Department of Economics, Brigham Young University, Email: samuelpmcintyre@gmail.com.

1 Introduction

Two-stage least squares (2SLS) suffers from a well-known bias towards ordinary least squares (OLS) when there are many instruments (Bekker, 1994). The bias stems from the influence each observation has on the estimation of its own first-stage fitted value. The jackknife instrumental variables estimator (JIVE), introduced by Angrist et al. (1999), reduces the bias by omitting each observation from the estimation of its first-stage fitted value. JIVE and its refinements, often referred to as "leave-out" estimation, have found widespread application in applied economics. For example, 67 out of 74 recent studies published in prominent economics journals using the popular judge fixed-effects instrumental variables strategy employed jackknife-based estimation.

JIVE eliminates many-instruments bias when observations are independent, but not when they have a clustered covariance structure. When observations are clustered, endogenous variation leaks into a given observation's first-stage fitted value from other observations in its cluster, even if the observation itself is left out. Consequently, JIVE also suffers from bias in clustered settings, and that bias can be nearly as large as that of 2SLS. The vast majority of recently published applied economics papers using JIVE-based estimation employ data with a clustered covariance structure, and so are vulnerable to this bias. Out of 99 instrumental variables papers using a jackknife or leave-out approach published in prominent economics journals, 88 use cluster-robust inference.

This article proposes a modification to jackknife instrumental variables methods that eliminates many-instruments bias in the presence of clustering: leave out each observation's entire cluster, not only the observation itself, when estimating that observation's first-stage fitted value. The estimator—which we dub CJIVE, for cluster-jackknife instrument variables estimator—is consistent and asymptotically unbiased for the familiar IV estimand under the many-instruments approximation of Bekker (1994), and therefore has a local weighted average treatment effect interpretation in a treatment effects setting with heterogeneous treatment effects.

The following section formally develops the CJIVE procedure and shows its consistency, and the inconsistency of 2SLS and JIVE under many instruments and a clustered variance structure. Section 3 demonstrates via Monte Carlo simulations the finite-sample bias of 2SLS and the usual

JIVE estimator with clustered data and shows that CJIVE performs well in finite samples. Section 4 illustrates CJIVE's use with an application to the effects of pre-trial release in Miami-Dade County.

2 Econometric Framework

We are interested in the causal effect, β, of a potentially endogenous regressor X_i (for example, an indicator for pre-trial release) on outcome Y_i, say, conviction, expressed in the following equation:

$$Y = X\beta + \varepsilon,$$

where Y is an $n \times 1$ vector of observations on Y_i, X is an $n \times 1$ vector of observations on X_i, and ε is an $n \times 1$ vector of error terms. If X_i is correlated with other factors that influence the outcome (denoted by ε_i) then least squares will fail to consistently estimate β. Identification relies on an $n \times p$ matrix Z of excluded instruments—for example, indicators for the bail judge assigned to each individual—that are related to X via the following first-stage regression equation:

$$X = Z\pi + v.$$

We take Z to be non-stochastic and define the first-stage coefficients to be the best population linear projection of X on Z, so $\pi = (Z'Z)^{-1} Z'E[X]$ and $E[Z'v] = 0$ by definition. Any additional exogenous covariates, including a constant, have already been partialed out of the outcome, the endogenous regressor, and the instruments. We adopt the familiar many-instruments asymptotic sequence where the number of instruments, p, increases with the sample size, $p/n \to \alpha$.

We assume the standard exclusion and relevance conditions for Z to serve as valid instruments:

Condition 1. *(Instrument Validity)*

(a) Exclusion: $(Z\pi)' \varepsilon/n \to_p 0$

(b) Relevance: $\pi' Z' Z\pi/n \to \kappa$ *for some* $\kappa > 0$.

Given the exclusion and relevance conditions, the causal effect β is identified by the familiar instrumental variables estimand:

$$\beta = \text{plim} \frac{(Z\pi)' Y/n}{(Z\pi)' X/n}.$$

In a constant effects model the IV estimand corresponds to the constant effect. With heterogeneous treatment effects the IV estimand identifies a convex combination of treatment effects (Kolesár, 2013).

We consider the properties of IV estimators under the assumption that the error terms have a clustered variance structure:

Condition 2. *(Clustered variance structure) $E[\varepsilon\varepsilon']$, $E[vv']$, and $E[\varepsilon v']$ are block diagonal matrices, each composed of G blocks indexed by $g \in \{1,\ldots,G\}$, where block g is of size $n_g \times n_g$ and $\sum_{g=1}^{G} n_g = n$.*

How do 2SLS, IJIVE, and CJIVE behave with many instruments and a clustered varaince structure? These three IV estimators can be written as

$$\begin{aligned}
\hat{b}^{2SLS} &= \frac{X'P_Z Y}{X'P_Z X} \\
\hat{b}^{IJIVE} &= \frac{X'C'_{IJIVE} Y}{X'C'_{IJIVE} X} \\
\hat{b}^{CJIVE} &= \frac{X'C'_{CJIVE} Y}{X'C'_{CJIVE} X},
\end{aligned}$$

where

$$\begin{aligned}
P_Z &= Z\left(Z'Z\right)^{-1} Z' \\
C_{IJIVE} &= \left(I - \text{diag}\left(P_Z\right)\right)^{-1}\left(P_Z - \text{diag}\left(P_Z\right)\right) \\
C_{CJIVE} &= \left(I - \mathbb{D}\left(P_Z, \{n_1,\ldots,n_G\}\right)\right)^{-1}\left(P_Z - \mathbb{D}\left(P_Z, \{n_1,\ldots,n_G\}\right)\right),
\end{aligned}$$

and $\mathbb{D}\left(P_Z, \{n_1,\ldots,n_G\}\right)$ is equal to P_Z, but where the elements outside of diagonal blocks of size $\{n_1,\ldots,n_G\}$ are set to zero. IJIVE stands for "improved jackknife instrumental variables estimator" and is equivalent to Angrist et al.'s (1999) JIVE in the absence of covariates. When there are covariates, the IJIVE procedure partials them out of the outcome, endogenous regressor, and instruments, as we have done here, leading to improved performance relative to JIVE (Ackerberg and Devereux, 2009). CJIVE is equivalent to a procedure where first-stage fitted values are calculated for each observation from a regression that leaves out the observation's entire cluster.

Our main result, proved in the appendix, is that with many instruments and a clustered variance structure, 2SLS and IJIVE are inconsistent, but CJIVE is not. That is, under many-instrument asymptotics and the conditions described above there are data-generating processes for which

$$\operatorname{plim} \hat{b}^{2SLS} \neq \beta$$

and

$$\operatorname{plim} \hat{b}^{IJIVE} \neq \beta,$$

while for all such data-generating processes

$$\operatorname{plim} \hat{b}^{CJIVE} = \beta.$$

The result means that whenever clustered inference is required, leave-one-out jackknife IV estimators may fail to eliminate many-instruments bias. The intuition is that because unobserved determinants of the outcome and the endogenous regressor (that is, ε_i and v_i) are correlated across observations within clusters, it is not sufficient to leave out only a given observation when estimating that observation's first-stage fitted value; the influence of observations in the same cluster contaminates the fitted value with endogenous variation. The remedy is to leave out the entire cluster when estimating a given observation's fitted value.

3 Monte Carlo Simulations

To illustrate the finite-sample performance of the proposed CJIVE procedure and compare it to existing 2SLS and JIVE estimators, we run a series of Monte Carlo simulations. In the simulation, each individual i belongs to a cluster c, and has an outcome determined by

$$Y_{ic} = \beta X_{ic} + \varepsilon_{ic}$$

and endogenous regressor determined by

$$X_{ic} = \Phi\left(-v_{ic}\right) \leq g_c.$$

Φ denotes the standard normal cdf. The error terms are generated by

$$\begin{aligned}
\varepsilon_{ic} &= \rho v_{ic} + \sqrt{1-\rho^2} N\left(0,1\right) \\
v_{ic} &= v_i + v_c \\
v_i &\sim N\left(0, \sigma_I^2\right) \\
v_c &\sim N\left(0, \sigma_C^2\right).
\end{aligned}$$

The terms $N\left(0,1\right), N\left(0,\sigma_I^2\right), N\left(0,\sigma_C^2\right)$ are independent normal random variables. The p-vector of instruments Z_c are indicators for each value of the integer-valued random variable $J_c \in \{0,\ldots,p\}$, which is assigned to each cluster c with a uniform distribution. Finally, $g_c = \left(J_c + 1\right)/p$.

This setup satisfies Conditions 1 and 2. The parameter ρ determines the degree of endogeneity, which the simulations vary between zero and $.5$. When ρ is positive, OLS will be upward biased. The parameter σ_C^2 determines the importance of the cluster-level dependence in the error terms, which the simulations vary between zero and one. In the simulations we vary the number of instruments, p between 25 and 300 to assess how the number of instruments affects the relative performance of the estimators. We also vary the number of clusters between 100 and 5,000. In all cases we maintain a sample size of $10,000$, divided approximately equally among the clusters. We fix the true effect size at $\beta = .3$.

Figure 1 plots the means of 2SLS, IJIVE, and CJIVE across the simulation parameters described above. The upper-left panel shows how the relative performance of the estimators depends on the number of clusters, holding the number of instruments at $p = 50$. IJIVE and and 2SLS are consistently upward biased (towards OLS), while CJIVE is close to unbiased throughout the range. The bias of IJIVE and 2SLS is particularly pronounced when the number of clusters is below 1,000. As the number of clusters grows, holding the number of observations fixed, there are fewer observations per cluster, and the bias of IJIVE declines because clusters are closer and closer to observations.

The upper-right panel shows how the bias depends on the number of instruments, p, setting the number of clusters to 1,000 (10 observations per cluster). The bias of 2SLS increases steeply with the number of instruments, as is well known. IJIVE is also upward biased, only slightly less so than 2SLS.

The lower-left panel shows that the bias of 2SLS and IJIVE increases as the importance of clustering (parameterized by σ_C) increases. When $\sigma_C = 0$, IJIVE is unbiased, as expected. 2SLS is biased because of the usual many-instruments bias. But as σ_C increases, so does the bias of both 2SLS and IJIVE, while CJIVE remains unbiased.

Finally, the lower-right panel of Figure 1 shows that the bias of both 2SLS and IJIVE increases with the degree of endogeneity, parmeterized by ρ. When $\rho = 0$, all three estimators are unbiased. This is to be expected, since OLS itself would be unbiased in this case. As ρ increases, CJIVE remains unbiased, but 2SLS and IJIVE become biased. IJIVE is only slightly less biased than 2SLS.

4 Empirical Example

In this section, we illustrate the use of the cluster jackknife procedure in an empirical example studying the effect of pretrial detention using a judge leniency design. As in Dobbie et al. (2018), we use a sample of court records from Miami-Dade County in Florida over the period 2006-2014. For the 70 percent of defendants who do not post bail immediately, there is a bail hearing within 24 hours of arrest. The bail judge at the hearing may change the amount or impose additional conditions for bail.

The judge assignment pattern varies based on whether hearings are scheduled during weekdays or weekends. A single judge presides over all weekday hearing shifts. During the weekend, judges preside over shifts based on a rotating schedule. Thus, defendants in the same shift are assigned to the same judge as a group. There is little scope for manipulating judge or shift assignment given the short window between arrest and hearings. Note that bail hearings are unrelated to the process of trial judge assignment so there is no mechanical relationship between the pretrial hearing process and later stages of a case.

The court records contain information on arrest charges, bail judges, bail amount and type, if and when bail was posted, as well as defendant characteristics such as name, gender, and race. We restrict our attention to cases that are assigned weekend bail hearings as these are cases where bail judges are assigned based on a rotating schedule, and to cases whose bail judge saw at least 200 cases in our sample period. The sample consists of 91,421 defendants coming before 146 unique bail judges.

We want to estimate the causal effect of pretrial release on conviction, as specified in the following model:

$$Y_i = \beta X_i + W_i'\gamma + \varepsilon_i$$

where Y_i is an indicator for whether person i is convicted, X_i is an indicator for being released pretrial, and W_i is a vector of controls, including a constant. Defendants who are released pretrial are 26 percentage points less likely to be convicted of a crime in the current case. There are also systematic differences in the demographic and criminal history characteristics of defendants who are released compared with those who are detained pretrial. The coefficient of interest is β. OLS will produce biased estimates if there are unobserved factors that are correlated with both release and conviction. To estimate the causal effect of pretrial release on case outcomes without selection bias, we leverage variation in judge leniency by using the vector of judge dummies Z_i as instruments for pretrial release. We cluster at the shift level, because this is the level at which judges are assigned (Abadie et al., 2023).

For the judge dummies to be valid instruments, it must be the case that judge assignment affects pretrial status (relevance) and that judges are as good as randomly assigned and do not influence case outcomes through channels other than pretrial status assignment (exclusion). Bail judges have wide discretion in setting the terms of pretrial release, and are therefore well-positioned to influence whether defendants are released or detained. Given the short timeline from arrest to bail hearing, the variability in judge scheduling, and the limited scope of bail hearings described above, it is plausible that judges are as good as randomly assigned to shifts and do not have a meaningful impact on case outcomes except through pretrial status assignment. We include year by day-of-week by court (felony or misdemeanor) fixed effects (“time-by-place fixed effects") to allow for the possibility that some judges may not be available for bail hearing shifts in all years or on particular weekend days, or that some judges may work primarily in one court or the other. While we assume that judge assignment to shifts is random conditional on time-by-place fixed effects, patterns in criminal activity or policing could give rise to within-shift correlations in defendant characteristics.

Table 1 displays the results from estimating the effect of pretrial release on conviction. Column 1 shows the results from an OLS regression of conviction on release. After controlling for time-by-place fixed effects, defendants who are released are about 23 percentage points less likely to be convicted. Instrumenting for release with the vector of judge dummies, without jackknifing, produces an estimated effect that is somewhat larger in magnitude (28 percentage points). Jackknifing at the individual level changes the result only slightly (30 percentage points). Cluster jackknifing results in a noticeably different estimate (44 percentage points).

The pattern of results is consistent with the possibility that defendant or case characteristics that are relevant for pretrial release and case outcomes are correlated within shift. The median judge in our sample presides at a dozen shifts during the sample period, so each shift contributes meaningfully to the estimated propensities to release. When we jackknife at the individual level, we are allowing the other defendants in the same shift to influence the value of the leniency instrument for each person, even though shift-peers may have systematically similar unobserved characteristics. This introduces the same kind of bias that is present in OLS or 2SLS without jackknifing. Jackknifing on clusters solves this endogeneity problem. Table 2 reports results from a balance test demonstrating that defendant demographics, prior offender status, and details of the current case do not (jointly) have a significant correlation with the continuous CJIVE instrument used in column 4 of Table 1.

5 Conclusion

Jackknife, or leave-out procedures are routinely used to combat the well-known bias of 2SLS towards OLS when there are many instruments. At the same time, the vast majority of studies using jackknife IV procedures also cluster their inference. When inference must be clustered, observation-level jackknife IV procedures fail to eliminate many-instruments bias. We proposed a cluster-level jackknife procedure (CJIVE) that is consistent under many instruments and a clustered variance structure. The procedure performs much better than existing estimators in simulations, and makes a substantive difference in our application on the effects of pre-trial release in Miami-Dade county.

Tables and Figures

Table 1: Results using various estimators

	OLS	Judge dummies	IJIVE	CJIVE
Released	-0.232	-0.275	-0.299	-0.444
	(0.007)	(0.066)	(0.107)	(0.206)

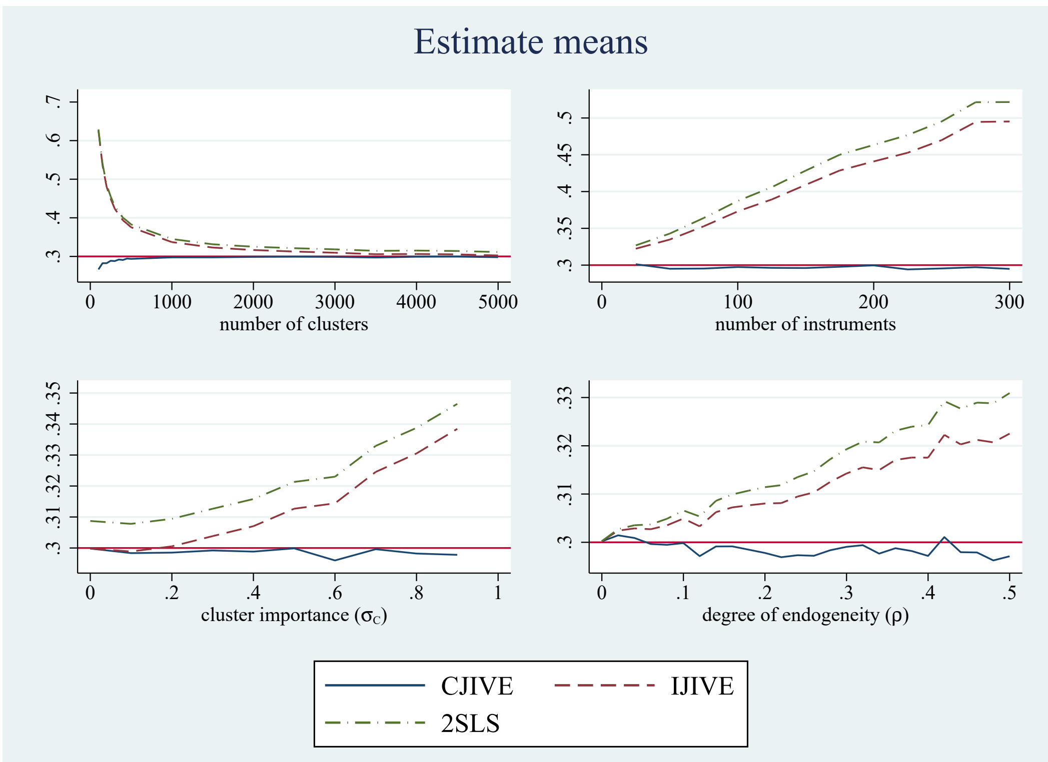
Note: Column 1 shows empirical example results from an OLS regression of an indicator for being convicted of any charge on an indicator for being released pretrial. Column 2 shows estimates from a 2SLS regression of the conviction indicator on the pretrial release indicator, where a vector of judge dummies serve as instruments for the pretrial release indicator. Column 3 shows estimates from using IJIVE to jackknife at the individual level. Column 4 shows estimates from using our proposed CJIVE estimator to jackknife at the cluster level. All specifications include a vector of time-by-place fixed effects.

Table 2: Balance test

Male	-0.000
	(0.000)
Black	-0.000
	(0.000)
Age	-0.000
	(0.000)
Prior offender	-0.000
	(0.000)
Number of counts	0.000
	(0.000)
Felony charge	-0.006
	(0.006)
Drug charge	0.001
	(0.000)
Violent charge	0.000
	(0.000)
Property charge	0.001
	(0.000)
Joint F stat	1.53
p-value	.132

Note: This table reports regression coefficients from regressing the CJIVE instrument constructed for the estimation in column 4 of Table 1 on a vector of demographic and case characteristics, along with a vector of time-by-place fixed effects. The joint F statistic a p-value correspond to a test of the joint significance of the demographic and case characteristics.

Figure 1: Monte Carlo Simulation Results



Note: The figure reports Monte Carlo means of the IV estimators indicated from the simulation design described in the text. Results are based on 1,000 iterations.

References

Alberto Abadie, Susan Athey, Guido W Imbens, and Jeffrey M Wooldridge. When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35, 2023.

Daniel A Ackerberg and Paul J Devereux. Improved JIVE Estimators for Overidentified Linear Models with and without Heteroskedasticity. *The Review of Economics and Statistics*, 91(2): 351–362, 05 2009. ISSN 0034-6535. doi: 10.1162/rest.91.2.351. URL `https://doi.org/10.1162/rest.91.2.351`.

J. D. Angrist, G. W. Imbens, and A. B. Krueger. Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67, 1999. ISSN 08837252, 10991255. URL `http://www.jstor.org/stable/223249`.

Paul A. Bekker. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica*, 62(3):657–681, 1994. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/2951662`.

Will Dobbie, Jacob Goldin, and Crystal S. Yang. The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *The American Economic Review*, 108(2):201–240, 2018. ISSN 00028282, 19447981. URL `https://www.jstor.org/stable/26527904`.

Michal Kolesár. Estimation in an instrumental variables model with treatment effect heterogeneity, 2013. unpublished working paper.

Appendix

Before proving the main result we establish the following useful lemma:

Lemma 1. *$X'C' = \pi'Z' + v'C'$ for $C \in \{P_Z, C_{IJIVE}, C_{CJIVE}\}$.*

Proof. Substituting $X = Z\pi + v$, $X'C' = \pi'Z'C' + v'C'$ and straightforward algebra shows that $\pi'Z'C' = \pi'Z'$ for each $C \in \{P_Z, C_{IJIVE}, C_{CJIVE}\}$. □

We assume the following high-level condition on the probability limits of $X'C'X/n$ and $v'C'\varepsilon/n$:

Condition 3. *For $C \in \{P_Z, C_{IJIVE}, C_{CJIVE}\}$, $\frac{1}{n}X'C'X$ converges in probability to a finite strictly positive value and $\operatorname{plim} \frac{1}{n}v'C'\varepsilon = \lim_{n\to\infty} E\left[\frac{1}{n}v'C'\varepsilon\right] < \infty$.*

With the lemma and regularity condition in hand, we can establish our main result:

Theorem 1. *Suppose Conditions 1, 2, and 3 hold. Then* $\hat{b}^{2SLS}$ *and* $\hat{b}^{IJIVE}$ *are inconsistent and* $\hat{b}^{CJIVE}$ *is consistent under many-instrument asymptotics.*

Proof. The probablity limit of any IV estimator of the form $\hat{b} = X'C'Y/\left(X'C'X\right)$ can be written

$$\text{plim}\,\hat{b} = \beta + \frac{\text{plim}\,\frac{1}{n}X'C'\varepsilon}{\text{plim}\,\frac{1}{n}X'C'X}.$$

By Lemma 1 and the exclusion restriction in Condition 1 we have, for $\hat{b} \in \left\{\hat{b}^{2SLS}, \hat{b}^{IJIVE}, \hat{b}^{CJIVE}\right\}$

$$\begin{aligned} \text{plim}\,\hat{b} - \beta &= \frac{\text{plim}\,\frac{1}{n}\left(\pi'Z' + v'C'\right)\varepsilon}{\text{plim}\,\frac{1}{n}X'C'X} \\ &= \frac{\text{plim}\,\frac{1}{n}v'C'\varepsilon}{\text{plim}\,\frac{1}{n}X'C'X}, \end{aligned}$$

where the second equality follows from the exclusion restriction. By Condition 3 the denominator is finite and strictly positive, so consistency of the three IV estimators hinges on the numerator, which by Condition 3 is equal to

$$\lim_{n\to\infty} E\left[\frac{1}{n}v'C'\varepsilon\right] = \lim_{n\to\infty}\frac{1}{n}trace\left(C'E\left[\varepsilon v'\right]\right),$$

where the equality uses the fact that a scalar is equal to its trace, and the trace is invariant to conformable cyclical permutations.

First take 2SLS, so $C = P_Z$. To show that 2SLS is inconsistent, it suffices to show that $\lim_{n\to\infty}\frac{1}{n}trace\left(C'E\left[\varepsilon v'\right]\right) \neq 0$ for some data-generating process that satisfies the conditions. For this purpose, let $E\left[\varepsilon v'\right]$ be diagonal where each diagonal element is a constant: $\sigma_{\varepsilon v}$ (corresponding to homoskedastic iid data). Then $C'E\left[\varepsilon v'\right] = \sigma_{\varepsilon v}P_Z$. Noting that the trace of P_Z is equal to p (the number of instruments), we have

$$\lim_{n\to\infty}\frac{1}{n}trace\left(C'E\left[\varepsilon v'\right]\right) = \lim_{n\to\infty}\frac{1}{n}\sigma_{\varepsilon v}p = \alpha\sigma_{\varepsilon v} \neq 0.$$

This establishes the familiar many-instruments bias of 2SLS towards OLS.

Next, take IJIVE, so $C = C_{IJIVE}$. Again, to show that IJIVE is inconsistent, it suffices to show that $\lim_{n\to\infty} \frac{1}{n} trace\left(C' E\left[\varepsilon v'\right]\right) \neq 0$ for some data-generating process that satisfies the conditions. For this purpose, take the special case where $E\left[\varepsilon v'\right]$ is block diagonal with a constant $\sigma_{\varepsilon v}$ within the clusters defined by the blocks, and each cluster has $n_g = 2$ observations. Let Z be a set of p mutually exclusive and collectively exhaustive binary indicators for equal-sized groups from which a constant has been partialed out. Let Z_i be constant within clusters, as in an IV design where all observations within a cluster have the same value of the instrument. The elements of C in this case equal zero on the diagonal by construction, $p/\left(n-p\right)$ off the diagonal in elements (i, j) where $Z_i = Z_j$ (within groups defined by the instruments) and $-1/\left(n-p\right)$ in all other off-diagonal elements. We therefore have that

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} trace\left(C' E\left[\varepsilon v'\right]\right) &= \lim_{n\to\infty} \frac{p}{n-p}\sigma_{\varepsilon v} \\ &= \frac{\alpha}{1-\alpha}\sigma_{\varepsilon v} \neq 0.\end{aligned}$$

Finally, let $C = C_{CJIVE}$. By construction, C_{CJIVE} has zeros along the block diagonal elements where $E\left[\varepsilon v'\right]$ is nonzero (by Condition 2). Letting $C'_{i,j}$ and $E\left[\varepsilon v'\right]_{i,j}$ denote the (i, j)-elements of C' and $E\left[\varepsilon v'\right]$, the product $C'_{i,j} E\left[\varepsilon v'\right]_{i,j} = 0$ for every element (i, j), and therefore

$$trace\left(C' E\left[\varepsilon v'\right]\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} C'_{i,j} E\left[\varepsilon v'\right]_{i,j} = 0,$$

and CJIVE is consistent. □