Jackknife Estimation of a Cluster-Sample IV Regression Model with Many Weak Instruments*

John C. Chao, Norman R. Swanson, and Tiemen Woutersen

First Draft: September 5, 2020 *This Version*: December 20, 2022

Abstract

This paper proposes new jackknife IV estimators that are robust to the effects of many weak instruments and error heteroskedasticity in a cluster sample setting with cluster-specific effects and possibly many included exogenous regressors. The estimators that we propose are designed to properly partial out the cluster-specific effects and included exogenous regressors while preserving the re-centering property of the jackknife methodology. To the best of our knowledge, our proposed procedures provide the first consistent estimators under many weak instrument asymptotics in the setting considered. We also present results on the asymptotic normality of our estimators and show that t-statistics based on said estimators are asymptotically normal under the null and consistent under fixed alternatives. Monte Carlo results show that our t-statistics perform better in controlling size in finite samples than those based on alternative jackknife IV procedures previously introduced in the literature.

Keywords: Cluster sample, instrumental variables, heteroskedasticity, jackknife, many weak instruments, panel data

JEL classification: C12, C13, C23, C26, C38

1 Introduction

The problem of endogeneity remains central to research in economics and econometrics. The key reason for this is that there are many different regression settings for which endogeneity is

*Corresponding author: John C. Chao, Department of Economics, 7343 Preinkert Drive, University of Maryland, chao@econ.umd.edu. Norman R. Swanson, Department of Economics, 9500 Hamilton Street, Rutgers University, nswanson@econ.rutgers.edu. Tiemen Woutersen, Department of Economics, 1130 E Helen Street, University of Arizona, woutersen@arizona.edu. The authors would like to thank the Editor, Xiaohong Chen, an Associate Editor, and an anonymous referee for very helpful comments and suggestions. The authors also owe special thanks to Jerry Hausman and Whitney Newey for many discussions on the topic of this paper over a number of years. In addition, thanks are owed to the particpants of the 2019 MIT Conference Honoring Whitney Newey for comments on and advice given on an earlier version of this work. Finally, the authors wish to thank Miriam Arden for excellent research assistance. Chao thanks the University of Maryland for research support, and Woutersen's work was supported by an Eller College of Management Research Grant.

an issue, but for which valid estimators are not currently available. One such setting involves the case where the objective is to estimate an IV regression with fixed effects using panel or cluster-sampled data in situations where the number of available instruments may be large, but where the instruments themselves are all only weakly correlated with the endogenous regressors. There is now a substantial literature on estimation and inference under many weak instruments, including Chao and Swanson (2005), Stock and Yogo (2005), Hansen, Hausman, and Newey (2008), Hausman et al. (2012), Chao et al. (2012, 2014), Bekker and Crudu (2015), Crudu, Mellace, and Sándor (2020), and Mikusheva and Sun (2021). However, the analyses given in these papers are for cross-sectional data, thus precluding panel data or cluster sampling settings where there is additional unobserved heterogeneity modeled by fixed or cluster-specific effects. Moreover, even in the cross-sectional context, 2SLS and the LIML estimators are not well behaved under many weak instruments. In particular, Chao and Swanson (2005) and Stock and Yogo (2005) show that the 2SLS estimator is inconsistent under many weak instrument asymptotics, even when the errors are homoskedastic. In addition, Hausman et al. (2012) and Chao et al. (2012) point out that LIML is also inconsistent under many weak instruments when there is error heteroskedasticity. Estimators which are currently known to be robust to the effects of many weak instruments in cross sectional settings with error heteroskedasticity all have a jackknife form, as discussed in Chao and Swanson (2004), Chao et al. (2012), and Hausman et al. (2012). These include the JIVE1 and JIVE2 estimators studied in Angrist, Imbens, and Krueger (1999), for example. For further discussion, see Phillips and Hale (1977), Blomquist and Dahlberg (1999), Ackerberg and Devereux (2009), and Bekker and Crudu (2015). These papers again only study various versions of the jackknife IV estimator in a cross-sectional setup without fixed effects.

The goal of this paper is to consider the problem of many weak instruments in a panel data or cluster-sampling framework with fixed or cluster specific effects. In addition to the presence of unobserved heterogeneity, our setup allows for additional (included) exogenous regressors which appear in both the outcome, or structural, equation and in the first-stage equations. To consistently estimate the structural parameter vector of interest in an IV regression with fixed or cluster-specific effects, we propose three new estimators, which we refer to by the acronyms FEJIV, FELIM, and FEFUL. These estimators are so named as they are modified versions and generalizations, respectively, of the jackknife IV (JIV), the LIML, and the Fuller (1977) estimators. In contrast to the original JIV, LIML, and Fuller estimators, our new estimators are designed to be robust to the effects of many weak instruments and error heteroskedasticity, even in the presence of additional complications caused by having fixed or cluster-specific effects and many included exogenous regressors. To achieve consistency in our setting requires an estimator that not only properly partials out additional covariates and cluster-specific effects, but at the same time must also be properly centered in a form similar to a degenerate U-statistic. It turns out that accomplishing both of these objectives simultaneously is quite challenging. While a number of innovative JIV-type estimators have been proposed recently (see, for example, the improved jackknife estimator IJIVE of Ackerberg and Devereux (2009), as well as the UJIVE estimator of Kolesár (2013)), due to the aforementioned difficulties, these estimators are not consistent when applied to our setting under many weak instrument asymptotics, as we shall elaborate on in greater detail in Section 3. On the other hand, the estimation procedures that we introduce here are carefully designed to properly partial out fixed or cluster-specific effects and included exogenous regressors, while preserving the re-centering property of the jackknife methodology. To the best of our knowledge, the estimators presented here are the first consistent estimators under many weak instrument asymptotics in an IV regression model with fixed or cluster-specific effects and possibly many included exogenous regressors. In addition to consistency, we also establish the asymptotic normality of the FELIM and FEFUL estimators[1].

This paper also provides a number of results showing that hypothesis testing procedures based on FELIM and FEFUL are robust to the effects of many weak instruments. In particular, we construct t-statistics based on these two estimators and show that, when the null hypothesis is true, these t-statistics converge to an asymptotic standard normal distribution under both many weak instrument asymptotics and also standard asymptotics. Moreover, our t-statistics are shown to be consistent in the sense that under fixed alternatives they diverge, with probability approaching one, in the direction of the alternative hypothesis.

The many-weak-instrument asymptotic framework used in the sequel to analyze the performance of FELIM and FEFUL was first proposed in Chao and Swanson (2005). This framework extends earlier work by Morimune (1983) and Bekker (1994) on what has become known in the IV literature as the many-instrument asymptotics or "Bekker asymptotics", whereby a large sample approximation is carried out by considering an alternative sequence where the number of instruments is allowed to approach infinity as the sample size grows to infinity. A key difference between the Bekker asymptotic framework and the many-weak-instrument asymptotic framework is the rate of growth of the so-called concentration parameter. As has been pointed out by Phillips (1983) and Rothenberg (1984), among others, the concentration parameter is the natural measure of instrument strength in a linear IV model. In the original papers by Morimune (1983) and Bekker (1994),

[1] We do not provide a formal proof of the asymptotic normality of the FEJIV estimator because the results of our Monte Carlo study, as reported in Section 6, show that FELIM and FEFUL tend to have better finite sample properties than FEJIV. For this reason, we shall focus the presentation of our theoretical results on FELIM and FEFUL only. However, one can easily show, by slightly modifying the arguments that we give for FELIM and FEFUL, that FEJIV is also asymptotically normal, under many weak instrument asymptotics. Note also that our simulation finding regarding the properties of FEJIV are consistent with the findings of Davidson and MacKinnon (2006).

the concentration parameter is assumed to grow at the same rate as the sample size, which is also what is assumed under standard (strong but fixed number of instruments) asymptotics, whereas the many-weak-instrument asymptotic framework allows the concentration parameter to grow at a rate much slower than the sample size, thus allowing for much weaker instruments. Let μ_n^2 be a sequence that gives the rate of growth of the concentration parameter, and let $K_{2,n}$ denote the number of instruments. Chao and Swanson (2005) show that for consistent point estimation to be possible, a sufficient condition is $\sqrt{K_{2,n}}/\mu_n^2 \to 0$, as $K_{2,n},\ \mu_n^2, n \to \infty$. This allows for the possibility that μ_n^2 is of an order smaller than $K_{2,n}$ which, in turn, can be of an order much smaller than the sample size n. The original Bekker framework, on the other hand, requires $K_{2,n}, \mu_n^2$, and n to all be of the same order of magnitude. Recent work by Mikusheva and Sun (2021) indicates that the condition $\sqrt{K_{2,n}}/\mu_n^2 \to 0$, as $K_{2,n},\ \mu_n^2, n \to \infty$ is not only sufficient but also necessary for consistency in point estimation and hypothesis testing.[2]

The rest of the paper is organized as follows. Section 2 provides some brief motivation for our paper. Section 3 states the model, defines the FELIM, FEFUL, and FEJIV estimators, and provides an explanation of how our estimators improve upon various alternative jackknife IV estimators that have previously been proposed in the literature. Analytical results presented in Section 4 establish that our estimators are consistent and asymptotically normally distributed. Section 5 shows how to estimate the variances of the estimators and also provides asymptotic results for t-statistics based on our estimators. Section 6 contains the results of a series of Monte Carlo experiments in which the relative performance of our estimators is compared with that of extant estimators in the literature. Section 7 concludes. Proofs of Theorem 1, Corollary 1, Theorems 4-5, and Corollaries 2-3 are presented in the Appendix to this paper. The proofs of Theorems 2 and 3 are longer and are given in a supplemental Appendix[3].

Before proceeding, we will first say a few words about some of the commonly used notations in this paper. In what follows, we use $\lambda_{\min}(A)$, $\lambda_{\max}(A)$, and $tr(A)$ to denote, respectively, the minimal eigenvalue, the maximal eigenvalue, and the trace of a square matrix A, whereas A' denotes the transpose of a (not necessarily square) matrix A. $\|a\|_2$ denotes the usual Euclidean norm when applied to a (finite-dimensional) vector a. On the other hand, for a matrix A, $\|A\|_2 \equiv \max\left\{\sqrt{\lambda(A'A)} : \lambda(A'A) \text{ is an eigenvalue of } A'A\right\}$ denotes the matrix spectral norm, while $\|A\|_F \equiv \sqrt{tr\{A'A\}}$ denotes the Frobenius norm and $\|A\|_\infty \equiv \max_{1 \le i \le m_n} \sum_{j=1}^{m_n} |a_{ij}|$ (i.e.,

[2] An alternative to the asymptotic framework considered here is the weak instrument asymptotic framework proposed in Staiger and Stock (1997). The Staiger-Stock framework considers a setting where $\mu_n^2 = O(1)$, in which case the IV model is not point identified. We do not consider the Staiger-Stock framework in this paper because our focus is on consistency of point estimation and on test consistency.

[3] The supplemental Appendix can be viewed at the URL: http://econweb.umd.edu/~chao/Research/research_files/Supplemental_Appendix_to_Jackknife_Estimation_Cluster_Sample_IV_Model_December_20_2022.pdf

the maximal row sum of an $m_n \times m_n$ matrix). In addition, we use $A \circ B$ to denote the Hadamard product of two conformable matrices A and B (i.e., $A \circ B \equiv [a_{ij} b_{ij}]$, for $A = [a_{ij}]$ and $B = [b_{ij}]$). We take $D(a)$ to be a diagonal matrix whose diagonal elements correspond with the elements of the vector a, while $D(A)$ is taken to be a diagonal matrix whose diagonal elements are the same as the diagonal elements of the square matrix A. Furthermore, we will let $\iota_p = (1, 1, ..., 1)'$ denote a $p \times 1$ vector of ones, and we take the shorthand *a.s.n.* to mean almost surely for all n sufficiently large. Finally, we use CS and T, respectively, to denote the Cauchy-Schwarz and the triangle inequality, and the abbreviation w.p.a.1 stands for "with probability approaching one".

2 Some Background and Motivation

In this section, we briefly discuss some of the issues that arise when one needs to partial out additional covariates in a setting with many weak instruments, with the hope that such a discussion will provide the necessary background to help readers gain a stronger intuitive feel for the estimation procedures which we will introduce in subsequent sections. To offer a point of contrast, we will start by first reviewing some basic aspects of IV estimation under many weak instruments in the context of a simple, cross-sectional model with a single endogenous regressor and no additional covariate, i.e.,

$$\begin{aligned} \underset{n\times 1}{y} &= \underset{1\times 1}{\delta_0} \underset{n\times 1}{x} + \underset{n\times 1}{\varepsilon}, \\ \underset{n\times 1}{x} &= \underset{n\times K_2}{Z_2} \underset{K_2\times 1}{\pi_n} + \underset{n\times 1}{u} \end{aligned}$$

Here, y is vector of observations on the outcome variable, x is the vector of observations on the endogenous regressor, and Z_2 is a non-random matrix of observations on the K_2 instruments. In addition, we intentionally specify the coefficient vector π_n of the first-stage equation to depend on n to allow for local-to-zero modeling of weak instruments[4]. Even in this simple setup, it is well-known that, in the presence of many weak instruments and error heteroskedasticity, the usual IV-type estimator such as 2SLS and LIML will not have desirable asymptotic properties. To see this, consider the case of the 2SLS estimator, which in this case, can be decomposed as

$$\widehat{\delta}_{2SLS} - \delta_0 = \left(x' P^{Z_2} x\right)^{-1} x' P^{Z_2} \varepsilon = \left(\pi_n' Z_2' Z_2 \pi_n + 2\pi_n' Z_2' u + u' P^{Z_2} u\right)^{-1} \left(\pi_n' Z_2' \varepsilon + u' P^{Z_2} \varepsilon\right) \tag{1}$$

[4]See Assumption 3 in section 3 for the type of (generalized) local-to-zero structure which we assume for the more general cluster-sample/panel-data IV regression setting studied in this paper.

where $\widehat{\delta}_{2SLS}$ is of course obtained by minimizing the objective function $\widehat{Q}_{2SLS}(\delta) = (y - x'\delta)' P^{Z_2} (y - x'\delta)$ with $P^{Z_2} = Z_2 (Z_2' Z_2)^{-1} Z_2'$. Under conventional asymptotics with a fixed number of strong instruments, the asymptotic behavior of the denominator $x' P^{Z_2} x$ will be dominated by the concentration parameter $\pi_n' Z_2' Z_2 \pi_n$ which in this case grows at the rate of the sample size n, whereas $\pi_n' Z_2' \varepsilon = O_p(\sqrt{n})$ and $u' P^{Z_2} \varepsilon = O_p(1)$ so that, in some sense, the signal in the denominator overwhelms the noise elements in the numerator, leading to the consistency of the 2SLS estimator. Viewed from this perspective, the problem caused by having weak instruments is that the signal component as represented by the concentration parameter $\pi' Z_2' Z_2 \pi$, is now weaker and grows at some rate μ_n^2 which is much slower than n. On the other hand, the problem caused by many instruments is that it inflates one of the noise components $u' P^{Z_2} \varepsilon$ which now grows, in probability, at the rate K_2. This combination of having stronger noise and a weaker signal can then lead to inconsistency of the 2SLS estimator when $\mu_n^2 / K_2 = O(1)$. Note also that under conventional, strong-instrument asymptotics the term $u' P^{Z_2} \varepsilon$ is of a lower order relative to $\pi_n' Z_2' \varepsilon$ but this will no longer be true when $\mu_n^2 / K_2 = O(1)$, so having sufficiently many weak instruments leads to a reshuffling of the order of magnitude of the terms in the numerator of expression (1).

Now, one way to fix this problem in the case with no additional covariates is to use one of the JIVE estimators proposed in Angrist, Imbens, and Krueger (1999). As an illustration, consider the JIVE2 estimator proposed in that paper which can be obtained by minimizing a modified 2SLS objective function whereby the diagonal elements of the projection matrix P^{Z_2} are removed; that is, the JIVE2 estimator is obtained by minimizing the objective function

$$\widehat{Q}_{JIVE2}(\delta) = \left(y - x'\delta\right)' \left[P^{Z_2} - D\left(P^{Z_2}\right)\right] \left(y - x'\delta\right)$$

where $D\left(P^{Z_2}\right)$ is the diagonal matrix whose diagonal elements are the same as those of P^{Z_2}. The reason why such "jackknife-type" modification helps is that if we do a decomposition of JIVE2 similar to the decomposition given for 2SLS in expression (1) above, we obtain

$$\widehat{\delta}_{JIVE2} - \delta_0 = \left(x' \left[P^{Z_2} - D\left(P^{Z_2}\right)\right] x\right)^{-1} \left(\pi_n' Z_2' \left[P^{Z_2} - D\left(P^{Z_2}\right)\right] \varepsilon + u' \left[P^{Z_2} - D\left(P^{Z_2}\right)\right] \varepsilon\right).$$

Comparing the JIVE2 bilinear term $u' \left[P^{Z_2} - D\left(P^{Z_2}\right)\right] \varepsilon$ with its counterpart $u' P^{Z_2} \varepsilon$ for the 2SLS estimator, we see that the former has a smaller order of magnitude than the latter under a many instrument asymptotic regime, so that, in particular, $u' \left[P^{Z_2} - D\left(P^{Z_2}\right)\right] \varepsilon = O_p\left(\sqrt{K_2}\right)$ whereas $u' P^{Z_2} \varepsilon = O_p(K_2)$. The reason why this is the case is related to the so-called concentration of measure phenomenon that has been studied in the probability literature. Note that, under the assumption that (ε_i, u_i) is independent of (ε_j, u_j) for all $i \neq j$ (where ε_i and u_i denote the i^{th}

component of ε and u respectively); $E\left[u'\left[P^{Z_2} - D\left(P^{Z_2}\right)\right]\varepsilon\right] = 0$, even under heteroskedasticity, whereas $E\left[u'P^{Z_2}\varepsilon\right] \neq 0$, so that the former, being a properly centered bilinear form, will have a lower order of magnitude than the latter, which is not properly centered at zero[5]. It follows that JIVE2 will be more robust to the effects of many weak instruments in the sense that it will be consistent as long as the concentration parameter grows fast enough so that $\sqrt{K_2}/\mu_n^2 \to 0$, whereas the consistency of the 2SLS requires the stronger condition that $K_2/\mu_n^2 \to 0$.

Consider next a more realistic model with additional covariates

$$
\begin{aligned}
\underset{n\times 1}{y} &= \underset{1\times 1}{\delta_0}\underset{n\times 1}{x} + \underset{n\times K_1}{Z_1}\underset{K_1\times 1}{\varphi} + \underset{n\times 1}{\varepsilon}, \\
\underset{n\times 1}{x} &= \underset{n\times K_1}{Z_1}\underset{K_1\times 1}{\beta} + \underset{n\times K_2}{Z_2}\underset{K_2\times 1}{\pi_n} + \underset{n\times 1}{u}
\end{aligned}
$$

To see why it is not as straightforward as one might think to generalize the JIVE2 estimator discussed previously to this setting, consider the IJIVE2 (the improved JIVE2) estimator discussed in Evdokimov and Kolesár (2018). To construct the IJIVE2 estimator, one first partials out the covariates Z_1 to obtain the system of equations

$$
\widetilde{y} = \delta_0\widetilde{x} + \widetilde{\varepsilon} \tag{2}
$$

$$
\widetilde{x} = \widetilde{Z}_2\pi + \widetilde{u} \tag{3}
$$

(where $\widetilde{y} = M^{Z_1}y$, $\widetilde{x} = M^{Z_1}x$, $\widetilde{Z}_2 = M^{Z_1}Z_2$, $\widetilde{\varepsilon} = M^{Z_1}\varepsilon$, $\widetilde{u} = M^{Z_1}u$ and $M^{Z_1} = I_n - Z_1\left(Z_1'Z_1\right)^{-1}Z_1'$) and then construct a JIVE2 estimator based on the representation given in expressions (2)-(3). It is easy to see that this estimation strategy leads equivalently to an estimator that minimizes that objective function

$$
\widehat{Q}_{IJIVE2}(\delta) = \left(\widetilde{y} - \widetilde{x}'\delta\right)'\left[P^{\widetilde{Z}_2} - D\left(P^{\widetilde{Z}_2}\right)\right]\left(\widetilde{y} - \widetilde{x}'\delta\right)
$$

and the deviation of this IJIVE2 estimator, $\widehat{\delta}_{IJIVE2}$, from the true value, δ_0, can be decomposed

[5] To give perhaps a more familiar example of the concentration of measure phenomenon, we can consider a simple case where $W_1, ..., W_n$ is a sequence of independent random variables such that $\sup_i E\left[W_i^2\right] < \infty$ and $E\left[W_i\right] \neq 0$ for all i. In this case, it is well-known that $\sum_{i=1}^n W_i = O_p(n)$ whereas $\sum_{i=1}^n (W_i - \mu_i) = O_p(\sqrt{n})$, so that the order of magnitude in probability of the uncentered sum $\sum_{i=1}^n W_i$ is much larger than that of the properly centered sum $\sum_{i=1}^n (W_i - \mu_i)$. In other words, the sum of an independent sequence of random variables will concentrate more sharply in a much narrower range around its mean. It follows also that if it had been the case that $E\left[W_i\right] = 0$ for all i; then, we would have $\sum_{i=1}^n W_i = O_p(\sqrt{n})$, so the order of magnitude in this case is smaller than in the case where $E\left[W_i\right] \neq 0$. Moreover, the concentration of measure phenomenon is known to exist more generally, not just for sums of independent random variables but also for Lipschitz functions of such variables and for multilinear forms. See, for example, Tao (2012) for additional discussion.

as

$$\widehat{\delta}_{IJIVE2}-\delta_0=\left(\widetilde{x}'\left[P^{\widetilde{Z}_2}-D\left(P^{\widetilde{Z}_2}\right)\right]\widetilde{x}\right)^{-1}\left(\pi_n'\widetilde{Z}_2'\left[P^{\widetilde{Z}_2}-D\left(P^{\widetilde{Z}_2}\right)\right]\widetilde{\varepsilon}+\widetilde{u}'\left[P^{\widetilde{Z}_2}-D\left(P^{\widetilde{Z}_2}\right)\right]\widetilde{\varepsilon}\right)$$

Again, if we focus on the term $\widetilde{u}'\left[P^{\widetilde{Z}_2}-D\left(P^{\widetilde{Z}_2}\right)\right]\widetilde{\varepsilon}$, we can show by simple manipulation that, since $P^{\widetilde{Z}_2}=\widetilde{Z}_2\left(\widetilde{Z}_2'\widetilde{Z}_2\right)^{-1}\widetilde{Z}_2'=M^{Z_1}Z_2\left(Z_2'M^{Z_1}Z_2\right)Z_2'M^{Z_1}$,

$$\begin{aligned}\widetilde{u}'\left[P^{\widetilde{Z}_2}-D\left(P^{\widetilde{Z}_2}\right)\right]\widetilde{\varepsilon} &= u'M^{Z_1}\left[M^{Z_1}Z_2\left(Z_2'M^{Z_1}Z_2\right)Z_2'M^{Z_1}-D\left(P^{\widetilde{Z}_2}\right)\right]M^{Z_1}\varepsilon \\ &= u'\left[P^{\widetilde{Z}_2}-M^{Z_1}D\left(P^{\widetilde{Z}_2}\right)M^{Z_1}\right]\varepsilon\end{aligned}$$

By an easy calculation, one can show that $E\left\{u'\left[P^{\widetilde{Z}_2}-M^{Z_1}D\left(P^{\widetilde{Z}_2}\right)M^{Z_1}\right]\varepsilon\right\}\neq 0$, so this term is not properly centered at zero, even under the usual assumption that (ε_i,u_i) is independent of (ε_j,u_j) for all $i\neq j$, as long as there is error heteroskedasticity. Note that, the matrix $P^{\widetilde{Z}_2}-M^{Z_1}D\left(P^{\widetilde{Z}_2}\right)M^{Z_1}$ in the middle of the bilinear form in u and ε turns out not to have zero diagonal elements because, in some sense, the process of partialing out Z_1 has interfered with the process of jackknife recentering in this case. Our basic point in presenting this example here is simply to show that it is not as easy as it might seem to construct an IV estimator which simultaneously partial out all additional covariates and at the same time preserve the recentering property of the jackknife methodology. As we will show in the remaining sections of this paper, such estimators can be constructed, however, even in a more general cluster-sample/panel-data setting with fixed effects and with many stochastic instruments and included exogenous regressors.

3 Model, Assumptions, and Estimation Procedures

The more general model that we consider in this paper is a cluster-sample IV regression model

$$\underset{1\times 1}{y_{(i,t)}} = X_{(i,t)}'\delta_0+\varphi_n'Z_{1,(i,t)}+\alpha_i+\varepsilon_{(i,t)}, \tag{4}$$

$$X_{(i,t)} = \Phi_n'Z_{1,(i,t)}+\Pi_n'Z_{2,(i,t)}+\xi_i+U_{(i,t)}, \tag{5}$$

where $i=1,...,n$, $t=1,...,T_i$, and the total sample size is given by $m_n=\sum_{i=1}^n T_i$. The notation $(i,t):\mathbb{N}\times\mathbb{N}\to\mathbb{N}$ denotes a pairing function which maps an ordered pair of natural numbers into a natural number, so that, in particular, we have $(1,1)=1,...,(1,T_1)=T_1,(2,1)=T_1+1,...,(n,T_n)=m_n$. This is just a notational device used to convert a double index into a single index, thus, facilitating certain vectorization and summation operations while still allowing one to keep track of both i and t. In this setup, we take $X_{(i,t)}$ to be a $d \times 1$ vector of endogenous regressors, and we let $Z_{1,(i,t)}$ denote a $K_{1,n} \times 1$ vector of included exogenous variables and let $Z_{2,(i,t)}$ denote a $K_{2,n} \times 1$ vector of instruments, for $i = 1, 2, ..., n$ and $t = 1, ..., T_i$ (or, equivalently, for $(i,t) = 1, ..., m_n$). To allow for the possibility that $Z_{1,(i,t)}$ and $Z_{2,(i,t)}$ may be weakly correlated with the endogenous variables $y_{(i,t)}$ and $X_{(i,t)}$, we let each of the coefficient parameters φ_n, Φ_n, and Π_n to possibly have a (generalized) local-to-zero structure which we will specify more precisely later in Assumptions 3 and 4. In addition, α_i and ξ_i in the above equations denote unobserved or individual effects interpreted as "fixed effects" in the sense that although we do not necessarily require α_i and ξ_i to be (non-random) constants, they are allowed to be correlated with the exogenous variables $Z_{1,(i,t)}$ and $Z_{2,(i,t)}$, unlike the typical assumptions specified in a traditional "random effects" model. More precise assumptions on the model given by equations (4) and (5) are given below.

We will develop some additional notations before proceeding. First, let $Z_1 = \left(Z_{1,(1,1)}, .., Z_{1,(1,T_1)}, ..., Z_{1,(n,1)}, .., Z_{1,(n,T_n)}\right)'$ be an $m_n \times K_{1,n}$ matrix of observations on the include exogenous variables and let $Z_2 = \left(Z_{2,(1,1)}, .., Z_{2,(1,T_1)}, ..., Z_{2,(n,1)}, .., Z_{2,(n,T_n)}\right)'$ be an $m_n \times K_{2,n}$ matrix of observations on the instruments. Also, define the $m_n \times K_n$ matrix $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$, where $K_n = K_{1,n} + K_{2,n}$. Now, let y and X be defined similar to Z_1 and Z_2 by stacking the observations across the index $(i,t) = 1, ..., m_n$; and we can write the model given by equations (4) and (5) more succinctly as

$$\underset{m_n \times 1}{y} = X\delta_0 + Z_1\varphi_n + Q\alpha + \varepsilon, \tag{6}$$

$$\underset{m_n \times d}{X} = Z_1\Phi_n + Z_2\Pi_n + Q\Xi + U, \tag{7}$$

where $\alpha = (\alpha_1, ..., \alpha_n)'$, $\Xi = (\xi_1, ..., \xi_n)'$, and $\underset{m_n \times n}{Q} = \begin{pmatrix} e_{1,n}\iota'_{T_1} & e_{2,n}\iota'_{T_2} & \cdots & e_{n,n}\iota'_{T_n} \end{pmatrix}'$ with $e_{j,n}$ being an $n \times 1$ elementary vector whose j^{th} component is 1 and all other components are 0. Note that our setup allows the clusters to be of possibly different sizes, so that our model can also be interpreted as a possibly unbalanced panel data model. For notational convenience, we have suppressed the dependence of y, X, Z_1, Z_2, Q, ε, and U on n but have made explicit the dependence of φ_n, Φ_n, and Π_n on n to highlight the fact that these parameters may have a local-to-zero structure.

Making use of these notations, we can write down the following assumptions for our model.

Assumption 1: Let $\mathcal{F}_n^Z = \sigma(Z)$ (i.e., the σ-algebra generated by Z). Assume the following conditions are satisfied (i) Conditional on $\mathcal{F}_n^Z$, $\left(\varepsilon_{(1,1)}, U'_{(1,1)}\right), ..., \left(\varepsilon_{(1,T_1)}, U'_{(1,T_1)}\right)$, ..., $\left(\varepsilon_{(n,1)}, U'_{(n,1)}\right), ..., \left(\varepsilon_{(n,T_n)}, U'_{(n,T_n)}\right)$ are mutually independent. (ii) $E\left[\varepsilon_{(i,t)} | \mathcal{F}_n^Z\right] = 0$ and

$E\left[U_{(i,t)}|\mathcal{F}_n^Z\right]=0$ $a.s.$, for $(i,t)=1,...,m_n$.

Assumption 2: There exists a constant $C\geq 1$ such that for all n
(i) $\max_{1\leq (i,t)\leq m_n} E\left[\varepsilon_{(i,t)}^8|\mathcal{F}_n^Z\right]\leq C<\infty$ $a.s.$and $\max_{1\leq (i,t)\leq m_n} E\left[\left\|U_{(i,t)}\right\|_2^8|\mathcal{F}_n^Z\right]\leq C<\infty$ $a.s.$ and (ii) $\inf_{1\leq (i,t)\leq m_n}\lambda_{\min}\left(\Omega_{(i,t)}\right)\geq 1/C>0$ $a.s.$, where $\Omega_{(i,t)}=E\left[\nu_{(i,t)}\nu_{(i,t)}'|\mathcal{F}_n^Z\right]$ with $\nu_{(i,t)}=\begin{pmatrix}\varepsilon_{(i,t)} & U_{(i,t)}'\end{pmatrix}'$.

Assumption 3: Let $\Pi_n=\Upsilon D_\mu/\sqrt{n}$, where $D_\mu=diag\left(\mu_{1,n},..,\mu_{d,n}\right)$. Also, let $\mu_n^{\min}=\min_{1\leq k\leq d}$ $\mu_{k,n}$ and let $K_{2,n}$ denote the number of instruments or the number of columns of Z_2. The following conditions are assumed on the diagonal elements $\mu_{1,n},..,\mu_{d,n}$, as $n\to\infty$. (i) Either $\mu_{k,n}=\sqrt{n}$ or $\mu_{k,n}/\sqrt{n}\to 0$, for $k\in\{1,...,d\}$. (ii) $\mu_n^{\min}\to\infty$, as $n\to\infty$, such that $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2\to 0$ (iii) $\lambda_{\min}\left(H_n\right)\geq 1/C>0$ and $\lambda_{\max}\left(\Upsilon' Z_2' Z_2\Upsilon/n\right)\leq C<\infty$ $a.s.$, for all n sufficiently large, where $H_n=\Upsilon' Z_2' M^{(Z_1,Q)}Z_2\Upsilon/n$. Here, we take $M^{(Z_1,Q)}=M^Q-M^Q Z_1\left(Z_1' M^Q Z_1\right)^{-1}Z_1' M^Q$ where $M^Q=I_{m_n}-Q\left(Q'Q\right)^{-1}Q'$, so that $M^{(Z_1,Q)}$ is a projection matrix which projects into the orthogonal complement of the space spanned by the columns of the matrix $\begin{bmatrix} Z_1 & Q\end{bmatrix}$.

Assumption 4: Let $\Phi_n=\Theta D_\kappa/\sqrt{n}$ and $\varphi_n=\gamma\tau_n/\sqrt{n}$, where $\underset{d\times d}{D_\kappa}=diag\left(\kappa_{1,n},..,\kappa_{d,n}\right)$ and τ_n is a sequence of positive real numbers. The following conditions are assumed on $\kappa_{1,n},..,\kappa_{d,n}$ and on τ_n as $n\to\infty$: (i) either $\kappa_{\ell,n}=\sqrt{n}$ or $\kappa_{\ell,n}/\sqrt{n}\to 0$, for $\ell\in\{1,...,d\}$; (ii) either $\tau_n=\sqrt{n}$ or $\tau_n/\sqrt{n}\to 0$.

Assumption 3 is general enough to accommodate a range of situations including both cases where there are strong instruments and cases where the instruments are weaker. In particular, when $\mu_{1,n}=\cdots=\mu_{d,n}=\mu_n^{\min}=\sqrt{n}$, our model specializes to the more classical situation where the instruments are strong. On the other hand, the cases where some of the $\mu_{j,n}$'s $(j=1,..,d)$ grow at a rate slower than $\sqrt{n}$ correspond to cases where at least some of the components of the parameter vector of interest δ are weakly identified. By allowing for the possibility that different $\mu_{j,n}$'s may grow at different rates, our setup also allows for heterogeneity in how strongly the different components of δ are identified. Note, however, that we do require that $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2\to 0$, since this is both a sufficient and a necessary condition for consistent estimation of δ.[6]

It should be noted that an interesting paper by Antoine and Renault (2012) has also modeled heterogeneity in instrument weakness in a way similar to Assumption 3. However, our setup here differs from that of Antoine and Renault (2012) in several respects. First of all, Antoine and Renault (2012) consider a GMM setup with a fixed number of moment conditions. Hence, Antoine

[6] The sufficiency part of this condition has been demonstrated in various settings by Chao and Swanson (2005), Hausman et al (2012), and Chao et al (2012); whereas the necessity part of this condition has been proved recently by Mikusheva and Sun (2021).

and Renault (2012) allow for nonlinearity in their framework but do not consider the case where the number of instruments/moment conditions may be large, as we do here in our linear setup. In addition, the parameter vector in the Antoine-Renault setup is of fixed dimension. In contrast, although our parameter vector of interest δ is also of fixed dimension; our model contains a large number of additional nuisance and incidental parameters, given that we allow for many included exogenous regressors and for the presence of fixed effects. Thus, the paper by Antoine and Renault (2012) does not consider the kind of problems associated with having to eliminate a large number of nuisance and incidental parameters that we do here in our paper. Given these differences, we view our analysis here as being largely complementary to that of Antoine and Renault (2012).

Assumption 4 allows for possible local-to-zero modeling of the coefficients of Z_1 both in the outcome (or structural) equation and in the first-stage equations. In the special case where $\kappa_{1,n} = \cdots = \kappa_{d,n} = \tau_n = \sqrt{n}$ and $\mu_{1,n} = \cdots = \mu_{d,n} = \mu_n^{\min} = \sqrt{n}$, our model becomes a standard textbook linear IV model (or limited information simultaneous equations model) with strong instruments. However, by allowing for the possibility that some of the $\kappa_{j,n}$'s and/or τ_n may grow at a rate slower than $\sqrt{n}$, we also accommodate situations where the additional covariates may only be weakly correlated with $y_{(i,t)}$ and/or with some elements of $X_{(i,t)}$.

Assumption 5: (i) $m_n \to \infty$ as $n \to \infty$, such that $m_n \sim n$. (ii) $K_{1,n}, K_{2,n} \to \infty$, as $n \to \infty$, such that $K_{1,n}^2/n = O(1)$ and $K_{2,n}^2/n = o(1)$. (iii) Let $M^Q = I_{m_n} - Q(Q'Q)^{-1}Q'$. There exists a positive constant $\underline{C}$ such that $\lambda_{\min}\left(Z'M^Q Z\right) \geq \underline{C} > 0$ $a.s.$, for all n sufficiently large. (iv) Let $P^{\perp} = P^{(Z,Q)} - P^{(Z_1,Q)} = M^{(Z_1,Q)}Z_2\left(Z_2'M^{(Z_1,Q)}Z_2\right)^{-1}Z_2'M^{(Z_1,Q)}$ and $P^{Z_1^{\perp}} = M^Q Z_1\left(Z_1'M^Q Z_1\right)^{-1}Z_1'M^Q$, where $M^{(Z_1,Q)}$ is as defined in part (iii) of Assumption 3 and where $P^{(Z,Q)}$ and $P^{(Z_1,Q)}$ are projection matrices that project into the column space of $\begin{bmatrix} Z & Q \end{bmatrix}$ and $\begin{bmatrix} Z_1 & Q \end{bmatrix}$, respectively[7]. Assume that $\max_{1\leq(i,t)\leq m_n} P^{Z_1^{\perp}}_{(i,t),(i,t)} = O_{a.s.}(K_{1,n}/n)$ and $\max_{1\leq(i,t)\leq m_n} P^{\perp}_{(i,t),(i,t)} = O_{a.s.}(K_{2,n}/n)$[8].

Assumption 6: (i) $\min_{1\leq i\leq n} T_i \geq 3$ for all n; (ii) There exists a positive integer $\overline{T} \geq 3$, such that $\max_{1\leq i\leq n} T_i \leq \overline{T} < \infty$, for all n.

Assumption 7: Assume that $\max_{1\leq(i,t)\leq m_n} \left\|\Upsilon' Z_2' M^{(Z_1,Q)} e_{(i,t)}\right\|_2/\sqrt{n} = o_p(1)$, where $e_{(i,t)}$ is an $m_n \times 1$ elementary vector whose $(i,t)^{th}$ component is 1 and all components are 0 for $(i,t) \in \{1, 2, ..., m_n\}$.

[7] Note that $P^{(Z,Q)}$ and $P^{(Z_1,Q)}$can be given the explicit representations $P^{(Z,Q)} = P^Z + M^Z Q\left(Q'M^Z Q\right)^{-1}Q'M^Z$ and $P^{(Z_1,Q)} = P^{Z_1} + M^{Z_1}Q\left(Q'M^{Z_1}Q\right)^{-1}Q'M^{Z_1}$, where $P^Z = Z(Z'Z)^{-1}Z'$, $P^{Z_1} = Z_1(Z_1'Z_1)^{-1}Z_1'$, $M^{Z_1} = I_{m_n} - P^{Z_1}$, and $M^Z = I_{m_n} - P^Z$.

[8] More primitive, sufficient conditions for $\max_{1\leq(i,t)\leq m_n} P^{Z_1^{\perp}}_{(i,t),(i,t)} = O_{a.s.}(K_{1,n}/n)$ and $\max_{1\leq(i,t)\leq m_n} P^{\perp}_{(i,t),(i,t)} = O_{a.s.}(K_{2,n}/n)$ are given in Lemma OA-20 of the Additional Online Appendix, which can be found at the URL: http://econweb.umd.edu/~chao/Research/research_files/Additional_Online_Appendix_Jackknife_Estimation_Cluster_Sample_IV_Model_December_20_2022.pdf

Note that Assumption 7 is similar to a condition given in Assumption 3 of Cattaneo, Jansson, and Newey (2018). As noted in that paper, this assumption comes close to providing a minimal condition for the central limit theorem to hold.

Assumption 8: Let $\rho_n = E\left[U' M^Q \varepsilon\right] / E\left[\varepsilon' M^Q \varepsilon\right]$. Let the limit of ρ_n exists, so that $\rho_n \to \rho$, as $n \to \infty$, for some fixed $d \times 1$ vector $\rho \in \mathcal{S}_\rho$, where $\mathcal{S}_\rho$ denotes some compact subset of $\mathbb{R}^d$.

To estimate the parameter (vector) of interest δ in equation (4), we propose three new jackknife-type IV estimators. We shall use the acronyms FEJIV, FELIM, and FEFUL to denote, respectively, the Fixed Effect Jackknife IV, the Fixed Effect LIML, and the Fixed Effect Fuller estimator.

1. **FEJIV:**

$$\widehat{\delta}_J = \left(X' A X\right)^{-1} X' A y,$$

   where $A = P^{\perp} - M^{(Z,Q)} D_{\widehat{\vartheta}} M^{(Z,Q)}$, with $M^{(Z,Q)} = I_{m_n} - P^{(Z,Q)}$ and with $P^{\perp}$ as previously defined in Assumption 5. In addition, $D_{\widehat{\vartheta}}$ denotes an $m_n \times m_n$ diagonal matrix, whose diagonal elements $\widehat{\vartheta} = \begin{pmatrix} \widehat{\vartheta}_1 & \widehat{\vartheta}_2 & \cdots & \widehat{\vartheta}_{m_n} \end{pmatrix}'$, when stacked into a vector, correspond to the solution of the system of linear equations $d_{P^{\perp}} = \left(M^{(Z,Q)} \circ M^{(Z,Q)}\right) \vartheta$, where $d_{P^{\perp}}$ is an $m_n \times 1$ vector containing the diagonal elements of the projection matrix $P^{\perp}$.[9]

2. **FELIM:** The FELIM estimator $\widehat{\delta}_L$ is the estimator that minimizes the objective function

$$\widehat{Q}_{FELIM}(\delta) = \frac{(y - X\delta)' A (y - X\delta)}{(y - X\delta)' M^{(Z_1,Q)} (y - X\delta)}, \tag{8}$$

   where A is as defined above in the definition of FEJIV and where $M^{(Z_1,Q)}$ is as defined in Assumption 3. $\widehat{\delta}_L$ has the explicit representation

$$\widehat{\delta}_L = \left(X' \left[A - \widehat{\ell}_L M^{(Z_1,Q)}\right] X\right)^{-1} \left(X' \left[A - \widehat{\ell}_L M^{(Z_1,Q)}\right] y\right), \tag{9}$$

   where $\widehat{\ell}_L$ is the smallest root of the determinantal equation $\det\left\{\overline{X}' A \overline{X} - \ell \overline{X}' M^{(Z_1,Q)} \overline{X}\right\}$ $= 0$ with $\overline{X} = \begin{bmatrix} y & X \end{bmatrix}$.

3. **FEFUL:** The FEFUL estimator $\widehat{\delta}_F$ is defined as follows:

$$\widehat{\delta}_F = \left(X' \left[A - \widehat{\ell}_F M^{(Z_1,Q)}\right] X\right)^{-1} \left(X' \left[A - \widehat{\ell}_F M^{(Z_1,Q)}\right] y\right),$$

[9] In Lemma 1 below, we show that, under mild conditions, the system of linear equations, $d_{P^{\perp}} = \left(M^{(Z,Q)} \circ M^{(Z,Q)}\right) \vartheta$, always has a unique solution.

where $\widehat{\ell}_F = \left[\widehat{\ell}_L - \left(1 - \widehat{\ell}_L\right) C/m_n\right] / \left[1 - \left(1 - \widehat{\ell}_L\right) C/m_n\right]$ for some constant C and where$\widehat{\ell}_L$ is as previously defined in the definition of FELIM given above. For the Monte Carlo results reported in section 6, we shall take $C = 1$.

To help develop some intuition for these new estimators, it is easiest if we focus the discussion on FEJIV. To proceed, note first that, under our setup, it is not difficult to show that

$$\widehat{\delta}_J - \delta_0 = \left(X'AX\right)^{-1} X'A\varepsilon = \left(X'AX\right)^{-1} \left(\Pi_n' Z_2' A\varepsilon + U'A\varepsilon\right),$$

where the "numerator" of the right-hand side of this equation is again written in a familiar form as the sum of a linear form $\Pi_n' Z_2' A\varepsilon$ plus a bilinear form $U'A\varepsilon$. Next, note that an elementary result from linear algebra states that if $A = MDM$, where A is a square matrix, D is a diagonal matrix, and M is a symmetric matrix, then $a = (M \circ M)\, d$, where $a = (a_{11}, a_{22}, ..., a_{m_n,m_n})'$ and $d = (d_{11}, d_{22}, ..., d_{m_n,m_n})'$ are vectors whose elements are the diagonal elements of the matrices A and D, respectively. Put in words, this result states that the vector of diagonal elements of A is a linear transformation of the vector of diagonal elements of D, with the transformation matrix given by $(M \circ M)$. Since in the definition of $\widehat{\delta}_J$, we have specified $A = P^{\perp} - M^{(Z,Q)} D_{\widehat{\vartheta}} M^{(Z,Q)}$, it follows that by choosing the diagonal elements of $D_{\widehat{\vartheta}}$ to satisfy the system of linear equations $d_{P^{\perp}} = \left(M^{(Z,Q)} \circ M^{(Z,Q)}\right) \vartheta$, where $d_{P^{\perp}} = \left(P_{11}^{\perp}, P_{22}^{\perp}, ..., P_{m_n,m_n}^{\perp}\right)'$, we would, by construction, end up with a matrix A whose diagonal elements $A_{11}, ..., A_{m_n,m_n}$ are all zero. This, in turn, leads to the bilinear form $U'A\varepsilon$ having the characteristics of a degenerate U-statistic, with expectation that is properly centered at zero. As discussed in the previous section, this proper centering is important, as it reduces the order of magnitude of the bilinear term $U'A\varepsilon$ and, thus, allows $\widehat{\delta}_J$ to be both consistent and asymptotically normal under many weak instrument asymptotics so long as $\sqrt{K_{2,n}} / \left(\mu_n^{\min}\right)^2 \to 0$. In addition, write $\widehat{\delta}_J - \delta_0 = \left(X'AX\right)^{-1} X'A\left(Z_1\varphi_n + Q\alpha + \varepsilon\right)$, and note that

$$\begin{aligned} X'A\left(Z_1\varphi_n + Q\alpha + \varepsilon\right) &= \left(Z_1\Phi_n + Z_2\Pi_n + Q\Xi + U\right)' \left[P^{\perp} - M^{(Z,Q)} D_{\widehat{\vartheta}} M^{(Z,Q)}\right] \left(Z_1\varphi_n + Q\alpha + \varepsilon\right) \\ &= \Pi_n' Z_2' P^{\perp}\varepsilon + U' \left[P^{\perp} - M^{(Z,Q)} D_{\widehat{\vartheta}} M^{(Z,Q)}\right] \varepsilon. \end{aligned} \tag{10}$$

Looking at equation (10), we see that the design of the matrix A allows fixed effects and the included exogenous regressors Z_1 to be partialed out on both sides of A in the above expression, and this is done in such a way so that the proper centering of the bilinear form $U' \left[P^{\perp} - M^{(Z,Q)} D_{\widehat{\vartheta}} M^{(Z,Q)}\right] \varepsilon$ is still preserved. FELIM and FEFUL are a bit more complicated than FEJIV to discuss, but they share the same basic design as FEJIV; and, in consequence, they will also be consistent and asymptotically normal under many weak instrument asymptotics, as we will show in the theorems below.

In contrast, jackknife IV estimators currently available in the literature do not fully accomplish the dual goals of being both properly centered and of having all cluster-specific effects and additional covariates properly partialed out. To be more specific, we will briefly discuss a number of jackknife IV estimators that have been proposed in the literature. The paper by Angrist, Imbens, and Krueger (1999) consider the JIVE1 and JIVE2 estimators of the parameter vector δ, but in a cross-sectional setup without either fixed effects or included exogenous regressors. Hence, these authors do not explicitly study the more general version of these estimators that partials out additional covariates. Hausman et al. (2012) introduce jackknife versions of LIML and Fuller estimators called HLIM and HFUL, but they do so in a cross-sectional context where there are no fixed effects and where only a small number of included exogenous regressors is allowed, so that the problem of having to partial out fixed effects and a potentially large number of included exogenous variables is not studied in that paper. In addition, the symmetric jackknife IV (SJIVE) estimator proposed by Bekker and Crudu (2015) is formulated in a setting without fixed effects and with no included exogenous regressors. Hence, that paper also does not consider issues related to having to partial out additional covariates.

In a recent paper, Evdokimov and Kolesár (2018) examine a number of interesting jackknife IV estimators that allow for partialing out of additional covariates. In the previous section, we have already discussed the IJIVE2 estimator from that paper in the context of a simple cross-sectional IV model. Here, we shall briefly examine the other estimators considered in Evdokimov and Kolesár (2018) and provide some discussion about how these estimators might perform under many weak instruments asymptotics when applied to our more general cluster-sample setting here with fixed effects. For this purpose, it is easiest to consider the case where there is only one endogenous regressor. In this case, note that $D_\mu = \mu_n = \mu_n^{\min}$ since $d = 1$, and we shall use x, π_n, ϕ_n, υ, and u in lieu of X, Π_n, Φ_n, Υ, and U to emphasize the fact that, in the one endogenous regressor case; x, π_n, ϕ_n, and u are vectors and not matrices.

Consider first the IJIVE1 estimator studied in that paper. This estimator was originally proposed by Ackerberg and Devereux (2009) and is further analyzed in the grouped data setting by

Evdokimov and Kolesár (2018)[10]. Using our notation, the estimator can be written in the form

$$\begin{aligned}\widehat{\delta}_{IJIVE1} &= \left(x'M^{(Z_1,Q)}\left[P^{\perp}-D\left(P^{\perp}\right)\right]\left[I_{m_n}-D\left(P^{\perp}\right)\right]^{-1}M^{(Z_1,Q)}x\right)^{-1} \\ &\quad\times\left(x'M^{(Z_1,Q)}\left[P^{\perp}-D\left(P^{\perp}\right)\right]\left[I_{m_n}-D\left(P^{\perp}\right)\right]^{-1}M^{(Z_1,Q)}y\right).\end{aligned}$$

Now, it is easily seen that the deviation of this estimator from the true value δ_0 can be written as

$$\begin{aligned}\widehat{\delta}_{IJIVE1}-\delta_0 &= \left(x'A_{IJ1}x\right)^{-1}x'A_{IJ1}\left(Z_1\varphi_n+Q\alpha+\varepsilon\right) \\ &= \left(x'A_{IJ1}x\right)^{-1}\left(\Pi_n'Z_2'A_{IJ1}\varepsilon+U'A_{IJ1}\varepsilon\right), \qquad (11)\end{aligned}$$

where $A_{IJ1}=M^{(Z_1,Q)}\left[P^{\perp}-D\left(P^{\perp}\right)\right]\left[I_{m_n}-D\left(P^{\perp}\right)\right]^{-1}M^{(Z_1,Q)}$. Straightforward calculations further show that the $(i,t)^{th}$ diagonal element of the matrix A_{IJ1} is given by

$$A_{IJ1,(i,t),(i,t)}=\sum_{(j,s)=1}^{m_n}\frac{M^{(Z_1,Q)}_{(j,s),(i,t)}}{1-P^{\perp}_{(j,s),(j,s)}}\left[P^{\perp}_{(i,t),(j,s)}-M^{(Z_1,Q)}_{(i,t),(j,s)}P^{\perp}_{(j,s),(j,s)}\right]\neq 0,$$

for $(i,t)=1,...,m_n$, so that $u'A_{IJ1}\varepsilon$, the bilinear form on the right-hand side of equation (11) above, will not be a degenerate U-statistic and will not be properly centered at the origin. Hence, similar to what we have pointed out previously about IJIVE2, the problem here is that, although the matrix $\left[P^{\perp}-D\left(P^{\perp}\right)\right]\left[I_{m_n}-D\left(P^{\perp}\right)\right]^{-1}$ does have a "jackknife form" in the sense that the elements of its main diagonal are all zero, it defines a bilinear form not with respect to u and ε but with respect to the projected vectors $\widehat{u}=M^{(Z_1,Q)}u$ and $\widehat{\varepsilon}=M^{(Z_1,Q)}\varepsilon$. Note, however, that in general the $(i,t)^{th}$ element of $\widehat{u}$ will contain not just the $(i,t)^{th}$ element of u but other elements as well, and similarly for $\widehat{\varepsilon}$. In consequence, merely having the diagonal elements zeroed out in this case is not sufficient for the bilinear form $u'A_{IJ1}\varepsilon=\widehat{u}'\left[P^{\perp}-D\left(P^{\perp}\right)\right]\left[I_{m_n}-D\left(P^{\perp}\right)\right]^{-1}\widehat{\varepsilon}$ to have expectation equal to zero. Again, we have a situation where the process of partialing out the covariates has interfered with the process of jackknife recentering.

Another estimator studied in Evdokimov and Kolesár (2018) is the UJIVE estimator, which was first introduced in Kolesár (2013) and then further analyzed in the grouped data setting by

[10] It should be noted that this estimator was originally referred to in Ackerberg and Devereux (2009) as simply IJIVE. However, since Edokimov and Kolesár (2018) introduced a variant of this estimator in their paper which they called IJIVE2, they renamed the original IJIVE estimator IJIVE1.

Evdokimov and Kolesár (2018). This estimator takes the form

$$\begin{aligned}\widehat{\delta}_{UJIVE} &= \left(x'\left[\widetilde{P}^{(Z,Q)}D\left(M^{(Z,Q)}\right)^{-1} - \widetilde{P}^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}\right]x\right)^{-1} \\ &\quad \times\left(x'\left[\widetilde{P}^{(Z,Q)}D\left(M^{(Z,Q)}\right)^{-1} - \widetilde{P}^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}\right]y\right),\end{aligned}$$

where $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$, $\widetilde{P}^{(Z,Q)} = P^{(Z,Q)} - D\left(P^{(Z,Q)}\right)$, and $\widetilde{P}^{(Z_1,Q)} = P^{(Z_1,Q)} - D\left(P^{(Z_1,Q)}\right)$. Now, the deviation of the UJIVE estimator from the true value δ_0 can be written as

$$\begin{aligned}\widehat{\delta}_{UJIVE} - \delta_0 &= \left(\frac{x'A_{UJ}x}{\mu_n^2}\right)^{-1}\left(\frac{x'A_{UJ}Z_1\varphi_n + x'A_{UJ}Q\alpha + \phi_n'Z_1'A_{UJ}\varepsilon + \pi_n'Z_2'A_{UJ}\varepsilon + u'A_{UJ}\varepsilon}{\mu_n^2}\right) \\ &= \left(\frac{x'A_{UJ}x}{\mu_n^2}\right)^{-1}\left(\frac{x'A_{UJ}Z_1\varphi_n + x'A_{UJ}Q\alpha + \pi_n'Z_2'A_{UJ}\varepsilon + u'A_{UJ}\varepsilon}{\mu_n^2}\right)\end{aligned}$$

where $A_{UJ} = \left[P^{(Z,Q)} - D\left(P^{(Z,Q)}\right)\right]D\left(M^{(Z,Q)}\right)^{-1} - \left[P^{(Z_1,Q)} - D\left(P^{(Z_1,Q)}\right)\right]D\left(M^{(Z_1,Q)}\right)^{-1}$. Note first that the diagonal elements of the matrix A_{UJ} are all equal to zero, so the bilinear term for this estimator, $U'A_{UJ}\varepsilon$, is properly centered. However, this estimator has a bias problem that arises from the presence of the term $x'A_{UJ}Z_1\varphi_n/\mu_n^2$, which can be nonnegligible and even large in order of magnitude. To see this, observe first that simple manipulation shows that $A_{UJ} = M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1} - M^{(Z,Q)}D\left(M^{(Z,Q)}\right)^{-1}$. Using this identity, we can write

$$\begin{aligned}\frac{x'A_{UJ}Z_1\varphi_n}{\mu_n^2} &= \frac{\pi_n'Z_2'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}Z_1\varphi_n}{\mu_n^2} \\ &\quad + \frac{u'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}Z_1\varphi_n}{\mu_n^2} - \frac{u'M^{(Z,Q)}D\left(M^{(Z,Q)}\right)^{-1}Z_1\varphi_n}{\mu_n^2}. \qquad (12)\end{aligned}$$

Note that the term on the right-hand side of (12) which can be particularly large in order of magnitude is $\pi_n'Z_2'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}Z_1\varphi_n/\mu_n^2$. In fact, one can show that

$$\frac{\pi_n'Z_2'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}Z_1\varphi_n}{\mu_n^2} = \frac{\tau_n}{\mu_n}\frac{\upsilon'Z_2'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}Z_1\gamma}{n} = O_{a.s.}\left(\frac{\tau_n}{\mu_n}\right).$$

Hence, this estimator will be inconsistent as long as $\mu_n = O(\tau_n)$. This will certainly be true in weak instrument cases where $\mu_n = o(\tau_n)$, but can also occur even in strong instrument cases where $\mu_n \sim \sqrt{n}$ if the included exogenous regressors enter significantly into the structural equation of interest, in which case $\tau_n \sim \sqrt{n}$. Our Monte Carlo results reported in section 6 also confirm that UJIVE can have a large median bias relative to its competitors when there are included exogenous regressors that enter significantly into the structural equation of interest[11].

Since our setup essentially has a panel data structure, one may also wonder if it is possible to simply first difference away the fixed effects and then do a jackknife-type recentering. A problem with this strategy occurs if the IV regression contains, in addition to fixed effects, other included exogenous regressors which cannot be eliminated by first-differencing. In that case, one will have to do a projection to partial out these included exogenous regressors, leading to the same problem as we have discussed previously with regard to IJIVE1 and IJIVE2. In fact, the problem will be worse in this case due to the serial correlation in the errors induced by the first-differencing. Moreover, even if there are no additional included exogenous regressors, the serial correlation induced by first differencing causes additional complications. In particular, let $P^Z = Z\left(Z'Z\right)^{-1}Z'$ denote the projection matrix of the instruments[12]. Then, to achieve proper jackknife recentering in this case requires the removal not only of the elements on the main diagonal of P^Z but also the elements on the superdiagonal and the subdiagonal of P^Z, so that with serial correlation proper recentering is attained only at the cost of greater information loss. Finally, the presence of serial correlation also makes the large sample covariance matrix of a jackknife IV estimator under many weak instrument asymptotics both more complicated and more difficult to estimate. Hence, we believe that our approach for removing fixed or cluster-specific effects has certain advantages over any alternative procedure that is based on first-differencing. It should be noted that a recent panel data paper by Hsiao and Zhou (2018) does take the approach of constructing a jackknife IV estimator after first-differencing the data. However, the objective and focus of that paper differs greatly from ours. First of all, the panel data simultaneous equations model specified in Hsiao and Zhou (2018) does not allow for the degree of instrument weakness that we consider. In addition, the model that they consider does not have error heteroskedasticity or included exogenous regressors. If we apply their estimator to our setting, the estimator will not be consistent in the case where $K_{2,n} \sim \left(\mu_n^{\min}\right)^2$ or in the case where $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$, but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. Still, it should be stressed that in their setting with strong instruments and error homoskedasticity their estimator does have good

[11] It should be noted, however, that UJIVE may perform well under many weak instrument asymptotics in the special case where the equation of interest contains no included exogenous regressors and only fixed effects. This is not only because in this case there is no term of the form $x'A_{UJ}Z_1\varphi_n/\mu_n^2 = \tau_n x'A_{UJ}Z_1\gamma/\left(\mu_n^2\sqrt{n}\right)$, but also because, in this case,

$$\frac{\pi_n' Z_2' A_{UJ} Q\alpha}{\mu_n^2} = \frac{\pi_n' Z_2' \left[M^Q D\left(M^Q\right)^{-1} - M^{(Z_2,Q)} D\left(M^{(Z_2,Q)}\right)^{-1}\right] Q\alpha}{\mu_n\sqrt{n}} = 0$$

so that, without the contaminating effects of the included exogenous regressors, UJIVE does properly partial out the fixed effects. We conjecture that, in this setting, UJIVE might be consistent so long as $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$, but we have yet to obtain a formal proof of this result.

[12] Here, we let Z denote the matrix of observations on the instruments because we are referring to a case where there are no included exogenous variables, Z_1.

asymptotic properties.

Turning our attention back to the equation $d_{P^{\perp}} = \left(M^{(Z,Q)} \circ M^{(Z,Q)}\right)\vartheta$, note that in order for this system of linear equations to have a unique solution, we need the matrix $\left(M^{(Z,Q)} \circ M^{(Z,Q)}\right)$ to be invertible. The following lemma provides sufficient conditions for the invertibility of $\left(M^{(Z,Q)} \circ M^{(Z,Q)}\right)$.

Lemma 1: Suppose that Assumptions 5 and 6(i) are satisfied. Then, there exists a positive constant C such that $\lambda_{\min}\left(M^{(Z,Q)} \circ M^{(Z,Q)}\right) \geq C > 0$ $a.s.$, for all n sufficiently large[13].

It should be noted that a more general result on conditions for the invertibility of Hadamard products has been given previously in Cattaneo, Jansson, and Newey (2018)[14]. However, we choose to present a specialization of their result because it shows that, in the context of our cluster-sampling setup, a key condition for ensuring the invertibility of $\left(M^{(Z,Q)} \circ M^{(Z,Q)}\right)$ is $\min_{1\leq i\leq n} T_i \geq 3$, which we explicitly assume in Assumption 6 part (i) above.

A further observation is that, in analyzing estimators that are obtained from minimizing a variance ratio (e.g., FELIM), it is often convenient to first consider the objective function in the form $Q(\beta) = \left(\beta'\overline{X}'A\overline{X}\beta\right)/\left(\beta'\overline{X}'M^{(Z_1,Q)}\overline{X}\beta\right)$, where $\overline{X} = [y, X]$ and where β is a $(d+1)\times 1$ vector, not initially normalized to identify the dependent variable from the regressors. In this setting, one would first minimize the objective function $Q(\beta)$ to obtain a minimizer $\widetilde{\beta} = \begin{pmatrix} \widetilde{\beta}_1 & \widetilde{\beta}_2' \end{pmatrix}'$, with $\widetilde{\beta}_1$ being a scalar and $\widetilde{\beta}_2$ a $d\times 1$ vector and subsequently normalize the last d components of $\widetilde{\beta}$ to obtain an estimator $\widetilde{\delta} = -\widetilde{\beta}_2/\widetilde{\beta}_1$ for the coefficients of the endogenous regressors X. The following assumption ensures that this subsequent normalization is well-defined. Moreover, in the proof of Lemma S2-11 given in the Additional Online Appendix to this paper, we show that, by following this procedure, we end up with exactly the FELIM estimator $\widehat{\delta}_L$, that satisfies the first-order conditions of the objective function given by (8) and that also has explicit representation given by equation (9) above[15].

Assumption 9: Consider the variance-ratio objective function $Q(\beta) = \left(\beta'\overline{X}'A\overline{X}\beta\right)/\left(\beta'\overline{X}'M^{(Z_1,Q)}\overline{X}\beta\right)$, where $\beta \in \overline{B} = \left\{\beta \in \mathbb{R}^{d+1} : \|\beta\|_2 = 1\right\}$. Let $\widetilde{\beta}$ be a $(d+1)\times 1$ vector that minimizes the objective function $Q(\beta)$, among all $\beta \in \overline{B}$ (i.e., $\widetilde{\beta} = \arg\min_{\beta\in\overline{B}} Q(\beta)$). Partition $\widetilde{\beta} = \left(\widetilde{\beta}_1, \widetilde{\beta}_2'\right)'$ as defined above and assume that there exists a positive

[13] A proof of Lemma 1 is given in section 2 of the Additional Online Appendix for this paper. This online appendix can be viewed at the URL:
http://econweb.umd.edu/~chao/Research/research_files/Additional_Online_Appendix_Jackknife_Estimation_Cluster_Sample_IV_Model_December_20_2022.pdf

[14] See, in particular, the analysis given in Section 3 of their Supplemental Appendix.

[15] The proof of Lemma S2-11 is given in section 1 of the Additional Online Appendix, which, in turn, can be found at the URL:
http://econweb.umd.edu/~chao/Research/research_files/Additional_Online_Appendix_Jackknife_Estimation_Cluster_Sample_IV_Model_December_20_2022.pdf

constant $\underline{C}$ such that

$$\left|\widetilde{\beta}_1\right| \geq \underline{C} > 0 \;\; a.s. \text{ for all } n \text{ sufficiently large.} \tag{13}$$

Note that constraining β (so that $\|\beta\|_2 = 1$) is not restrictive since we are dealing with an objective function $Q(\beta)$ that is a ratio of quadratic forms in β. More precisely, let $\overline{\beta} = \arg\min_{\beta \in \mathbb{R}^{d+1}} Q(\beta)$, where $\overline{\beta} \neq 0$, and let $\widetilde{\beta} = \overline{\beta}/\|\overline{\beta}\|_2$ so that $\left\|\widetilde{\beta}\right\|_2 = 1$. Then, $Q\left(\overline{\beta}\right) = \left(\overline{\beta}'\overline{X}'A\overline{X}\overline{\beta}\right) / \left(\overline{\beta}'\overline{X}'M^{(Z_1,Q)}\overline{X}\overline{\beta}\right)$ $= \left(\|\overline{\beta}\|_2^{-1}\overline{\beta}'\overline{X}'A\overline{X}\overline{\beta}\,\|\overline{\beta}\|_2^{-1}\right) / \left(\|\overline{\beta}\|_2^{-1}\overline{\beta}'\overline{X}'M^{(Z_1,Q)}\overline{X}\overline{\beta}\,\|\overline{\beta}\|_2^{-1}\right) = Q\left(\widetilde{\beta}\right)$, so any minimal value of $Q(\beta)$ obtained by minimizing β over all $\beta \in \mathbb{R}^{d+1}$ can also be achieved by some $\widetilde{\beta}$ such that $\left\|\widetilde{\beta}\right\|_2 = 1$.

4 Consistency and Asymptotic Normality of Point Estimators

Theorem 1: Let $\overline{\delta}_n = \left(X'\left[A - \overline{\ell}_n M^{(Z_1,Q)}\right]X\right)^{-1}\left(X'\left[A - \overline{\ell}_n M^{(Z_1,Q)}\right]y\right)$, for some sequence $\overline{\ell}_n$, such that $\overline{\ell}_n = o_p\left(\left[\mu_n^{\min}\right]^2/n\right) = o_p(1)$. Then, under Assumptions 1-6, $\left\|D_\mu\left(\overline{\delta}_n - \delta_0\right)/\mu_n^{\min}\right\|_2 \xrightarrow{p} 0$ and $\left\|\overline{\delta}_n - \delta_0\right\|_2 \xrightarrow{p} 0$, as $n \to \infty$

Special cases of the class of estimators that satisfy the conditions of Theorem 1, and are thus consistent in the sense described in the theorem, include FEJIV $\widehat{\delta}_{J,n}$, FELIM $\widehat{\delta}_{L,n}$, and FEFUL $\widehat{\delta}_{F,n}$. Evidently, the main difference between these estimators is the different specifications of $\overline{\ell}_n$. $\widehat{\delta}_{J,n}$ takes $\overline{\ell}_n = 0$, for all n; $\widehat{\delta}_{L,n}$ takes $\overline{\ell}_n = \widehat{\ell}_{L,n}$, where $\widehat{\ell}_{L,n}$ is the smallest root of the determinantal equation $\det\left\{\overline{X}'A\overline{X} - \ell\overline{X}'M^{(Z_1,Q)}\overline{X}\right\} = 0$; and $\widehat{\delta}_{F,n}$ takes $\overline{\ell}_n = \widehat{\ell}_{F,n}$ $= \left[\widehat{\ell}_L - \left(1 - \widehat{\ell}_L\right)C/m_n\right] / \left[1 - \left(1 - \widehat{\ell}_L\right)C/m_n\right]$, as described earlier. Hence, by verifying that, in all three cases, $\overline{\ell}_n$ satisfies the condition $\overline{\ell}_n = o_p\left(\left[\mu_n^{\min}\right]^2/n\right) = o_p(1)$, we can easily specialize the consistency result of Theorem 1 to establish the consistency of FEJIV, FELIM, and FEFUL. These results are given in the following corollary.

Corollary 1: Under Assumptions 1-6 and 9, the following results hold as $n \to \infty$.
(a) $\left\|D_\mu\left(\widehat{\delta}_{J,n} - \delta_0\right)/\mu_n^{\min}\right\|_2 \xrightarrow{p} 0$ and $\left\|\widehat{\delta}_{J,n} - \delta_0\right\|_2 \xrightarrow{p} 0$. (b) $\left\|D_\mu\left(\widehat{\delta}_{L,n} - \delta_0\right)/\mu_n^{\min}\right\|_2 \xrightarrow{p} 0$ and $\left\|\widehat{\delta}_{L,n} - \delta_0\right\|_2 \xrightarrow{p} 0$. (c) $\left\|D_\mu\left(\widehat{\delta}_{F,n} - \delta_0\right)/\mu_n^{\min}\right\|_2 \xrightarrow{p} 0$ and $\left\|\widehat{\delta}_{F,n} - \delta_0\right\|_2 \xrightarrow{p} 0$.

The next two results establish asymptotic normality for the FELIM and FEFUL estimators, under two different cases: (i) Case I: $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$ and (ii) Case II: $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$, but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. The FEJIV estimator can also be shown to have an asymptotic normal distribution under both Cases I and II. However, we choose to focus our theoretical analysis on

FELIM and FEFUL because, as noted previously, the results of our Monte Carlo study indicate that FELIM and FEFUL have better finite sample properties than FEJIV.

To facilitate the statement of the next two results, define

$$\Lambda_{I,n} = H_n^{-1}\left(\Sigma_{1,n}+\Sigma_{2,n}\right)H_n^{-1} = H_n^{-1}\Sigma_n H_n^{-1}, \tag{14}$$

$$\Lambda_{II,n} = \frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} H_n^{-1}\Sigma_{2,n}H_n^{-1}, \tag{15}$$

where $H_n = \Upsilon' Z_2' M^{(Z_1,Q)} Z_2 \Upsilon / n$, $\Sigma_{1,n} = VC\left(\Upsilon' Z_2' M^{(Z_1,Q)} \varepsilon/\sqrt{n} | \mathcal{F}_n^Z\right)$ $= \Upsilon' Z_2' M^{(Z_1,Q)} D_{\sigma^2} M^{(Z_1,Q)} Z_2 \Upsilon / n$, and $\Sigma_{2,n} = D_\mu^{-1}\Sigma_{2,n}^* D_\mu^{-1}$, with

$$\begin{aligned}
\Sigma_{2,n}^* &= VC\left(\underline{U}' A\varepsilon | \mathcal{F}_n^Z\right) \\
&= \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2 E\left[\varepsilon_{(i,t)}^2 | \mathcal{F}_n^Z\right] E\left[\underline{U}_{(j,s)}\underline{U}_{(j,s)}' | \mathcal{F}_n^Z\right] \\
&\quad + \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2 E\left[\underline{U}_{(i,t)}\varepsilon_{(i,t)} | \mathcal{F}_n^Z\right] E\left[\varepsilon_{(j,s)}\underline{U}_{(j,s)}' | \mathcal{F}_n^Z\right].
\end{aligned} \tag{16}$$

In addition, $\Sigma_n = \Sigma_{1,n} + \Sigma_{2,n}$ and $\underline{U}_{(i,t)} = U_{(i,t)} - \rho\varepsilon_{(i,t)}$ for $(i,t) = 1,...,m_n$. Here, for any random vector x, $VC\left(x|\mathcal{F}_n^Z\right)$ denotes the conditional variance-covariance matrix of x given $\mathcal{F}_n^Z$. Moreover, let $D_{\sigma^2} = diag\left(\sigma_{(1,1)}^2, ..., \sigma_{(n,T_n)}^2\right) = diag\left(\sigma_1^2, ..., \sigma_{m_n}^2\right)$, where $\sigma_{(i,t)}^2 = \left[\varepsilon_{(i,t)}^2 | \mathcal{F}_n^Z\right]$, for $(i,t) = 1,...,m_n$ and where, for notational convenience, we suppress the dependence of $\sigma_{(i,t)}^2$ on $\mathcal{F}_n^Z$.

As evident from the results given below, $\Lambda_{I,n}$ and $\Lambda_{II,n}$ are the (conditional) covariance matrices of FELIM (and also of FEFUL) in large samples under Cases I and II, respectively.

Theorem 2: Let Assumptions 1-9 be satisfied. Then, under Case I where $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$, the following results hold: $\Lambda_{I,n}$ is positive definite $a.s.$ for all n sufficiently large; and, as $n \to \infty$, $\Lambda_{I,n}^{-1/2} D_\mu\left(\widehat{\delta}_{L,n} - \delta_0\right) \xrightarrow{d} N(0, I_d)$ and $\Lambda_{I,n}^{-1/2} D_\mu\left(\widehat{\delta}_{F,n} - \delta_0\right) \xrightarrow{d} N(0, I_d)$.

Theorem 3: Let Assumptions 1-9 be satisfied, let $\widetilde{L}_n$ be a $q \times d$ matrix with $1 \le q \le d$, and let there exists a positive constant C such that $\left\|\widetilde{L}_n\right\|_2 \le C < \infty$ and $\lambda_{\min}\left(\widetilde{L}_n \Lambda_{II,n} \widetilde{L}_n'\right) \ge 1/C > 0$ $a.s.n.$ Then, under Case II where $\left(\mu_n^{\min}\right)^2/K_{2,n} = o(1)$, but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$, the following results hold: as $n \to \infty$, $\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left(\widetilde{L}_n \Lambda_{II,n} \widetilde{L}_n'\right)^{-1/2} \widetilde{L}_n D_\mu\left(\widehat{\delta}_{L,n} - \delta_0\right) \xrightarrow{d} N(0, I_q)$ and $\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left(\widetilde{L}_n \Lambda_{II,n} \widetilde{L}_n'\right)^{-1/2} \widetilde{L}_n D_\mu\left(\widehat{\delta}_{F,n} - \delta_0\right) \xrightarrow{d} N(0, I_q)$.

5 Covariance Matrix Estimation and Hypothesis Testing

To consistently estimate the asymptotic covariance matrix of FELIM and FEFUL, we propose the following estimators

$$\widehat{V}_L = \widehat{H}_L^{-1}\widehat{\Sigma}_L\widehat{H}_L^{-1} \text{ and } \widehat{V}_F = \widehat{H}_F^{-1}\widehat{\Sigma}_F\widehat{H}_F^{-1}, \tag{17}$$

where

$$\begin{aligned}
\widehat{H}_L &= X'\left[A - \widehat{\ell}_{L,n}M^{(Z_1,Q)}\right]X,\ \widehat{H}_F = X'\left[A - \widehat{\ell}_{F,n}M^{(Z_1,Q)}\right]X \\
\widehat{\Sigma}_L &= X'AD\left(J\left[\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right]\right)AX - \widehat{\rho}_L\left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_L\iota_d' \circ M^{(Z,Q)}X\right) \\
&\quad - \left(\widehat{\varepsilon}_L\iota_d' \circ M^{(Z,Q)}X\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)\widehat{\rho}_L' + \widehat{\rho}_L\widehat{\rho}_L'\left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right) \\
&\quad + \left(\widehat{\varepsilon}_L\iota_d' \circ \underline{\widehat{U}}_L\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_L\iota_d' \circ \underline{\widehat{U}}_L\right), \\
\widehat{\Sigma}_F &= X'AD\left(J\left[\widehat{\varepsilon}_F \circ \widehat{\varepsilon}_F\right]\right)AX - \widehat{\rho}_F\left(\widehat{\varepsilon}_F \circ \widehat{\varepsilon}_F\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_F\iota_d' \circ M^{(Z,Q)}X\right) \\
&\quad - \left(\widehat{\varepsilon}_F\iota_d' \circ M^{(Z,Q)}X\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_F \circ \widehat{\varepsilon}_F\right)\widehat{\rho}_F' + \widehat{\rho}_F\widehat{\rho}_F'\left(\widehat{\varepsilon}_F \circ \widehat{\varepsilon}_F\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_F \circ \widehat{\varepsilon}_F\right) \\
&\quad + \left(\widehat{\varepsilon}_F\iota_d' \circ \underline{\widehat{U}}_F\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_F\iota_d' \circ \underline{\widehat{U}}_F\right).
\end{aligned}$$

and where $J = \left[M^Q \circ M^Q\right]^{-1}$, $\widehat{\varepsilon}_L = M^{(Z,Q)}\left(y - X\widehat{\delta}_L\right)$, $\widehat{\varepsilon}_F = M^{(Z,Q)}\left(y - X\widehat{\delta}_F\right)$, $\underline{\widehat{U}}_L = M^{(Z,Q)}X - \widehat{\varepsilon}_L\widehat{\rho}_L'$, and $\underline{\widehat{U}}_F = M^{(Z,Q)}X - \widehat{\varepsilon}_F\widehat{\rho}_F'$. In addition, let $\widehat{\rho}_L = \left[X'M^{(Z,Q)}\left(y - X\widehat{\delta}_L\right)\right] / \left[\left(y - X\widehat{\delta}_L\right)'M^{(Z,Q)}\left(y - X\widehat{\delta}_L\right)\right]$ and $\widehat{\rho}_F = \left[X'M^{(Z,Q)}\left(y - X\widehat{\delta}_F\right)\right] / \left(y - X\widehat{\delta}_F\right)'M^{(Z,Q)}\left(y - X\widehat{\delta}_F\right)$ denote estimators of the parameter $\rho = \lim_{n\to\infty} E\left[U'M^Q\varepsilon\right] / E\left[\varepsilon'M^Q\varepsilon\right]$, based on $\widehat{\delta}_L$ and $\widehat{\delta}_F$, respectively.

Our next result shows the consistency of the covariance matrix estimators given in equation (17) under both Cases I and II[16].

Theorem 4: If Assumptions 1-6 and 8-9 are satisfied; then, the following statements are true:

[16]It can be shown that an estimator of the asymptotic covariance matrix of FEJIV, which will be consistent under both Case I and II, is given by

$$\widehat{V}_{J,n} = \widehat{H}^{-1}\widehat{\Sigma}_J\widehat{H}^{-1} = \left(X'AX\right)^{-1}\left[X'AD_{\widehat{\varsigma}_J}AX + \left(\widehat{\varepsilon}_J \circ \widehat{U}\right)'J\left(A \circ A\right)J\left(\widehat{\varepsilon}_J \circ \widehat{U}\right)\right]\left(X'AX\right)^{-1},$$

where $D_{\widehat{\varsigma}_J} = diag\left(\widehat{\varsigma}_{J,(1,1)}, .., \widehat{\varsigma}_{J,(1,T_1)}, .., \widehat{\varsigma}_{J,(n,1)}, .., \widehat{\varsigma}_{J,(n,T_n)}\right)$, $\widehat{\varsigma}_{J,(i,t)} = e_{(i,t)}'J\left(\widehat{\varepsilon}_J \circ \widehat{\varepsilon}_J\right)$, $\widehat{\varepsilon}_J = M^{(Z,Q)}\left(y - X\widehat{\delta}_J\right)$, and $\widehat{U} = M^{(Z,Q)}X$. Note also that the standard error used for FEJIV in our Monte Carlo study given in section 6 is based on the above formula.

(a) For Case I, where $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$, $D_\mu \widehat{V}_L D_\mu = \Lambda_{I,n} + o_p(1)$ and $D_\mu \widehat{V}_F D_\mu = \Lambda_{I,n} + o_p(1)$, where $\Lambda_{I,n}$ is as defined in equation (14).

(b) For Case II, where $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$, $\left[\left(\mu_n^{\min}\right)^2 / K_{2,n}\right] D_\mu \widehat{V}_L D_\mu = \Lambda_{II,n} + o_p(1)$ and $\left[\left(\mu_n^{\min}\right)^2 / K_{2,n}\right] D_\mu \widehat{V}_F D_\mu = \Lambda_{II,n} + o_p(1)$, where $\Lambda_{II,n}$ is as defined in equation (15).

Theorem 5 below provides results on the asymptotic properties of t-statistics associated with the FELIM and FEFUL estimators when testing a general linear hypothesis of the form $H_0 : c'\delta_0 = r$. We show that the t-ratio based on our estimators has an asymptotic standard normal distribution under the null hypothesis, as long as $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. Moreover, our results show that, under the same rate condition, the tests are also consistent, as our test statistics diverge with probability approaching one under fixed alternatives. Some additional conditions are needed to obtain these results if we were to allow for general heterogeneity in instrument weakness where the diagonal elements $\mu_{g,n}$ $(g = 1, .., d)$ of D_μ can diverge at different rates. These conditions are given in Assumption 10.

Assumption 10: Consider testing the null hypothesis $H_0 : c'\delta_0 = r$. Let

$$\mu_n^*(c) = \min\left\{\mu_{g,n} | g \in \{1, ..., d\}, c_g \neq 0\right\}$$

where c_g is the g^{th} component of the vector c, and let $\mu_n^*(c) D_\mu^{-1} c \to c_* \neq 0$ as $n \to \infty$. Assume that $c_*' \Lambda_{II,n} c_* \geq \underline{C}$ $a.s.$ for all n sufficiently large for some positive constant $\underline{C}$.

Theorem 5: If Assumptions 1-10 are satisfied; then, the following statements are true for the t-statistics $\mathbb{T}_L = \left(c'\widehat{\delta}_L - r\right)/\sqrt{c'\widehat{V}_L c}$ and $\mathbb{T}_F = \left(c'\widehat{\delta}_F - r\right)/\sqrt{c'\widehat{V}_F c}$.

a. For Case I, where $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$:

(i) Under $H_0 : c'\delta_0 = r$, $\mathbb{T}_L \xrightarrow{d} N(0,1)$ and $\mathbb{T}_F \xrightarrow{d} N(0,1)$.

(ii) Under $H_1 : c'\delta_0 \neq r$, with probability approaching one, as $n \to \infty$, the following results hold: $\mathbb{T}_L \to +\infty$ if $c'\delta_0 > r$; $\mathbb{T}_L \to -\infty$ if $c'\delta_0 < r$; $\mathbb{T}_F \to +\infty$ if $c'\delta_0 > r$; and $\mathbb{T}_F \to -\infty$ if $c'\delta_0 < r$.

b. For Case II, where $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$:

(i) Under $H_0 : c'\delta_0 = r$, $\mathbb{T}_L \xrightarrow{d} N(0,1)$ and $\mathbb{T}_F \xrightarrow{d} N(0,1)$.

(ii) Under $H_1 : c'\delta_0 \neq r$, with probability approaching one, as $n \to \infty$, the following results hold: $\mathbb{T}_L \to +\infty$ if $c'\delta_0 > r$; $\mathbb{T}_L \to -\infty$ if $c'\delta_0 < r$; $\mathbb{T}_F \to +\infty$ if $c'\delta_0 > r$; and $\mathbb{T}_F \to -\infty$ if $c'\delta_0 < r$.

.

In looking over the proof of Theorem 5, one can see that the condition stipulating $c_*'\Lambda_{II,n}c_* \geq \underline{C}$ $a.s.n.$ for some constant $\underline{C} > 0$, given in Assumption 10, is only needed in Case II where $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. This is because in this case the covariance matrix estimator is dominated by the contribution of the bilinear term and, when appropriately normalized, this matrix takes the form

$$
\begin{aligned}
&\Lambda_{II,n} \\
= \;& \frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} D_\mu \left(\Pi_n' Z_2' M^{(Z_1,Q)} Z_2 \Pi_n\right)^{-1} \Sigma_{2,n}^* \left(\Pi_n' Z_2' M^{(Z_1,Q)} Z_2 \Pi_n\right)^{-1} D_\mu \\
= \;& \frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} \left(D_\mu^{-1} \frac{D_\mu \Upsilon' Z_2' M^{(Z_1,Q)} Z_2 \Upsilon D_\mu}{n} D_\mu^{-1}\right)^{-1} D_\mu^{-1} \Sigma_{2,n}^* D_\mu^{-1} \\
& \times \left(D_\mu^{-1} \frac{D_\mu \Upsilon' Z_2' M^{(Z_1,Q)} Z_2 \Upsilon D_\mu}{n} D_\mu^{-1}\right)^{-1} \\
= \;& H_n^{-1} \left[\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} D_\mu^{-1} \Sigma_{2,n}^* D_\mu^{-1}\right] H_n^{-1} \qquad (18)
\end{aligned}
$$

where $H_n = \Upsilon' Z_2' M^{(Z_1,Q)} Z_2 \Upsilon / n$ and $\Sigma_{2,n}^*$ is as defined in expression (16) above. Now, the matrix Λ_{II} is singular in large sample when heterogeneity in instrument weakness of a general form is allowed because, even though $K_{2,n}^{-1}\Sigma_{2,n}^*$ can be shown to be positive definite almost surely as $K_{2,n}$, $n \to \infty$[17]; the matrix $\left(\mu_n^{\min}\right) D_\mu^{-1}$ converges to a singular diagonal matrix where some of the diagonal elements are zero, except in the special case where $D_\mu = \left(\mu_n^{\min}\right) \cdot I_d$. It follows that the matrix $\left[\left(\mu_n^{\min}\right)^2 / K_{2,n}\right] D_\mu^{-1} \Sigma_{2,n}^* D_\mu^{-1}$in expression (18) will in general be a singular matrix asymptotically. By following through the derivation given in expression (18), we see that this problem occurs because, under Case II, the covariance matrix has a “denominator" term, i.e., $D_\mu \Upsilon' Z_2' M^{(Z_1,Q)} Z_2 \Upsilon D_\mu / n$, which depends on D_μ but the “numerator" term $\Sigma_{2,n}^*$ does not. Due to this asymmetry, in trying to properly standardize $D_\mu \Upsilon' Z_2' M^{(Z_1,Q)} Z_2 \Upsilon D_\mu / n$ so that its inverse will exist asymptotically, we end up, in some sense, transferring the singularity to the

[17] A proof of the asymptotically positive definiteness of $K_{2,n}^{-1}\Sigma_{2,n}^*$ is given in Lemma S2-3 part (b) of the Additional Online Appendix to this paper, which can be found at the URL:
http://econweb.umd.edu/~chao/Research/research_files/Additional_Online_Appendix_Jackknife_Estimation_Cluster_Sample_IV_Model_December_20_2022.pdf

"numerator". This also explains why this same problem does not arise under Case I, where $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$, since in that case the covariance matrix is dominated in the "numerator" by $D_\mu \Upsilon' Z_2' M^{(Z_1,Q)} D_{\sigma^2} M^{(Z_1,Q)} Z_2 \Upsilon D_\mu / n$, the contribution of the linear term, which is affected by heterogeneity in instrument weakness in the same way as the "denominator", so that, upon proper normalization, the ill effect of this heterogeneity is removed via cancellation.

It should be pointed out that there are important special cases of interest where Assumption 10 either holds automatically or can be shown to hold under mild additional conditions. One such special case is where instrument weakness is homogeneous, i.e., the case where $\mu_{1,n} = \cdots = \mu_{d,n} = \mu_n^{\min}$. In this case, the asymptotic singularity of $\Lambda_{II,n}$ does not arise, so that Assumption 10 is fulfilled without additional side conditions, allowing us to easily obtain the following corollary to Theorem 5.

Corollary 2: Let Assumptions 1-9 be satisfied. Assume further that the diagonal matrix D_μ in Assumption 3 takes the form $D_\mu = \mu_n^{\min} \cdot I_d$ (i.e., $\mu_{1,n} = \cdots = \mu_{d,n} = \mu_n^{\min}$). Then, the following statements are true for the t-statistics $\mathbb{T}_L = \left(c'\widehat{\delta}_L - r\right)/\sqrt{c'\widehat{V}_L c}$ and $\mathbb{T}_F = \left(c'\widehat{\delta}_F - r\right)/\sqrt{c'\widehat{V}_F c}$.

a. For Case I, where $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$:

   (i) Under $H_0 : c'\delta_0 = r$, $\mathbb{T}_L \xrightarrow{d} N(0,1)$ and $\mathbb{T}_F \xrightarrow{d} N(0,1)$.

   (ii) Under $H_1 : c'\delta_0 \neq r$, with probability approaching one as $n \to \infty$, the following results hold: $\mathbb{T}_L \to +\infty$ if $c'\delta_0 > r$; $\mathbb{T}_L \to -\infty$ if $c'\delta_0 < r$; $\mathbb{T}_F \to +\infty$ if $c'\delta_0 > r$; and $\mathbb{T}_F \to -\infty$ if $c'\delta_0 < r$.

b. For Case II, where $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$:

   (i) Under $H_0 : c'\delta_0 = r$, $\mathbb{T}_L \xrightarrow{d} N(0,1)$ and $\mathbb{T}_F \xrightarrow{d} N(0,1)$.

   (ii) Under $H_1 : c'\delta_0 \neq r$, with probability approaching one as $n \to \infty$, the following results hold: $\mathbb{T}_L \to +\infty$ if $c'\delta_0 > r$; $\mathbb{T}_L \to -\infty$ if $c'\delta_0 < r$; $\mathbb{T}_F \to +\infty$ if $c'\delta_0 > r$; and $\mathbb{T}_F \to -\infty$ if $c'\delta_0 < r$.

Corollary 2 is of interest because the case where the degree of instrument weakness is homogeneous and does not vary across the different first-stage equations is one which is often assumed in previous papers on weak and/or many instruments. This includes the well-known papers by Bekker (1994), Staiger and Stock (1997) and Kleibergen (2002). In addition, note that the case where there is only one endogenous regressor is also obviously a special case of the setup considered in Corollary 2.

Another special case of interest is where we test a null hypothesis involving only one coefficient, such as testing the significance of a particular parameter. This case is important because it is the most frequent use of the t-statistic by empirical researchers. In this case, we show in the corollary below that, under mild additional conditions, the t-test based on our proposed estimators will be robust to many weak instruments, even if there is heterogeneity in instrument weakness of a general form.

Assumption 10*: Let e_ℓ denote a $d \times 1$ elementary vector whose ℓ^{th} component is 1 and all other components are 0, and write D_μ in the form

$$D_\mu = \begin{pmatrix} \underset{d_1 \times d_1}{D_1} & 0 \\ 0 & \left(\mu_n^{\min}\right) \cdot I_{d_2} \end{pmatrix}, \tag{19}$$

where $D_1 = diag\left(\mu_{1,n}, .., \mu_{d_1,n}\right)$, where d_1 and d_2 are positive integers such that $d_1 + d_2 = d$, and where $\left(\mu_n^{\min}\right)/\mu_{g,n} \to 0$, as $n \to \infty$, for $g \in \{1, ..., d_1\}$. Partition H_n^{-1} as $H_n^{-1} = \overline{H}_n = \begin{pmatrix} \overline{H}'_{1\cdot} & \overline{H}'_{2\cdot} \end{pmatrix}'$, where $\overline{H}_{1\cdot}$ is $d_1 \times d$ and $\overline{H}_{2\cdot}$ is $d_2 \times d$. Assume that there exists a positive constant C_* such that $e'_\ell \overline{H}'_{2\cdot} \overline{H}_{2\cdot} e_\ell \geq C_* > 0$ $a.s.$for all n sufficiently large.

Corollary 3: Let Assumptions 1-9 and 10* be satisfied; and let e_ℓ be as defined in Assumption 10* above. Consider testing the null hypothesis $H_0 : e'_\ell \delta_0 = r$, using either the t-statistic, $\mathbb{T}_L = \left(e'_\ell \widehat{\delta}_L - r\right) / \sqrt{e'_\ell \widehat{V}_L e_\ell}$ or the t-statistic, $\mathbb{T}_F = \left(e'_\ell \widehat{\delta}_F - r\right) / \sqrt{e'_\ell \widehat{V}_F e_\ell}$.

a. For Case I, where $K_{2,n} / \left(\mu_n^{\min}\right)^2 = O(1)$, the following results hold for $\ell \in \{1, ..., d\}$.

   (i) Under $H_0 : e'_\ell \delta_0 = r$, $\mathbb{T}_L \xrightarrow{d} N(0,1)$ and $\mathbb{T}_F \xrightarrow{d} N(0,1)$.

   (ii) Under $H_1 : e'_\ell \delta_0 \neq r$, with probability approaching one as $n \to \infty$, the following results hold: $\mathbb{T}_L \to +\infty$ if $e'_\ell \delta_0 > r$; $\mathbb{T}_L \to -\infty$ if $e'_\ell \delta_0 < r$; $\mathbb{T}_F \to +\infty$ if $e'_\ell \delta_0 > r$; and $\mathbb{T}_F \to -\infty$ if $e'_\ell \delta_0 < r$.

b. For Case II, where $K_{2,n} / \left(\mu_n^{\min}\right)^2 \to \infty$ but $\sqrt{K_{2,n}} / \left(\mu_n^{\min}\right)^2 \to 0$, the following results hold, for $\ell \in \{1, ..., d\}$.

   (i) Under $H_0 : e'_\ell \delta_0 = r$, $\mathbb{T}_L \xrightarrow{d} N(0,1)$ and $\mathbb{T}_F \xrightarrow{d} N(0,1)$.

   (ii) Under $H_1 : e'_\ell \delta_0 \neq r$, with probability approaching one as $n \to \infty$, the following results hold: $\mathbb{T}_L \to +\infty$ if $e'_\ell \delta_0 > r$; $\mathbb{T}_L \to -\infty$ if $e'_\ell \delta_0 < r$; $\mathbb{T}_F \to +\infty$ if $e'_\ell \delta_0 > r$; and $\mathbb{T}_F \to -\infty$ if $e'_\ell \delta_0 < r$.

Note that writing D_μ in the way specified in equation (19) does not really lead to any loss of generality. In fact, a seemingly more general D_μ matrix, where not all of the diagonal elements grow at the same rate, as $n \to \infty$, can always be rewritten in the form given in equation (19), via repermutation of the rows and columns of D_μ. To see this, suppose that $\mu_{1n}, .., \mu_{d_1,n}, \mu_n^{\min}$ are not ordered as in equation (19), so that we have some diagonal matrix D_μ^*, whose diagonal elements are $\mu_{1n}, .., \mu_{d_1,n}, \mu_n^{\min}$, but in some other order. Then, there exists some permutation matrix P such that $D_\mu = P D_\mu^* P'$, where D_μ is the diagonal matrix given in equation (19). Moreover, let the elements of $\widehat{\delta}^*, \delta_0^*$, c^*, and $\widehat{V}^*$ be ordered in a way that is conformable with D_μ^*, and let $\widehat{\delta}, \delta_0$, c, and $\widehat{V}$ be the corresponding vectors and matrix but with elements ordered conformably with D_μ. Then, it is easy to see that $\widehat{\delta} = P\widehat{\delta}^*$, $\delta_0 = P\delta_0^*$, $c = Pc^*$, $\widehat{V} = P\widehat{V}^* P'$. Hence, by making use of these relations and of the fact that P is an orthogonal matrix, we further obtain that $\mathbb{T}_L^* = c^{*\prime}\left(\widehat{\delta}^* - \delta_0^*\right)/\sqrt{c^{*\prime}\widehat{V}^* c^*} = c^{*\prime}P'P\left(\widehat{\delta}^* - \delta_0^*\right)/\sqrt{c^{*\prime}P'P\widehat{V}^*P'Pc^*} = c'\left(\widehat{\delta} - \delta_0\right)/\sqrt{c'\widehat{V}c} = \mathbb{T}_L$. It follows that the value of the t-statistic is invariant to repermutation of the order of the elements of $\widehat{\delta}$, δ_0, c, and $\widehat{V}$, so that the asymptotic distribution which we derive for $\mathbb{T}_L$, under an assumed ordering of the elements of $\widehat{\delta}$, δ_0, c, and $\widehat{V}$ that is conformable with equation (19) will still apply, even if the t-statistic computed by the empirical researcher is based on some other ordering.

Given that the representation of D_μ given in equation (19) does not result in any loss of generality, the only real restriction imposed by Assumption 10* is the condition that $e_\ell' \overline{H}_{2\cdot}' \overline{H}_{2\cdot} e_\ell \geq C_* > 0$ $a.s.n.$ for some positive constant C_*. We show in the proof of Corollary 3 that this latter condition implies the more general conditions given in Assumption 10 if the null hypothesis we are testing involves only one coefficient. It follows that hypotheses involving only one coefficient can be tested under very general assumptions about the heterogeneity of instrument weakness since the violation of the condition $e_\ell' \overline{H}_{2\cdot}' \overline{H}_{2\cdot} e_\ell \geq C_* > 0$ $a.s.n$ will only occur if the ℓ^{th} column of $\overline{H}_{2\cdot}$ does not have a single nonzero entry, which seems unlikely in most practical applications.

To date, papers in the weak instrument literature have focused primarily on size control, with little attention paid to test consistency under weak identification. One exception is a recent paper by Mikusheva and Sun (2021), which shows that a condition similar to $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$ is both necessary and sufficient for the existence of a consistent test. Interpreted in light of their result, the results presented in Theorems 5 as well as Corollaries 2 and 3 above prove that t-tests based on FELIM and FEFUL are consistent as long as instruments are strong enough so that consistency in hypothesis testing is possible. In contrast, t-tests based on estimators such as the 2SLS estimator will only be consistent if $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to 0$ (i.e., under stronger instruments). Test statistics based on LIML also have undesirable properties under many weak instrument asymptotics, when there is error heteroskedasticity. In addition, note that one advantage of t-tests is that they are particularly easy to apply if one is interested in testing against one-sided alternatives. The results of Theorems 5 as well as Corollaries 2 and 3 show that, when the null hypothesis is incorrect, t-tests based on FELIM and FEFUL diverge in the direction of the true alternative, with probability approaching one, even in situations where identification is weaker than that typically assumed under standard large sample theory, provided of course that $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. Hence, the test statistics proposed in this paper should be particularly useful to empirical researchers interested in testing whether an effect in a particular direction is statistically significant.

6 Monte Carlo Results

In this section, we report some Monte Carlo results based on the following data generating process:

$$\begin{aligned} y_{(i,t)} &= \underset{1\times 1}{\delta}\underset{1\times 1}{x_{(i,t)}} + \underset{1\times 10}{\varphi'}\underset{10\times 1}{Z_{1,(i,t)}} + \alpha_i + \varepsilon_{(i,t)}, \\ x_{(i,t)} &= \underset{1\times 10}{\Phi'}\underset{10\times 1}{Z_{1,(i,t)}} + \underset{1\times K_2}{\Pi}\underset{K_2\times 1}{Z_{2,(i,t)}} + \xi_i + u_{(i,t)}. \end{aligned}$$

where we specify $\varphi = \iota_{10}$, $\Phi = \iota_{10}$, and $\Pi = (\iota_{K_2} \otimes \pi)$ with ι_{10} and ι_{K_2} being, respectively, a 10×1 and a $K_2 \times 1$ vector of ones. Here, π is taken to be a scalar parameter, and we choose π so that the concentration parameter $\mu^2 = 25$, 35, 45, and 55. Moreover, in our experiments, we consider two choices of K_2: $K_2 = 10, 30$. Additionally, we set $n = 200$ and $T_i = 3$, for each $i \in \{1, 2, ..., 200\}$, so that $m_n = 600$. The $(i,t)^{th}$ observation of the vector of included exogenous regressors, or covariates, is taken to be $Z_{1,(i,t)} = \begin{pmatrix} z_{1,(i,t)} & z_{1,(i,t)}^2 & z_{1,(i,t)}^3 & z_{1,(i,t)}^4 & z_{1,(i,t)}D_{(i,t),1} & \cdots & z_{1,(i,t)}D_{(i,t),6} \end{pmatrix}'$, where $\left\{z_{1,(i,t)}\right\}_{(i,t)=1}^{600} \equiv i.i.d.N\left(0,1\right)$ and where $D_{(i,t),k} \in \{0,1\}$ for $k \in \{1, 2, .., 6\}$ is a binary variable such that $\Pr\left(D_{(i,t),k} = 1\right) = 1/2$, with $\left\{D_{(i,t),k}\right\}$ specified to be independent across both (i,t) and k. We also take $\left\{Z_{2,(i,t)}\right\}_{(i,t)=1}^{600} \equiv i.i.d.N\left(0, I_{K_2}\right)$, $\left\{u_{(i,t)}\right\}_{(i,t)=1}^{600} \equiv i.i.d.N\left(0,1\right)$, $\{\alpha_i\}_{i=1}^{200}\, i.i.d.N\left(0,1\right)$, and $\{\xi_i\}_{i=1}^{200}\, i.i.d.N\left(0,1\right)$; with $z_{1,(i,t)}$, $D_{(i,t),k}$, $Z_{2,(i,t)}$, $u_{(i,t)}$, α_i and ξ_i all specified to be independent of each other. We allow the structural disturbance, $\varepsilon_{(i,t)}$, to exhibit conditional heteroskedasticity in a manner similar to the design given in Hausman et al. (2012). In particular, we let

$$\varepsilon_{(i,t)} = \rho u_{(i,t)} + \sqrt{\frac{1-\rho^2}{\phi^2 + (0.86)^2}}\left(\phi v_{1,(i,t)} + 0.86 v_{2,(i,t)}\right), \tag{20}$$

where $v_{1,(i,t)}|Z_{1,(i,t)}, Z_{2,(i,t)} \sim N\left(0, \kappa\left[1 + \left(\iota_{10}' Z_{1,(i,t)} + \iota_{K_2}' Z_{2,(i,t)}\right)^2\right]\right)$ and $v_{2,(i,t)} \sim N\left(0,1\right)$. Both of these distributions are specified to be independent across the index (i,t), and κ is a normalization constant chosen so that the unconditional variance, $Var\left(v_{1,(i,t)}\right)$, is equal to 1. For all experiments reported below, we set $\rho = 0.3$ and choose the parameter ϕ, so that the R-squared for the regression of ε^2 on the instruments and the included exogenous variables take the values 0, 0.1, and 0.2.

Our simulation study examines the finite sample properties of our three estimators (FEJIV, FELIM, and FEFUL) and their associated t-statistics. Additionally, we compare the performance of our estimators with the 2SLS estimator, the IJIVE1 estimator originally proposed in Ackerberg and Devereux (2009), the IJIVE2 estimator introduced in Evdokimov and Kolesár (2018), and the UJIVE estimator originally proposed in Kolesár (2013) and further studied in Evdokimov and Kolesár (2018). The comparison of these point estimators is made on the basis of median bias and nine decile range. We also evaluate the associated t-statistics for these estimators on the basis of size control, as measured by their rejection frequencies under the null hypothesis $H_0 : \delta = 0$.

The results of our Monte Carlo study are reported in Tables 1-6 below.

Table 1: Median Bias, $K_2 = 10$

μ^2	$\mathcal{R}^2_{\varepsilon^2 \mid z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
25	0	0.1092	0.0440	0.0441	0.3811	-0.0058	0.0042	0.0161
	0.1	0.1080	0.0429	0.0427	0.3719	-0.0071	0.0052	0.0167
	0.2	0.1125	0.0475	0.0480	0.3982	-0.0053	0.0051	0.0170
35	0	0.0857	0.0290	0.0290	0.2562	-0.0167	-0.0003	0.0090
	0.1	0.0860	0.0300	0.0301	0.2617	-0.0127	0.0018	0.0107
	0.2	0.0879	0.0328	0.0327	0.2486	-0.0105	-0.0002	0.0083
45	0	0.0733	0.0263	0.0263	0.1943	-0.0095	0.0009	0.0079
	0.1	0.0738	0.0280	0.0281	0.1984	-0.0072	0.0025	0.0094
	0.2	0.0690	0.0213	0.0216	0.1908	-0.0130	-0.0007	0.0057
55	0	0.0629	0.0210	0.0212	0.1586	-0.0074	0.0009	0.0068
	0.1	0.0627	0.0205	0.0206	0.1415	-0.0084	0.0017	0.0071
	0.2	0.0583	0.0167	0.0165	0.1429	-0.0136	-0.0041	0.0017

Results based on 10,000 simulations

Table 2: Nine Decile Range 0.05 to 0.95[18], $K_2 = 10$								
μ^2	$\mathcal{R}^2_{\varepsilon^2 \mid z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
25	0	0.6638	1.0286	1.0247	5.9842	1.5849	1.2900	1.1543
	0.1	0.6590	1.0475	1.0468	6.0422	1.6538	1.2810	1.1442
	0.2	0.6491	1.0235	1.0253	6.1624	1.6218	1.2788	1.1216
35	0	0.5936	0.8214	0.8200	5.6334	1.0861	0.9554	0.8921
	0.1	0.5952	0.8286	0.8288	5.8319	1.1026	0.9430	0.8876
	0.2	0.5755	0.7955	0.7939	5.7128	1.0402	0.9068	0.8561
45	0	0.5362	0.6960	0.6960	5.1587	0.8458	0.7769	0.7433
	0.1	0.5332	0.6876	0.6883	5.2225	0.8378	0.7665	0.7392
	0.2	0.5244	0.6751	0.6753	5.2851	0.8210	0.7542	0.7229
55	0	0.4929	0.6109	0.6114	4.8115	0.7132	0.6620	0.6418
	0.1	0.4929	0.6068	0.6069	4.7546	0.7076	0.6564	0.6387
	0.2	0.4857	0.6039	0.6029	4.8027	0,6972	0.6465	0.6279

Results based on 10,000 simulations

[18] By nine decile range we mean the range between the 0.05 and the 0.95 quantiles. It should also be noted that the reason we compare the estimators based on median bias and nine decile range instead of the usual criteria of (mean) bias and variance is because it is well-known that the exact finite sample (mean) bias and variance of LIML-type estimators do not exist under the assumption that errors are normally distributed. However, it is also well-known that LIML-type estimators tend to be better centered than the 2SLS estimator in terms of median bias and, in many ways, have better finite sample properties, in spite of the fact that they have fatter tails. Hence, the use of median bias and nine decile range allow us to conduct a broader based Monte Carlo comparison without restricting ourselves to only those estimators whose positive integer moments are known to exist.

Table 3: 0.05 Rejection Frequencies[19], $K_2 = 10$								
μ^2	$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
25	0	0.1784	0.0951	0.0884	0.5215	0.0253	0.0518	0.0535
	0.1	0.1842	0.0997	0.0937	0.5181	0.0326	0.0555	0.0561
	0.2	0.1797	0.0958	0.0896	0.5334	0.0275	0.0538	0.0547
35	0	0.1659	0.1064	0.0999	0.5347	0.0319	0.0481	0.0506
	0.1	0.1668	0.1017	0.0951	0.5345	0.0340	0.0489	0.0508
	0.2	0.1677	0.1021	0.0951	0.5369	0.0326	0.0484	0.0511
45	0	0.1584	0.1098	0.1034	0.5601	0.0354	0.0503	0.0528
	0.1	0.1592	0.1087	0.1023	0.5555	0.0351	0.0469	0.0493
	0.2	0.1611	0.1100	0.1042	0.5606	0.0350	0.0483	0.0504
55	0	0.1544	0.1127	0.1053	0.5853	0.0398	0.0476	0.0496
	0.1	0.1583	0.1157	0.1098	0.5835	0.0400	0.0547	0.0561
	0.2	0.1510	0.1123	0.1048	0.5881	0.0401	0.0524	0.0550

Results based on 10,000 simulations

Table 4: Median Bias, $K_2 = 30$								
μ^2	$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
25	0	0.1907	0.1105	0.1106	0.5648	0.0157	0.0042	0.0150
	0.1	0.1916	0.1136	0.1138	0.5828	0.0197	0.0085	0.0217
	0.2	0.1933	0.1159	0.1160	0.5890	0.0287	0.0076	0.0191
35	0	0.1702	0.0954	0.0954	0.4059	0.0069	0.0067	0.0150
	0.1	0.1666	0.0900	0.0901	0.4075	-0.0097	-0.0023	0.0061
	0.2	0.1699	0.0946	0.0941	0.4190	-0.0025	0.0032	0.0124
45	0	0.1501	0.0764	0.0763	0.2939	-0.0050	0.0010	0.0079
	0.1	0.1501	0.0789	0.0788	0.2928	-0.0079	-0.0017	0.0051
	0.2	0.1502	0.0775	0.0778	0.2820	-0.0065	-0.0001	0.0057
55	0	0.1357	0.0670	0.0672	0.2420	-0.0100	-0.0006	0.0051
	0.1	0.1335	0.0641	0.0642	0.2202	-0.0141	-0.0078	-0.0026
	0.2	0.1365	0.0679	0.0682	0.2246	-0.0031	0.0034	0.0092

Results based on 10,000 simulations

[19] See Ackerberg and Devereux (2009), Kolesár (2013), and Evdokimov and Kolesár (2018) for formulae for the estimators IJIVE1, IJIVE2, and UJIVE as well as for the standard errors used in constructing the t-statistics for these estimators.

Table 5: Nine Decile Range 0.05 to 0.95, $K_2 = 30$								
μ^2	$\mathcal{R}^2_{\varepsilon^2\vert z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
25	0	0.4785	0.9885	0.9909	5.7669	3.0848	2.1434	1.7483
	0.1	0.4760	0.9629	0.9634	6.2210	3.0543	2.1265	1.7549
	0.2	0.4693	0.9735	0.9764	6.0341	2.9121	2.1675	1.7602
35	0	0.4501	0.8155	0.8175	6.4148	1.7734	1.4271	1.2895
	0.1	0.4513	0.8083	0.8109	6.2439	1.8113	1.4544	1.3066
	0.2	0.4427	0.7871	0.7899	6.1090	1.7457	1.3613	1.2405
45	0	0.4186	0.6941	0.6939	5.8258	1.2562	1.0510	0.9935
	0.1	0.4254	0.6958	0.6969	5.9272	1.2409	1.0471	0.9948
	0.2	0.4186	0.6771	0.6779	5.9727	1.2126	1.0306	0.9764
55	0	0.4008	0.6206	0.6211	5.7132	0.9825	0.8625	0.8287
	0.1	0.3985	0.6087	0.6109	5.5675	0.9513	0.8614	0.8299
	0.2	0.4028	0.6196	0.6214	5.4996	0.9661	0.8661	0.8354

Results based on 10,000 simulations

Table 6: 0.05 Rejection Frequencies, $K_2 = 30$								
μ^2	$\mathcal{R}^2_{\varepsilon^2\vert z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
25	0	0.4113	0.1387	0.1214	0.5461	0.0249	0.0519	0.0534
	0.1	0.4242	0.1425	0.1226	0.5489	0.0220	0.0518	0.0545
	0.2	0.4350	0.1466	0.1266	0.5527	0.0251	0.0546	0.0565
35	0	0.3919	0.1526	0.1310	0.5387	0.0315	0.0531	0.0553
	0.1	0.3901	0.1527	0.1333	0.5355	0.0298	0.0577	0.0601
	0.2	0.4015	0.1535	0.1338	0.5489	0.0329	0.0572	0.0604
45	0	0.3624	0.1563	0.1362	0.5516	0.0339	0.0539	0.0559
	0.1	0.3639	0.1542	0.1339	0.5396	0.0357	0.0542	0.0564
	0.2	0.3764	0.1551	0.1344	0.5459	0.0370	0.0579	0.0601
55	0	0.3376	0.1485	0.1294	0.5676	0.0385	0.0514	0.0541
	0.1	0.3332	0.1455	0.1277	0.5638	0.0371	0.0534	0.0558
	0.2	0.3530	0.1638	0.1421	0.5686	0.0417	0.0593	0.0605

Results based on 10,000 simulations

Looking over the results reported in Tables 1-6, note first that, in terms of median bias, the performance of FEJIV, FELIM, and FEFUL are uniformly better across our experiments when compared to 2SLS, IJIVE1, IJIVE2, and UJIVE; although our experiments do show 2SLS, IJIVE1, and IJIVE2 to be less dispersed than the three estimators proposed in this paper. Comparing

FELIM and FEFUL in terms of nine decile range, we see that FEFUL tends to be less dispersed than FELIM, which is in accord with the motivation behind the original Fuller (1977) modification. Perhaps the most notable difference in performance is that t-statistics based on FELIM and FEFUL have much less size distortion than t-statistics constructed from any of the other five estimators. Finally, note that t-statistics based on the FEJIV estimator tend to be undersized, but the empirical rejection frequencies are still closer to the nominal level than t-statistics based on 2SLS, IJIVE1, IJIVE2, or UJIVE.

7 Conclusion

This paper considers an IV regression model with many weak instruments, cluster specific effects, error heteroskedasticity, and possibly many included exogenous regressors. To carry out point estimation in this setup, we propose three new jackknife-type IV estimators, which we refer to by the acronyms FEJIV, FELIM, and FEFUL. All three of these estimators are shown to be robust to the effects of many weak instruments, in the sense that they are consistent estimators within a framework broad enough to include both the standard situation with strong instruments and situations with many weak instruments. To the best of our knowledge, the estimators proposed in this paper are the first consistent estimators which have been developed in a many weak instrument framework when the IV regression under consideration has both cluster specific effects and possibly many included exogenous regressors. We establish asymptotic normality for FELIM and FEFUL under both strong instrument and many weak instrument asymptotics. In addition, we provide consistent standard errors for our estimators and show that, when the null hypothesis is true, t-statistics based on these standard errors are asymptotically normal under both strong instrument and many weak instrument asymptotics. Finally, we show that under both strong instrument and many weak instrument asymptotics, the t-statistics based on these standard errors are consistent under fixed alternatives. Thus, we underscore an interesting aspect of the many weak instrument setup. Namely, test consistency is still possible under this framework, as has been pointed out in a recent paper by Mikusheva and Sun (2021). In a series of Monte Carlo experiments, we find that t-statistics based on FELIM and FEFUL control size better in finite samples than t-statistics based on alternative jackknife-type IV estimators that have previously been proposed in the literature. Hence, based on the findings of this paper, we recommend that either FELIM or FEFUL be used in settings where there are many weak instruments, cluster specific effects, and possibly many included exogenous regressors.

References

[1] Ackerberg, D.A. and P.J. Devereux (2009): Improved JIVE Estimators for Overidentified Linear Models with and without Heteroskedasticity, Review of Economics and Statistics, 91, 351-362.

[2] Angrist, J.D., G.W. Imbens, and A. Krueger (1999): Jackknife Instrumental Variables Estimation, Journal of Applied Econometrics, 14, 57-67.

[3] Antoine, B. and E. Renault (2012): Efficient Minimum Distance Estimation with Multiple Rates of Convergence, Journal of Econometrics, 170, 350-357.

[4] Bekker, P. A. (1994): Alternative Approximations to the Distributions of Instrumental Variables Estimators, Econometrica, 62, 657-681.

[5] Bekker, P. A. and F. Crudu (2015): Jackknife Instrumental Variable Estimation with Heteroskedasticity, Journal of Econometrics, 185, 332-342.

[6] Blomquist, S. and M. Dahlberg (1999): Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments, Journal of Applied Econometrics, 14, 69-88.

[7] Cattaneo, M. D., M. Jansson, and W.K. Newey (2018): Inference in Linear Regression Models with Many Covariates and Heteroskedasticity, Journal of the American Statistical Association, 113, 1350-1361.

[8] Chao, J. C. and N. R. Swanson (2004): Estimation and Testing Using Jackknife IV in Heteroskedastic Regressions with Many Weak Instruments, Unpublished Manuscript, University of Maryland and Rutgers University.

[9] Chao, J. C. and N. R. Swanson (2005): Consistent Estimation with a Large Number of Weak Instruments, Econometrica, 73, 1673-1692.

[10] Chao, J.C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012): Asymptotic Distribution of JIVE in a Heteroskedastic IV Regression with Many Instruments, Econometric Theory, 28, 42-86.

[11] Crudu, F., G. Mellace. and Z. Sándor (2020): Inference in Instrumental Variables Models with Heteroskedasticity and Many Instruments, Econometric Theory, forthcoming.

[12] Davidson, R. and J. G. MacKinnon (2006): The Case Against JIVE, Journal of Applied Econometrics, 21, 827–833.

[13] Evdokimov, K. S. and M. Kolesár (2018): Inference in Instrumental Variables Analysis with Heterogeneous Treatment Effects, Unpublished Manuscript, Princeton University.

[14] Fuller, W. A. (1977): Some Properties of a Modification of the Limited Information Estimator, Econometrica, 45, 939-954.

[15] Hansen, C., J. A. Hausman, and W. K. Newey (2008): Estimation with Many Instrumental Variables, Journal of Business & Economic Statistics, 26, 398-422.

[16] Hastie, T. and R. Tibshirani (1990): Generalized Additive Models. London: Chapman and Hall.

[17] Hausman, J. A., W. K. Newey, T. Woutersen, J. C. Chao, and N. R. Swanson (2012): Instrumental Variable Estimation with Heteroskedasticity and Many Instruments, Quantitative Economics, 3, 211-255.

[18] Hsiao, C. and Q. Zhou (2018): JIVE for Panel Dynamic Simultaneous Equations Models, Econometric Theory, 34, 1325-1369.

[19] Kleibergen, F. (2002): Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression, Econometrica, 70, 1781-1803.

[20] Kolesár, M. (2013): Estimation in Instrumental Variables Models with Treatment Effect Heterogeneity, Unpublished Manuscript, Princeton University.

[21] Mikusheva, A. and L. Sun (2021): Inference with Many Weak Instruments, Review of Economic Studies, forthcoming.

[22] Morimune, K. (1983): Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability Is Large Compared with the Sample Size, Econometrica, 51, 821-841.

[23] Newey, W. K. (1997): Convergence Rates and Asymptotic Normality for Series Estimators, Journal of Econometrics, 79, 147-168.

[24] Phillips, G. D. A. and C. Hale (1977): The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems, International Economic Review, 18, 219-228.

[25] Phillips, P. C. B. (1983). Exact Small Sample Theory in the Simultaneous Equations Model, in Handbook of Econometrics, Vol. I, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland, 449-516.

[26] Rothenberg, T. (1984): Approximating the Distributions of Econometric Estimators and Test Statistics, in Handbook of Econometrics, Vol. II, ed. by Z. Griliches and M.D. Intriligator. Amsterdam: North-Holland, 881-935.

[27] Staiger, D. and J. H. Stock (1997): Instrumental Variables Regression with Weak Instruments, Econometrica, 65, 557-586.

[28] Stock, J. H. and M. Yogo (2005): Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments, in Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, ed. by D.W.K. Andrews and J. H. Stock, Chapter 6. Cambridge: Cambridge University Press.

[29] Tao, T. (2012): Topics in Random Matrix Theory. Graduate Studies in Mathematics, 132. Providence, RI: American Mathematical Society.

8 Appendix: Proofs of Main Theorems and Other Key Results

This appendix provides the proofs for Theorem 1, Corollary 1, Theorems 4-5, and Corollaries 2-3 of the paper. The proofs of Theorems 2 and 3 are longer and, thus, are given in Appendix S1 of a Supplemental Appendix to this paper. This Supplemental Appendix can be viewed at the URL: http://econweb.umd.edu/~chao/Research/research_files/Supplemental_Appendix_to_Jackknife _Estimation_Cluster_Sample_IV_Model_December_20_2022.pdf. In addition, the proofs provided below rely on a number of technical results that are stated without proof in Appendix S2 of the Supplemental Appendix. These results are designated in the derivations that follow by the use of the prefix S. So, for example, Lemma S2-2 will refer to the second lemma in Appendix S2 of the Supplemental Appendix. Proofs for these additional supporting lemmas (more specifically, Lemmas S2-1 to S2-18) are available in a separate online appendix which can be viewed at the URL:
http://econweb.umd.edu/~chao/Research/research_files/Additional_Online_Appendix_ Jackknife_Estimation_Cluster_Sample_IV_Model_December_20_2022.pdf

Proof of Theorem 1:

To proceed, note first that, by parts (a) and (b) of Lemma S2-2 and by the assumption on $\overline{\ell}_n$, we have $D_\mu^{-1}X'\left[A-\overline{\ell}_n M^{(Z_1,Q)}\right]XD_\mu^{-1}=D_\mu^{-1}X'AXD_\mu^{-1}-\overline{\ell}_n D_\mu^{-1}X'M^{(Z_1,Q)}XD_\mu^{-1}=H_n+o_p(1)$, where $H_n=\Upsilon' Z_2' M^{(Z_1,Q)}Z_2\Upsilon/n=O_p(1)$. By Assumption 3(iii), we also have that H_n is positive definite almost surely for n sufficiently large, so that $D_\mu^{-1}X'\left[A-\overline{\ell}_n M^{(Z_1,Q)}\right]XD_\mu^{-1}$ is invertible w.p.a.1. Hence, w.p.a.1., we can write

$$\frac{1}{\mu_n^{\min}} D_\mu \left(\overline{\delta}_n - \delta_0\right) = \left(D_\mu^{-1} X' \left[A - \overline{\ell}_n M^{(Z_1,Q)}\right] X D_\mu^{-1}\right)^{-1} \frac{1}{\mu_n^{\min}} D_\mu^{-1} X' \left[A - \overline{\ell}_n M^{(Z_1,Q)}\right] \varepsilon.$$

Applying Lemma S2-4 and Lemma S2-5, we get

$$\begin{aligned} \frac{1}{\mu_n^{\min}} D_\mu^{-1} X' \left[A - \overline{\ell}_n M^{(Z_1,Q)}\right] \varepsilon &= \frac{1}{\mu_n^{\min}} D_\mu^{-1} X' A \varepsilon - \overline{\ell}_n \frac{1}{\mu_n^{\min}} D_\mu^{-1} X' M^{(Z_1,Q)} \varepsilon \\ &= O_p\left(\max\left\{\frac{1}{\mu_n^{\min}}, \frac{\sqrt{K_{2,n}}}{\left(\mu_n^{\min}\right)^2}\right\}\right) + o_p(1) = o_p(1). \end{aligned}$$

It follows by the Slutsky's Theorem that $\left\| D_\mu \left(\overline{\delta}_n - \delta_0\right) / \left(\mu_n^{\min}\right) \right\|_2 = o_p(1)$, which gives the first result. To show the second result, note that, by straightforward calculations, we obtain $\left\| D_\mu \left(\overline{\delta}_n - \delta_0\right) / \left(\mu_n^{\min}\right) \right\|_2 \geq \sqrt{\left(\mu_n^{\min}\right)^2 / \left(\mu_n^{\min}\right)^2} \sqrt{\left(\overline{\delta}_n - \delta_0\right)' \left(\overline{\delta}_n - \delta_0\right)} = \left\| \overline{\delta}_n - \delta_0 \right\|_2$, which implies that $\left\| \overline{\delta}_n - \delta_0 \right\|_2 \xrightarrow{p} 0$, as required. □

Proof of Corollary 1:

In light of the results given in Theorem 1, it suffices that we verify the condition $\overline{\ell}_n = o_p\left(\left[\mu_n^{\min}\right]^2 / n\right) = o_p(1)$ for all three estimators. For the FEJIV estimator considered in part (a), $\overline{\ell}_n = 0$ for all n, so this condition is trivially satisfied. Now, part (b) considers the FELIM estimator. For this estimator, the result of Lemma S2-11 has shown that we can take $\overline{\ell}_n = \widehat{\ell}_{L,n} = \min_{\beta \in \overline{B}} \left(\beta' \overline{X}' A \overline{X} \beta\right) / \left(\beta' \overline{X}' M^{(Z_1,Q)} \overline{X} \beta\right)$ $= \left(y - X\widehat{\delta}_L\right)' A \left(y - X\widehat{\delta}_L\right) / \left[\left(y - X\widehat{\delta}_L\right)' M^{(Z_1,Q)} \left(y - X\widehat{\delta}_L\right)\right]$. By part (a) of Lemma S2-7, we then have $\widehat{\ell}_{L,n} = o_p\left(\left[\mu_n^{\min}\right]^2 / n\right)$, so FELIM also satisfies the needed condition. Finally, part (c) considers the FEFUL estimator, which takes $\overline{\ell}_n = \widehat{\ell}_{F,n}$ $= \left[\widehat{\ell}_{L,n} - \left(1 - \widehat{\ell}_{L,n}\right)(C/m_n)\right] / \left[1 - \left(1 - \widehat{\ell}_{L,n}\right)(C/m_n)\right]$. By part (b) of Lemma S2-7, we have that $\widehat{\ell}_{F,n} = o_p\left(\left[\mu_n^{\min}\right]^2 / n\right)$, so the needed condition is satisfied again. The consistency results given in parts (a)-(c) of this corollary then follow as a consequence of Theorem 1. □

Proof of Theorem 4:

We shall prove this theorem for the FELIM case since the proof for FEFUL is similar. To proceed, first define $S_{L,1} = X' A D \left(J \left[\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right]\right) A X$, $S_{L,2} = \left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)' J (A \circ A) J \left(\widehat{\varepsilon}_L \iota_d' \circ M^{(Z,Q)} X\right)$, $S_{L,3} = \left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)' J (A \circ A) J \left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)$, $\underline{S}_{L,4} = \left(\widehat{\varepsilon}_L \iota_d' \circ \widehat{\underline{U}}_L\right)' J (A \circ A) J \left(\widehat{\varepsilon}_L \iota_d' \circ \widehat{\underline{U}}_L\right)$, $\widehat{H}_L = X' \left[A - \widehat{\ell}_{L,n} M^{(Z_1,Q)}\right] X$, $\Sigma_{1,n} = \Upsilon' Z_2' M^{(Z_1,Q)} D_{\sigma^2} M^{(Z_1,Q)} Z_2 \Upsilon / n$. In addition, also define $\sigma_{(i,t)}^2 = E\left[\varepsilon_{(i,t)}^2 | \mathcal{F}_n^Z\right]$, $\phi_{(i,t)} = E\left[U_{(i,t)} \varepsilon_{(i,t)} | \mathcal{F}_n^Z\right]$, $\Psi_{(i,t)} = E\left[U_{(i,t)} U_{(i,t)}' | \mathcal{F}_n^Z\right]$, $\underline{\phi}_{(i,t)} = E\left[\underline{U}_{(i,t)} \varepsilon_{(i,t)} | \mathcal{F}_n^Z\right]$, and $\underline{\Psi}_{(i,t)} = E\left[\underline{U}_{(i,t)} \underline{U}_{(i,t)}' | \mathcal{F}_n^Z\right]$ where $\underline{U}_{(i,t)} = U_{(i,t)} - \rho \varepsilon_{(i,t)}$ and where for notational convenience we suppress the dependence of $\sigma^2_{(i,t)}$, $\phi_{(i,t)}$, $\Psi_{(i,t)}$, $\underline{\phi}_{(i,t)}$, and $\underline{\Psi}_{(i,t)}$ on $\mathcal{F}^Z_n = \sigma(Z)$.

Using these notations, to show part (a), we first write $D_\mu \widehat{V}_L D_\mu = \widehat{V}_{L,1} + \widehat{V}_{L,2} + \widehat{V}_{L,3} + \widehat{V}_{L,4}$, where $\widehat{V}_{L,1} = \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} D_\mu^{-1} S_{L,1} D_\mu^{-1} \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1}$,
$\widehat{V}_{L,2} = -\left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} D_\mu^{-1}\left(\widehat{\rho}_L S_{L,2} + S_{L,2}'\widehat{\rho}_L'\right) D_\mu^{-1} \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1}$,
$\widehat{V}_{L,3} = \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} D_\mu^{-1}\widehat{\rho}_L S_{L,3}\widehat{\rho}_L' D_\mu^{-1} \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1}$, and $\widehat{V}_{L,4} = \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1}$
$\times D_\mu^{-1}\underline{S}_{L,4} D_\mu^{-1} \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1}$. Now, consider $\widehat{V}_{L,1}$ first. Note that, by Lemma S2-17,

$$D_\mu^{-1} X' A D\left(\varepsilon \circ \varepsilon\right) A X D_\mu^{-1} = \Sigma_{1,n} + \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} \sigma^2_{(i,t)} D_\mu^{-1} \Psi_{(j,s)} D_\mu^{-1} + o_p(1),$$

from which we deduce that $D_\mu^{-1} X' A D\left(\varepsilon \circ \varepsilon\right) A X D_\mu^{-1} = O_p(1)$ using Assumptions 2(i) and 3(iii), Lemma S2-1 part (a), and the assumption that $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$ under Case I.

Next, note that by Lemma S2-11,
$\widehat{\ell}_L = \left(y - X\widehat{\delta}_L\right)' A \left(y - X\widehat{\delta}_L\right) / \left(y - X\widehat{\delta}_L\right)' M^{(Z_1,Q)} \left(y - X\widehat{\delta}_L\right)$. Moreover, by the result given in Lemma S2-10, we have that $D_\mu^{-1}\widehat{H}_L D_\mu^{-1} = H_n + o_p(1)$, where, by Assumption 3(iii), $H_n = \Upsilon' Z_2' M^{(Z_1,Q)} Z_2 \Upsilon / n$ is positive definite *a.s.n.* In addition, we can apply part (a) of Lemma S2-18 and Slutsky's theorem to deduce that

$$\begin{aligned} \widehat{V}_{L,1} &= \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} D_\mu^{-1} S_{L,1} D_\mu^{-1} \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} \\ &= H_n^{-1}\Sigma_{1,n}H_n^{-1} + H_n^{-1}\left(\sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} \sigma^2_{(i,t)} D_\mu^{-1} \Psi_{(j,s)} D_\mu^{-1}\right) H_n^{-1} + o_p(1). \end{aligned} \tag{21}$$

Next, consider $\widehat{V}_{L,2}$. Here, note that we can further decompose $\widehat{V}_{L,2}$ as $\widehat{V}_{L,2} = \widehat{V}_{L,2,1} + \widehat{V}_{L,2,2}$, where
$\widehat{V}_{L,2,1} = -\left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} D_\mu^{-1}\widehat{\rho}_L S_{L,2} D_\mu^{-1} \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1}$ and
$\widehat{V}_{L,2,2} = -\left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} D_\mu^{-1} S_{L,2}'\widehat{\rho}_L' D_\mu^{-1} \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1}$. Noting that $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$ under Case I and applying the result of Lemma S2-10, as well as parts (d) and (e) of Lemma S2-18 and Slutsky's theorem, we get

$$\begin{aligned} \widehat{V}_{L,2,1} &= -H_n^{-1}\frac{K_{2,n}}{\left(\mu_n^{\min}\right)}\left\{D_\mu^{-1}\rho + D_\mu^{-1}\left(\widehat{\rho}_L - \rho\right)\right\}\frac{\mu_n^{\min}}{K_{2,n}} S_{L,2} D_\mu^{-1} H_n^{-1}\left(1 + o_p(1)\right) \\ &= -H_n^{-1} D_\mu^{-1}\rho \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} \sigma^2_{(i,t)} \phi'_{(j,s)} D_\mu^{-1} H_n^{-1} + o_p(1). \end{aligned}$$

Moreover, since $\widehat{V}_{L,2,2} = \widehat{V}'_{L,2,1}$, we also have $\widehat{V}_{L,2,2} = -H_n^{-1} \sum_{(i,t),(j,s)=1:m_n,(i,t)\neq(j,s)} A^2_{(i,t),(j,s)} \sigma^2_{(i,t)} D_\mu^{-1} \phi_{(j,s)} \rho' D_\mu^{-1} H_n^{-1} + o_p(1)$. Given that $\widehat{V}_{L,2} = \widehat{V}_{L,2,1} + \widehat{V}_{L,2,2}$, it follows from these calculations that

$$\widehat{V}_{L,2} = -H_n^{-1} \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} D_\mu^{-1} \left(\rho \sigma^2_{(i,t)} \phi'_{(j,s)} + \sigma^2_{(i,t)} \phi_{(j,s)} \rho' \right) D_\mu^{-1} H_n^{-1} + o_p(1) \tag{22}$$

Turning our attention to $\widehat{V}_{L,3}$, note that, in this case, we can apply Lemma S2-10, parts (b) and (e) of Lemma S2-18, and Slutsky's theorem to obtain

$$\begin{aligned} \widehat{V}_{L,3} &= K_{2,n} H_n^{-1} \left[D_\mu^{-1} \rho + D_\mu^{-1} (\widehat{\rho}_L - \rho) \right] \frac{S_{L,3}}{K_{2,n}} \rho' D_\mu^{-1} H_n^{-1} (1 + o_p(1)) \\ &\quad + K_{2,n} H_n^{-1} \left[D_\mu^{-1} \rho + D_\mu^{-1} (\widehat{\rho}_L - \rho) \right] \frac{S_{L,3}}{K_{2,n}} (\widehat{\rho}_L - \rho)' D_\mu^{-1} H_n^{-1} (1 + o_p(1)) \\ &= H_n^{-1} D_\mu^{-1} \rho \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} \sigma^2_{(i,t)} \sigma^2_{(j,s)} \rho' D_\mu^{-1} H_n^{-1} + o_p(1). \end{aligned} \tag{23}$$

Lastly, we consider $\widehat{V}_{L,4}$. Here, we can apply Lemma S2-10, part (f) of Lemma S2-18, the fact that $K_{2,n} / \left(\mu_n^{\min} \right)^2 = O(1)$ under Case I, as well as Slutsky's theorem to obtain

$$\begin{aligned} \widehat{V}_{L,4} &= H_n^{-1} D_\mu^{-1} \underline{S}_{L,4} D_\mu^{-1} H_n^{-1} (1 + o_p(1)) \\ &= H_n^{-1} \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} D_\mu^{-1} \underline{\phi}_{(i,t)} \underline{\phi}'_{(j,s)} D_\mu^{-1} H_n^{-1} + o_p(1). \end{aligned} \tag{24}$$

It follows from equations (21), (22), (23), and (24) that

$$\begin{aligned} D_\mu \widehat{V}_L D_\mu &= H_n^{-1} \Sigma_{1,n} H_n^{-1} + H_n^{-1} \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} \sigma^2_{(i,t)} D_\mu^{-1} \underline{\Psi}_{(j,s)} D_\mu^{-1} H_n^{-1} \\ &\quad + H_n^{-1} \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} D_\mu^{-1} \underline{\phi}_{(i,t)} \underline{\phi}'_{(j,s)} D_\mu^{-1} H_n^{-1} + o_p(1) \\ &= H_n^{-1} (\Sigma_{1,n} + \Sigma_{2,n}) H_n^{-1} + o_p(1) = \Lambda_{I,n} + o_p(1). \end{aligned}$$

To show the same result for FEFUL, note that $\widehat{\delta}_F$ satisfies the conditions of both Lemma S2-12 and Lemma S2-18. Hence, we can make the same argument as given above for FELIM, except that we use the result of Lemma S2-12 in lieu of Lemma S2-10 to obtain $D_\mu \widehat{V}_F D_\mu = H_n^{-1}(\Sigma_{1,n}+\Sigma_{2,n})H_n^{-1}+o_p(1)=\Lambda_{I,n}+o_p(1)$.

To show part (b), we again only provide an explicit argument for $\widehat{V}_L$ since the proof of $\widehat{V}_F$ follows in a similar way. To proceed, write $\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right] D_\mu \widehat{V}_L D_\mu = \left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]\sum_{\ell=1}^4 \widehat{V}_{L,\ell}$, where $\widehat{V}_{L,1}$, $\widehat{V}_{L,2}$, $\widehat{V}_{L,3}$, and $\widehat{V}_{L,4}$ are as defined in the proof of part (a).

Considering $\widehat{V}_{L,1}$ first, note that, since $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$ under Case II, we have, upon applying the result of Lemma S2-10, part (a) of Lemma S2-18, and Slutsky's theorem,

$$\begin{aligned}
\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\widehat{V}_{L,1} &= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\left[\Sigma_{1,n}+\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2\sigma_{(i,t)}^2 D_\mu^{-1}\Psi_{(j,s)}D_\mu^{-1}\right]H_n^{-1}(1+o_p(1))\\
&= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2\sigma_{(i,t)}^2 D_\mu^{-1}\Psi_{(j,s)}D_\mu^{-1}H_n^{-1}+o_p(1)\,. \qquad (25)
\end{aligned}$$

Now, consider $\widehat{V}_{L,2}$. Here, we write $\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]\widehat{V}_{L,2} = \left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]\widehat{V}_{L,2,1}$ $+\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]\widehat{V}_{L,2,2}$, where $\widehat{V}_{L,2,1}$ and $\widehat{V}_{L,2,2}$ are again as defined in the proof of part (a). Making use of the results of Lemma S2-10, parts (d) and (e) of Lemma S2-18, and Slutsky's theorem while noting that $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ under Case II, we get

$$\begin{aligned}
\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\widehat{V}_{L,2,1} &= -H_n^{-1}\left(\mu_n^{\min}\right)\left\{D_\mu^{-1}\rho+D_\mu^{-1}(\widehat{\rho}_L-\rho)\right\}\frac{\mu_n^{\min}}{K_{2,n}}S_{L,2}D_\mu^{-1}H_n^{-1}(1+o_p(1))\\
&= -H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2 D_\mu^{-1}\rho\sigma_{(i,t)}^2\phi_{(j,s)}' D_\mu^{-1}H_n^{-1}+o_p(1)
\end{aligned}$$

Moreover, since $\widehat{V}_{L,2,2}=\widehat{V}_{L,2,1}'$, we also have
$\left[\left(\mu_n^{\min}\right)^2 K_{2,n}^{-1}\right]\widehat{V}_{L,2,2} = -H_n^{-1}\left[\left(\mu_n^{\min}\right)^2 K_{2,n}^{-1}\right]\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2 D_\mu^{-1}\phi_{(j,s)}\sigma_{(i,t)}^2\rho' D_\mu^{-1}H_n^{-1}+o_p(1)$.
It follows from these calculations that

$$\frac{\left(\mu_n^{\min}\right)^2\widehat{V}_{L,2}}{K_{2,n}} = -H_n^{-1}\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n}\frac{\left(\mu_n^{\min}\right)^2 A_{(i,t),(j,s)}^2}{K_{2,n}}D_\mu^{-1}\left(\rho\sigma_{(i,t)}^2\phi_{(j,s)}'+\phi_{(j,s)}\sigma_{(i,t)}^2\rho'\right)D_\mu^{-1}H_n^{-1}+o_p(1)\,. \qquad (26)$$

Next, consider $\widehat{V}_{L,3}$. Given that $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ under Case II, we get, upon applying the result given in Lemma S2-10, as well as parts (b) and (e) of Lemma S2-18 and Slutsky's theorem,

$$\begin{aligned}
\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\widehat{V}_{L,3} &= \left(\mu_n^{\min}\right)^2 H_n^{-1}\left[D_\mu^{-1}\rho + D_\mu^{-1}\left(\widehat{\rho}_L - \rho\right)\right]\frac{S_{L,3}}{K_{2,n}}\rho' D_\mu^{-1} H_n^{-1}\left(1 + o_p\left(1\right)\right) \\
&\quad + \left(\mu_n^{\min}\right)^2 H_n^{-1}\left[D_\mu^{-1}\rho + D_\mu^{-1}\left(\widehat{\rho}_L - \rho\right)\right]\frac{S_{L,3}}{K_{2,n}}\left(\widehat{\rho}_L - \rho\right)' D_\mu^{-1} H_n^{-1}\left(1 + o_p\left(1\right)\right) \\
&= H_n^{-1}\sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n}\frac{\left(\mu_n^{\min}\right)^2 A_{(i,t),(j,s)}^2}{K_{2,n}} D_\mu^{-1}\rho\sigma_{(i,t)}^2\sigma_{(j,s)}^2\rho' D_\mu^{-1} H_n^{-1} + o_p\left(1\right) \qquad (27)
\end{aligned}$$

Finally, we consider $\widehat{V}_{L,4}$. Again, noting that $K_{2,n}/\left(\mu_n^{\min}\right)^2 \to \infty$ under Case II, we have, upon applying the result given in Lemma S2-10, as well as part (f) of Lemma S2-18 and Slutsky's theorem,

$$\begin{aligned}
\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\widehat{V}_{L,4} &= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} D_\mu^{-1}\underline{S}_{L,4} D_\mu^{-1} H_n^{-1}\left(1 + o_p\left(1\right)\right) \\
&= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2 D_\mu^{-1}\underline{\phi}_{(i,t)}\underline{\phi}_{(j,s)}' D_\mu^{-1} H_n^{-1} + o_p\left(1\right). \qquad (28)
\end{aligned}$$

It follows from equations (25), (26), (27), and (28) that

$$\begin{aligned}
\frac{\left(\mu_n^{\min}\right)^2 D_\mu \widehat{V}_L D_\mu}{K_{2,n}} &= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2 D_\mu^{-1}\left(\sigma_{(i,t)}^2\underline{\Psi}_{(j,s)} + \underline{\phi}_{(i,t)}\underline{\phi}_{(j,s)}'\right) D_\mu^{-1} H_n^{-1} + o_p\left(1\right) \\
&= \frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} H_n^{-1}\Sigma_{2,n} H_n^{-1} + o_p\left(1\right) = \Lambda_{II,n} + o_p\left(1\right).
\end{aligned}$$

To show the same result for FEFUL, note again that $\widehat{\delta}_F$ satisfies the conditions of Lemmas S2-12 and S2-18. Hence, we can make the same argument as given above for FELIM, except using Lemma S2-12 in lieu of Lemma S2-10 to obtain $\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right] D_\mu \widehat{V}_F D_\mu =$
$\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right] H_n^{-1}\Sigma_{2,n} H_n^{-1} + o_p\left(1\right) = \Lambda_{II,n} + o_p\left(1\right)$. $\square$

Proof of Theorem 5:

To show part (a), first note that, by part (d) of Lemma S2-3 and Assumption 3(iii), $\Lambda_{I,n}$ is positive definite $a.s.n.$ In addition, making use of part (a) of Theorem 4, we have $D_\mu \widehat{V}_L D_\mu =$

$\Lambda_{I,n}+o_p(1)$, so that $D_\mu \widehat{V}_L D_\mu$ is positive definite w.p.a.1. Hence, under $H_0 : c'\delta_0 = r$, we can write

$$\mathbb{T}_L = \frac{c'\widehat{\delta}_{L,n}-r}{\sqrt{c'\widehat{V}_L c}} = \frac{c'\left(\widehat{\delta}_{L,n}-\delta_0\right)}{\sqrt{c'\widehat{V}_L c}} = \frac{\left(c'D_\mu^{-1}\mu_n^*(c)\right)\Lambda_{I,n}^{1/2}\left[\Lambda_{I,n}^{-1/2}D_\mu\left(\widehat{\delta}_{L,n}-\delta_0\right)\right]}{\sqrt{\left(c'D_\mu^{-1}\mu_n^*(c)\right)D_\mu\widehat{V}_L D_\mu\left(\mu_n^*(c)\,D_\mu^{-1}c\right)}}$$

Applying Theorem 2, we have $\Lambda_{I,n}^{-1/2}D_\mu\left(\widehat{\delta}_{L,n}-\delta_0\right)\xrightarrow{d} N(0,I_d)$. It follows by the definition of c_* given in Assumption 10, as well as by applying part (a) of Theorem 4 and the continuous mapping theorem that

$$\mathbb{T}_L = \frac{c_*'\Lambda_{I,n}^{1/2}\left[\Lambda_{I,n}^{-1/2}D_\mu\left(\widehat{\delta}_{L,n}-\delta_0\right)\right]}{\sqrt{c_*'\Lambda_{I,n}c_*}}\left[1+o_p(1)\right]\xrightarrow{d} N(0,1)\,. \tag{29}$$

On the other hand, under H_1, we have $c'\delta_0 = r+h$ for some $h\in\mathbb{R}\backslash\{0\}$, and we can write $\mathbb{T}_L = \left(c'\widehat{\delta}_{L,n}-r\right)/\sqrt{c'\widehat{V}_L c} = c'\left(\widehat{\delta}_{L,n}-\delta_0\right)/\sqrt{c'\widehat{V}_L c}+h/\sqrt{c'\widehat{V}_L c}$. The first term above is $O_p(1)$, as shown in (29) above, whereas application of part (a) of Theorem 4, Assumption 10, and the Slutsky's theorem shows that $(\mu_n^*(c))^2 c'\widehat{V}_L c = \left(c'D_\mu^{-1}\mu_n^*(c)\right)D_\mu\widehat{V}_L D_\mu\left(\mu_n^*(c)\,D_\mu^{-1}c\right) = c_*'\Lambda_{I,n}c_*+o_p(1)$, where $c_*'\Lambda_{I,n}c_*>0$ since $\Lambda_{I,n}$ is positive definite in light of part (d) of Lemma S2-3 and Assumption 3(iii) and since $c_*\neq 0$ by construction. In addition, by parts (a) and (c) of Lemma S2-3; Assumption 3(iii); and the fact that, under Case I, $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$; there exists a positive constant $C<\infty$ such that, almost surely for all n sufficiently large,

$$\lambda_{\max}\left(\Lambda_{I,n}\right)\leq\frac{\lambda_{\max}\left[VC\left(\Upsilon' Z_2' M^{(Z_1,Q)}\varepsilon/\sqrt{n}\right)|\mathcal{F}_n^Z\right]+\frac{K_{2,n}}{\left(\mu_n^{\min}\right)^2}\lambda_{\max}\left[VC\left(\underline{U}'A\varepsilon/\sqrt{K_{2,n}}\right)|\mathcal{F}_n^Z\right]}{\left[\lambda_{\min}\left(H_n\right)\right]^2}\leq C. \tag{30}$$

It follows that, in this case, $h/\sqrt{c'\widehat{V}_L c} = \mu_n^*(c)\,h/\sqrt{(\mu_n^*(c))^2 c'\widehat{V}_L c} = \left(\mu_n^*(c)\,h/\sqrt{c_*'\Lambda_{I,n}c_*}\right)\left[1+o_p(1)\right]$. So, w.p.a.1, $h/\sqrt{c'\widehat{V}_L c}\to+\infty$ if $h>0$, whereas $h/\sqrt{c'\widehat{V}_L c}\to-\infty$ if $h<0$, from which the stated result follows. Finally, note that the results for $\mathbb{T}_F$ can be shown in the same way, so to avoid redundancy, we omit the proof.

To show part (b), we first let $\widetilde{L}_n = \mu_n^*(c)\,c'D_\mu^{-1}$; and note that, by Assumption 10, there exist a constant vector $c_*\neq 0$ and a positive constant $\underline{C}$ such $\widetilde{L}_n = \mu_n^*(c)\,c'D_\mu^{-1}\to c_*'$ and $c_*'\Lambda_{II,n}c_*\geq\underline{C}>0$ *a.s.n.* It follows that, in this case, the conditions for $\widetilde{L}_n$ given in Theorem 3 are trivially satisfied. Applying Theorem 3, we then obtain
$\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left[\mu_n^*(c)\,c'D_\mu^{-1}\Lambda_{II,n}D_\mu^{-1}c\mu_n^*(c)\right]^{-1/2}\mu_n^*(c)\,c'D_\mu^{-1}\left[D_\mu\left(\widehat{\delta}_{L,n}-\delta_0\right)\right]$
$=\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left[c_*'\Lambda_{II,n}c_*\right]^{-1/2}c_*'\left[D_\mu\left(\widehat{\delta}_{L,n}-\delta_0\right)\right]\left[1+o_p(1)\right]\xrightarrow{d} N(0,1)$. Moreover,
$\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]D_\mu\widehat{V}_L D_\mu = \Lambda_{II,n}+o_p(1)$ by part (b) of Theorem 4. Now, under $H_0 : c'\delta_0 = r$, we

can write

$$\mathbb{T}_L = \frac{c'\widehat{\delta}_{L,n} - r}{\sqrt{c'\widehat{V}_L c}} = \frac{\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\mu_n^*(c)\,c' D_\mu^{-1}\left[D_\mu\left(\widehat{\delta}_{L,n} - \delta_0\right)\right]}{\sqrt{\left(\mu_n^*(c)\,c' D_\mu^{-1}\right)\left[\left\{\left(\mu_n^{\min}\right)^2/K_{2,n}\right\} D_\mu \widehat{V}_L D_\mu\right]\left(D_\mu^{-1} c \mu_n^*(c)\right)}}$$

from which it follows that

$$\mathbb{T}_L = \frac{\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right) c_*'\left[D_\mu\left(\widehat{\delta}_{L,n} - \delta_0\right)\right]}{\sqrt{c_*'\Lambda_{II,n} c_*}}\left[1 + o_p(1)\right] \xrightarrow{d} N(0,1). \tag{31}$$

Under H_1, we again write $c'\delta_0 = r + h$ for some $h \in \mathbb{R}\backslash\{0\}$, and note that, in this case, by applying Assumption 10, part (b) of Theorem 4, and Slutsky's theorem; we have $(\mu_n^*(c))^2\left\{\left(\mu_n^{\min}\right)^2/K_{2,n}\right\} c'\widehat{V}_L c = \left(\mu_n^*(c)\,c' D_\mu^{-1}\right)\left[\left\{\left(\mu_n^{\min}\right)^2/K_{2,n}\right\} D_\mu \widehat{V}_L D_\mu\right]\left(D_\mu^{-1} c \mu_n^*(c)\right)$ $= c_*'\Lambda_{II,n} c_* + o_p(1)$. Moreover, there exists a positive constant $\underline{C}$ such that $c_*'\Lambda_{II,n} c_* \geq \underline{C} > 0$ $a.s.n.$ by Assumption 10. In addition, by part (c) of Lemma S2-3 and Assumption 3(iii), there exists a positive constant C such that, almost surely for all n sufficiently large

$$\lambda_{\max}\left(\Lambda_{II,n}\right) \leq \frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} \frac{1}{\left[\lambda_{\min}\left(H_n\right)\right]^2} \frac{K_{2,n}}{\left(\mu_n^{\min}\right)^2} \lambda_{\max}\left[VC\left(\frac{\underline{U}' A \varepsilon}{\sqrt{K_{2,n}}}\right) | \mathcal{F}_n^Z\right] \leq C < \infty. \tag{32}$$

It follows that, for this case,

$$\frac{h}{\sqrt{c'\widehat{V}_L c}} = \frac{\mu_n^*(c)\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right) h}{\sqrt{(\mu_n^*(c))^2\left\{\left(\mu_n^{\min}\right)^2/K_{2,n}\right\} c'\widehat{V}_L c}} = \frac{\mu_n^*(c)\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right) h}{\sqrt{c_*'\Lambda_{II,n} c_*}}\left[1 + o_p(1)\right].$$

Hence, w.p.a.1, $h/\sqrt{c'\widehat{V}_L c} \to +\infty$ if $h > 0$ whereas $h/\sqrt{c'\widehat{V}_L c} \to -\infty$ if $h < 0$, given the condition that $\left(\mu_n^{\min}\right)^2/\sqrt{K_{2,n}} \to \infty$ and given that, by construction, $\mu_n^{\min}/\mu_n^*(c) = O(1)$. Finally, write

$$\mathbb{T}_L = \frac{c'\widehat{\delta}_{L,n} - r}{\sqrt{c'\widehat{V}_L c}} = \frac{c'\left(\widehat{\delta}_{L,n} - \delta_0\right)}{\sqrt{c'\widehat{V}_L c}} + \frac{h}{\sqrt{c'\widehat{V}_L c}}.$$

Since the first term on the right-hand side above is $O_p(1)$ as shown in (31), we deduce that w.p.a.1, $\mathbb{T}_L \to +\infty$ if $h > 0$ and $\mathbb{T}_L \to -\infty$ if $h < 0$. The results for $\mathbb{T}_F$ can be shown in the same way, so to avoid redundancy, we omit the proof. □

Proof of Corollary 2: Note that the assumptions and setup of Corollary 2 is essentially the same as that of Theorem 5, except that we do not assume the more general conditions given in Assumption

10 but rather we assume the specialized structure where $D_\mu = \mu_n^{\min} \cdot I_d$. Hence, to prove this corollary, we need to show that $D_\mu = \mu_n^{\min} \cdot I_d$ implies that Assumption 10 is satisfied. To proceed, note that, trivially in this case, $\mu_n^* (c) = \mu_n^{\min}$ so that $\mu_n^* (c) D_\mu^{-1} c = \mu_n^{\min} \left[\left(\mu_n^{\min} \right)^{-1} \cdot I_d \right] c = c$ for all n. Thus, $c_* = c \neq 0$ in this case. Moreover, there exists a positive constant $\underline{C}$ such that $c_*' \Lambda_{II,n} c_* = c' \Lambda_{II,n} c = \left(\mu_n^{\min} \right)^2 c' H_n^{-1} \Sigma_{2,n} H_n^{-1} c / K_{2,n} = \left(\mu_n^{\min} \right)^2 c' H_n^{-1} D_\mu^{-1} VC \left(\underline{U}' A \varepsilon / \sqrt{K_{2,n}} | \mathcal{F}_n^Z \right) D_\mu^{-1} H_n^{-1} c = c' H_n^{-1} VC \left(\underline{U}' A \varepsilon / \sqrt{K_{2,n}} | \mathcal{F}_n^Z \right) H_n^{-1} c \geq \underline{C} > 0$ *a.s.n.* for all $c \neq 0$, by the almost sure positive definiteness of $VC \left(\underline{U}' A \varepsilon / \sqrt{K_{2,n}} | \mathcal{F}_n^Z \right)$ as shown in part (b) of Lemma S2-3, which completes the proof. $\square$

Proof of Corollary 3: Note that the assumptions and setup of Corollary 3 is essentially the same as that of Theorem 5, except that we do not assume the more general conditions given in Assumption 10. Instead, we consider the special case where $c = e_\ell$ for $\ell \in \{1, ..., d\}$; and, in lieu of Assumption 10, we assume the condition that there exists a positive constant C_* such that $e_\ell' \overline{H}_{2\cdot}' \overline{H}_{2\cdot} e_\ell \geq C_* > 0$ *a.s.n.* Hence, to prove this corollary, we need to show that, in the case where the problem of interest is testing the null hypothesis $H_0 : c' \delta_0 = e_\ell' \delta_0 = r$, the condition that $e_\ell' \overline{H}_{2\cdot}' \overline{H}_{2\cdot} e_\ell \geq C_* > 0$ *a.s.n.* implies the conditions given in Assumption 10. To proceed, note first that since $c = e_\ell$ here, we have $\mu_n^* (c) = \min \left\{ \mu_{g,n} | g \in \{1, ..., d\} \text{ and } c_g \neq 0 \right\} = \mu_{\ell,n}$, so that $\mu_n^* (c) D_\mu^{-1} c = \mu_{\ell,n} D_\mu^{-1} e_\ell = \mu_{\ell,n} \left(\mu_{\ell,n} \right)^{-1} e_\ell = e_\ell$. Thus, $c_* = e_\ell \neq 0$ in this case. Moreover, note that

$$\left(\mu_n^{\min} \right) D_\mu^{-1} = \left(\mu_n^{\min} \right) \begin{pmatrix} D_1^{-1} & 0 \\ 0 & \left(\mu_n^{\min} \right)^{-1} \cdot I_{d_2} \end{pmatrix} \to \begin{pmatrix} 0 & 0 \\ 0 & I_{d_2} \end{pmatrix} = D_0, \quad (say).$$

It follows that, in this case, $c_*' \Lambda_{II,n} c_* = e_\ell' \Lambda_{II,n} e_\ell = \left(\mu_n^{\min} \right)^2 e_\ell' H_n^{-1} D_\mu^{-1} \Sigma_{2,n}^* D_\mu^{-1} H_n^{-1} e_\ell / K_{2,n}$
$= e_\ell' H_n^{-1} D_0 VC \left(\underline{U}' A \varepsilon / \sqrt{K_{2,n}} | \mathcal{F}_n^Z \right) D_0 H_n^{-1} e_\ell \left[1 + o_{a.s.} (1) \right] \geq C e_\ell' \overline{H}_{2\cdot}' \overline{H}_{2\cdot} e_\ell \geq C C_* = \underline{C} > 0$ *a.s.n.*, by the fact that $VC \left(\underline{U}' A \varepsilon / \sqrt{K_{2,n}} | \mathcal{F}_n^Z \right) \geq C I_d$ *a.s.n.* for some positive constant C, as shown in part (b) of Lemma S2-3, and by the assumption that $e_\ell' \overline{H}_{2\cdot}' \overline{H}_{2\cdot} e_\ell \geq C_* > 0$ *a.s.n.*. This completes the proof. $\square$