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*: September 15, 2021

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 our estimators are asymptotically normal under the null and consistent under fixed alternatives. Our Monte Carlo results further 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 economics, despite the vast literature on the topic. One key reason for this is that there are many different regression settings for which endogeneity is an issue, but for which valid estimators are not currently available. One such setting

*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 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 about the topic of this paper. 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.

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 (2020). 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) both point out that LIML is also inconsistent under many weak instruments, when there is error heteroskedasticity. Estimators that are robust to the effects of many weak instruments in cross sectional settings with error heteroskedasticity turn out to have a jackknife form, as discussed in Chao and Swanson (2004). 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 the structural equation of interest to have a partially linear form so that additional exogenous regressors can enter the equation nonlinearly. In this sense, our paper is also related to recent work by Cattaneo, Jansson, and Newey (2018a,b) on the partially linear model. However, the focus of these papers differs from ours, as they do not consider the problem of endogeneity. Thus, rather than employing IV estimators, estimation is done using an OLS estimator, with the nonlinear component being first approximated nonparametrically by a set of basis functions.

To consistently estimate the parameters of 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 is also 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 estimators, IJIVE1 and IJIVE2, of Ackerberg and Devereux (2009), and 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 in greater detail in Section 2. On the other hand, the estimation procedures that we introduce here are carefully designed to properly partial out the presence of 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 standard (strong but fixed number of instruments) asymptotics and also under many weak instrument 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 this paper 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

[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 5, 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).

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), 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 (2020) 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 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 3 establish that our estimators are consistent and asymptotically normally distributed. Section 4 shows how to estimate the variances of the estimators and also provides asymptotic results for t-statistics based on our estimators. Section 5 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 6 concludes. Proofs of Theorem 1, Corollary 1, and Theorems 4-6 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

[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.pdf

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 $\|A\|_\infty \equiv \max_{1\leq i\leq m_n} \sum_{j=1}^{m_n} |a_{ij}|$ (i.e., 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. 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 Model, Assumptions, and Estimation Procedures

The model that we consider is a cluster-sample IV regression model

$$\underset{1\times 1}{y_{(i,t)}} = X'_{(i,t)}\delta_0 + \varphi_n\left(W_{1,(i,t)}\right) + \alpha_i + \varepsilon_{(i,t)}, \tag{1}$$

$$X_{(i,t)} = \Phi_n\left(W_{1,(i,t)}\right) + \Upsilon_n\left(W_{2,(i,t)}\right) + \xi_i + U_{(i,t)}, \tag{2}$$

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 $(1,1) = 1$, $(1,T_1) = T_1$, $(2,1) = T_1 + 1$, and $(n,T_n) = m_n$. 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 $W_{1,(i,t)}$ and $W_{2,(i,t)}$ denote, respectively, a $p_1 \times 1$ vector and a $p_2 \times 1$ vector of exogenous variables, for $i = 1,2,...,n$ and $t = 1,...,T_i$ (or, equivalently, for $(i,t) = 1,...,m_n$). Note that $\varphi_n(\cdot)$, $\Phi_n(\cdot)$, and $\Upsilon_n(\cdot)$ are allowed to be nonlinear functions, so that the structural equation (1) can be taken to be a partially linear equation, and the system of first-stage equations given by (2) may be interpreted as a generalized additive model in the sense of Hastie and Tibshirani (1990). In addition, α_i and ξ_i in the above equations are 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 $W_{1,(i,t)}$ and $W_{2,(i,t)}$, unlike the typical assumptions specified in a traditional "random effects" model. More precise assumptions on the model given by equations (1) and (2) are given below.

We will develop some additional notations before proceeding. First, let $W_{(i,t)} = \left(W'_{1,(i,t)}, W'_{2,(i,t)}\right)'$, for $(i,t) = 1, ..., m_n$, and define $W_n = \left(W_{(1,1)}, .., W_{(1,T_1)}, ..., W_{(n,1)}, .., W_{(n,T_n)}\right)'$. Now, stacking the observations $(i,t) = 1, ..., m_n$, we can write the model given by equations (1) and (2) more succinctly as

$$\underset{m_n\times 1}{y} = \underset{m_n\times d}{X}\underset{d\times 1}{\delta_0} + \underset{m_n\times 1}{\varphi_n} + \underset{m_n\times n}{Q}\underset{n\times 1}{\alpha} + \underset{m_n\times 1}{\varepsilon}, \tag{3}$$

$$\underset{m_n\times d}{X} = \underset{m_n\times d}{\Phi_n} + \underset{m_n\times d}{\Upsilon_n} + \underset{m_n\times n}{Q}\underset{n\times d}{\Xi} + \underset{m_n\times d}{U}, \tag{4}$$

where $\alpha = (\alpha_1, ..., \alpha_n)'$, $\Xi = (\xi_1, ..., \xi_n)'$, and

$$\underset{m_n\times n}{Q} = \begin{pmatrix} \iota_{T_1} & 0 & \cdots & 0 \\ 0 & \iota_{T_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \iota_{T_n} \end{pmatrix}.$$

and where the other vectors and matrices are stacked similar to W_n. For notational convenience, we have suppressed the dependence of φ_n, Φ_n, and Υ_n on W_n. 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.

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

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

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

Assumption 3: Suppose that $\Upsilon_n\left(W_{2,(i,t)}\right) = D_\mu\gamma\left(W_{2,(i,t)}\right)/\sqrt{n}$, for $(i,t) = 1, ..., m_n$, where $D_\mu = diag\left(\mu_{1,n}, .., \mu_{d,n}\right)$. 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) Let $\mu_n^{\min} = \min_{1\leq k\leq d} \mu_{k,n}$, and suppose that $\mu_n^{\min} \to \infty$, as $n \to \infty$, such that $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. (iii) $\lambda_{\min}(H_n) \geq 1/C > 0$ and $\lambda_{\max}(\Gamma'\Gamma/n) \leq C < \infty$ $a.s.$, for all n sufficiently large, where $H_n = \Gamma' M^{(Z_1,Q)}\Gamma/n$ and

$$\underset{m_n\times d}{\Gamma}=\begin{pmatrix}\gamma\left(W_{2,(1,1)}\right) & \cdots & \gamma\left(W_{2,(1,T_1)}\right) & \cdots & \gamma\left(W_{2,(n,1)}\right) & \cdots & \gamma\left(W_{2,(n,T_n)}\right)\end{pmatrix}'.$$

Assumption 4: Suppose that $\Phi_n\left(W_{1,(i,t)}\right)=D_\kappa f\left(W_{1,(i,t)}\right)/\sqrt{n}$ and $\varphi_n\left(W_{1,(i,t)}\right)$ $=\tau_n g\left(W_{1,(i,t)}\right)/\sqrt{n}$, for $(i,t)=1,...,m_n$, 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 $\tau_n/\sqrt{n}\to 0$.

We shall consider approximating the function $\gamma\left(W_{2,(i,t)}\right)$ in Assumption 3 by the series expansion $\sum_{k=1}^{K_{2,n}}\pi_k z_{2,k}\left(W_{2,(i,t)}\right)$, for some family of approximating functions $z_{2,1}(\cdot), z_{2,2}(\cdot),.....$[4]. Stacking the observed values of these functions into a matrix, we obtain the $m_n\times K_{2,n}$ matrix, $Z_2=\left[Z_2\left(W_{2,(1,1)}\right),..,Z_2\left(W_{2,(1,T_1)}\right),...,Z_2\left(W_{2,(n,1)}\right),..,Z_2\left(W_{2,(n,T_n)}\right)\right]$, where $Z_2\left(W_{2,(i,t)}\right)=\left(z_{2,1}\left(W_{2,(i,t)}\right),....,z_{2,K_{2,n}}\left(W_{2,(i,t)}\right)\right)'$ is a $K_{2,n}\times 1$ vector, for each $(i,t)\in\{1,...,m_n\}$. Similarly, we shall approximate the functions $f\left(W_{1,(i,t)}\right)$ and $g\left(W_{1,(i,t)}\right)$ given in Assumption 4 by, respectively, the series expansions $\sum_{k=1}^{K_{1,n}}\Theta_k z_{1,k}\left(W_{2,(i,t)}\right)$ and $\sum_{k=1}^{K_{1,n}}\theta_k z_{1,k}\left(W_{2,(i,t)}\right)$, for a family of approximating functions $z_{1,1}(\cdot), z_{1,2}(\cdot),.....$ Stacking the observed values of these functions, we get the $m_n\times K_{1,n}$ matrix, $Z_1=\left[Z_1\left(W_{1,(1,1)}\right),..,Z_1\left(W_{1,(1,T_1)}\right),...,Z_1\left(W_{1,(n,1)}\right),..,Z_1\left(W_{1,(n,T_n)}\right)\right]$, with $Z_1\left(W_{1,(i,t)}\right)=\left(z_{1,1}\left(W_{1,(i,t)}\right),....,z_{1,K_{1,n}}\left(W_{1,(i,t)}\right)\right)'$ being a $K_{1,n}\times 1$ vector, for each $(i,t)\in\{1,...,m_n\}$. For notational convenience, we shall suppress the dependence of Z_1 and Z_2 on W_n. In addition, let $\underset{m_n\times K_n}{Z}=\begin{bmatrix}Z_1 & Z_2\end{bmatrix}$, with $K_n=K_{1,n}+K_{2,n}$, and let $\Pi^{K_{2,n}}=\left(\pi_1,...,\pi_{K_{2,n}}\right)'$, $\Theta^{K_{1,n}}=\left(\Theta_1,...,\Theta_{K_{1,n}}\right)'$, and $\theta^{K_{1,n}}=\left(\theta_1,...,\theta_{K_{1,n}}\right)'$.

Analogous to the linear IV model, we could interpret Z_2 as the matrix of observations on the instruments of the model and Z_1 as the matrix of observations on the additional covariates or included exogenous regressors. Viewed in this light, we see that 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 (as discussed in Chao and Swanson (2005), Hausman et al. (2012), and Chao et al. (2012)), if this condition is not fulfilled,

[4] The approximating functions $z_{2,k}(\cdot)$, for $k=1,..,K_{2,n}$, may be polynomials, splines, or some other family of functions as discussed in Newey (1997).

then consistent estimation of at least some of the components of δ may not be possible; and, in this paper, we focus only on situations where we can consistently estimate δ. To interpret this condition, it is easiest to consider the special case where $\mu_{1,n} = \cdots = \mu_{d,n} = \mu_n^{\min}$. In this case, $\left(\mu_n^{\min}\right)^2$ can be interpreted as giving the order of magnitude of the signal component of the IV model, whereas $\sqrt{K_{2,n}}$ measures the order of magnitude of a leading noise term, so that, in order for consistent estimation to be possible, the signal-to-noise ratio $\left(\mu_n^{\min}\right)^2 / \sqrt{K_{2,n}}$ must diverge to infinity.

Likewise, Assumption 4 allows for possible local-to-zero modeling of the nonlinear components $g\left(W_{1,(i,t)}\right)$ and $f\left(W_{1,(i,t)}\right)$. 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 structural equation of interest becomes a standard partially linear model, whereas the system of first-stage equations becomes a standard multivariate generalized additive model . 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 $M^Q = I_{m_n} - Q\left(Q'Q\right)^{-1} Q'$. 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 let $P^{Z_1^{\perp}} = M^Q Z_1 \left(Z_1' M^Q Z_1\right)^{-1} Z_1' M^Q$, where $M^{(Z_1,Q)} = M^Q - M^Q Z_1 \left(Z_1' M^Q Z_1\right)^{-1} Z_1' M^Q$ with M^Q as defined in part (iii) above. Suppose that $\max_{1 \leq (i,t) \leq m_n} P^{Z_1^{\perp}}_{(i,t),(i,t)} = O_{a.s.}\left(K_{1,n}/n\right)$ and $\max_{1 \leq (i,t) \leq m_n} P^{\perp}_{(i,t),(i,t)} = O_{a.s.}\left(K_{2,n}/n\right)$.

Note that part (ii) of Assumption 5 requires that both $K_{1,n}$ and $K_{2,n}$ grow at a rate slower than the sample size n. Hence, our setup does not include the type of many-instrument setup of Morimune (1983) and Bekker (1994), where the number of instruments grows at the same rate as n, nor the type of many-regressor setup of Cattaneo, Jansson, and Newey (2018a,b), where the number of exogenous regressors grows to infinity on the same order as n. However, note that our assumptions are in accord with the interpretation of the structural equation as a partially linear model with the nonlinear component being estimated nonparametrically by a series estimator, as Newey (1997) has given results which show that consistent series estimation requires κ_n, the number of approximating functions, to grow slower than the sample size n. In fact, for consistent series estimation using regression spline functions, Newey (1997) provides an explicit rate restriction, where $\kappa_n^2/n \to 0$ as $n \to \infty$, and our assumption is in accord with this rate condition. Moreover, note that our setup does allow the number of fixed or cluster-specific effects to be on the order of n, so that the number of overall covariates in the structural equation of interest will be on the order of n. Hence, we believe that our framework is general enough to accommodate a wide range of empirical problems of interest.

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: Let $\mathcal{W}_1 \subseteq \mathbb{R}^{p_1}$ and $\mathcal{W}_2 \subseteq \mathbb{R}^{p_2}$ denote the support of $W_{1,(i,t)}$ and $W_{2,(i,t)}$, respectively. The following rates of approximation are assumed. (i) There exists a positive real number ϱ_g and a vector of coefficients $\theta^{K_{1,n}}$, such that

$$\left\|g(\cdot) - \theta^{K_{1,n}\prime} Z_1(\cdot)\right\|_{\infty} = O_{a.s.}\left(K_{1,n}^{-\varrho_g}\right), \text{ as } K_{1,n} \to \infty,$$

where $\tau_n / K_{1,n}^{\varrho_g} = o(1)$ and $\left\|g(\cdot) - \theta^{K_{1,n}\prime} Z_1(\cdot)\right\|_{\infty} = \sup_{w_1 \in \mathcal{W}_1} \left|g(w_1) - \theta^{K_{1,n}\prime} Z_1(w_1)\right|$.
(ii) There exists a positive real number ϱ_f and a matrix of coefficients $\Theta^{K_{1,n}}$, such that

$$\left\|f(\cdot) - \Theta^{K_{1,n}\prime} Z_1(\cdot)\right\|_{\infty,d} = O_{a.s.}\left(K_{1,n}^{-\varrho_f}\right), \text{ as } K_{1,n} \to \infty,$$

where $\kappa_n^{\max} / K_{1,n}^{\varrho_f} = O(1)$, with $\kappa_n^{\max} = \max_{1\leq \ell \leq d} \kappa_{\ell,n}$ and $\left\|f(\cdot) - \Theta^{K_{1,n}\prime} Z_1(\cdot)\right\|_{\infty,d}$
$= \max_{\ell \in \{1,2,\dots,d\}} \sup_{w_1 \in \mathcal{W}_1} \left|f_\ell(w_1) - e_{\ell,d}' \Theta^{K_{1,n}\prime} Z_1(w_1)\right|$, with $e_{\ell,d}$ denoting a $d \times 1$ elementary vector with 1 in the ℓ^{th} component and 0 in all other components.
(iii) There exists a positive real number ϱ_γ and a matrix of coefficients $\Pi^{K_{2,n}}$, such that

$$\left\|\gamma(\cdot) - \Pi^{K_{2,n}\prime} Z_2(\cdot)\right\|_{\infty,d} = O_{a.s.}\left(K_{2,n}^{-\varrho_\gamma}\right), \text{ as } K_{2,n} \to \infty,$$

where $\mu_n^{\max} / K_{2,n}^{\varrho_\gamma} = O(1)$, with $\mu_n^{\max} = \max_{1\leq \ell \leq d} \mu_{\ell,n}$ and $\left\|\gamma(\cdot) - \Pi^{K_{2,n}\prime} Z_2(\cdot)\right\|_{\infty,d}$
$= \max_{\ell \in \{1,2,\dots,d\}} \sup_{w_2 \in \mathcal{W}_2} \left|\gamma_\ell(w_2) - e_{\ell,d}' \Pi^{K_{2,n}\prime} Z_2(w_2)\right|$.
(iv) Assume that $\max_{1\leq (i,t) \leq m_n} \left\|\Gamma' M^{(Z_1,Q)} e_{(i,t)}\right\|_2 / \sqrt{n} = o_p(1)$.

A few comments about Assumption 7 are in order. Parts (i)-(iii) of this assumption place conditions on the rate at which the error in approximating the functions $g(\cdot)$, $f(\cdot)$, $\gamma(\cdot)$, and $\mu_\Gamma(\cdot)$ must vanish uniformly. A similar assumption has been specified in Newey (1997) in studying convergence rates for nonparametric series estimators (see Assumption 3 of that paper). As noted in that paper, the size of the exponents, such as ϱ_g, ϱ_f, ϱ_γ, and ϱ_μ, depends both on the degree of smoothness of the function to be approximated (i.e., the number of continuous derivatives that the functions has) and on the dimension of the argument of the function (i.e., the dimension of $W_{1,(i,t)}$ or $W_{2,(i,t)}$ in our case). For example, under Assumption 7(i), if the approximating functions used are splines or polynomials, then $\left\|g(\cdot) - \theta^{K_{1,n}\prime} Z_1(\cdot)\right\|_{\infty} = O_{a.s.}\left(K_{1,n}^{-\varrho_g}\right)$, with $\varrho_g = s/p_1$, where s is the number of continuous derivatives of the function $g(\cdot)$ and p_1 is the dimension of $W_{1,(i,t)}$. Since our results require that these approximation errors vanish sufficiently fast, this, in turn, places certain requirements on the smoothness of the functions $g(\cdot)$, $f(\cdot)$, $\gamma(\cdot)$, and $\mu_\Gamma(\cdot)$ and on the dimension of $W_{1,(i,t)}$ or $W_{2,(i,t)}$, with some trade-offs between the two. Finally, note that part (iv) of Assumption 7 is similar to a condition given in Assumption 3 of Cattaneo, Jansson, and Newey (2018b). As noted in that paper, this condition comes close to being minimal for the central limit theorem to hold.

Although our specification allows the structural equation (1) and the system of first-stage equations (2) to have nonlinear components, the results that we give in this paper will also, of course, hold under a linear specification with (possibly) many weak instruments and/or many weak covariates taking the form

$$y_{(i,t)} = X'_{(i,t)}\delta_0 + \frac{\tau_n \theta^{K_{1,n}\prime} Z_{1,(i,t)}}{\sqrt{n}} + \alpha_i + \varepsilon_{(i,t)}, \tag{5}$$

$$X_{(i,t)} = \frac{D_\kappa \Theta^{K_{1,n}\prime} Z_{1,(i,t)}}{\sqrt{n}} + \frac{D_\mu \Pi^{K_{2,n}\prime} Z_{2,(i,t)}}{\sqrt{n}} + \xi_i + U_{(i,t)}, \tag{6}$$

where $i = 1, ..., n$ and $t = 1, ..., T_i$. In this case, $Z_{1,(i,t)}$ and $Z_{2,(i,t)}$ will be exogenous regressors/instruments that need not depend on other variables such as $W_{1,(i,t)}$ and $W_{2,(i,t)}$. In addition, in a strictly linear setup, parts (i)-(iii) of Assumption 7 that impose conditions on the approximation errors by series estimation will no longer be needed.

Assumption 8: Let $\rho_n = E\left[U'M^Q\varepsilon\right]/E\left[\varepsilon'M^Q\varepsilon\right]$. Suppose that 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 (1), 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:**
$$\widehat{\delta}_J = \left(X'AX\right)^{-1} X'Ay,$$
where $A = P^\perp - M^{(Z,Q)} D_{\widehat{\vartheta}} M^{(Z,Q)}$, $P^\perp = P^{(Z,Q)} - P^{(Z_1,Q)}$, and $M^{(Z,Q)} = I_{m_n} - P^{(Z,Q)}$, with $P^{(Z,Q)}$ and $P^{(Z_1,Q)}$ being projection matrices that project into the column space of $\begin{bmatrix} Z & Q \end{bmatrix}$ and $\begin{bmatrix} Z_1 & Q \end{bmatrix}$, respectively. 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 $P^\perp$.

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

$$\widehat{Q}_{FELIM}(\delta) = \frac{(y - X\delta)' A (y - X\delta)}{(y - X\delta)' M^{(Z_1,Q)} (y - X\delta)}, \tag{7}$$

where A is as defined above in the definition of FEJIV and $M^{(Z_1,Q)} = I_{m_n} - P^{(Z_1,Q)}$, with $P^{(Z_1,Q)}$ as defined above. $\widehat{\delta}_L$ has the explicit representation

$$\widehat{\delta}_L = \left(X' \left[A - \widehat{\ell}_L M^{(Z_1,Q)}\right] X\right)^{-1} \left(X' \left[A - \widehat{\ell}_L M^{(Z_1,Q)}\right] y\right), \tag{8}$$

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 $\overline{X} = \left[\begin{array}{cc} y & X \end{array}\right]$.

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

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

where $M^{(Z_1,Q)} = I_{m_n} - P^{(Z_1,Q)}$ and $\widehat{\ell}_F = \left[\widehat{\ell}_L - \left(1 - \widehat{\ell}_L\right) C/m_n\right] / \left[1 - \left(1 - \widehat{\ell}_L\right) C/m_n\right]$ for some constant C. Here, $\widehat{\ell}_L$ is the smallest root of equation $\det\left\{\overline{X}' A \overline{X} - \ell \overline{X}' M^{(Z_1,Q)} \overline{X}\right\} = 0$. For the Monte Carlo results reported in section 5, 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

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

so that, at least in large samples, the "numerator" of the right-hand side of this equation has a familiar form (i.e., it is in terms of a linear form $\Upsilon_n' A \varepsilon$ 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 $d = (d_{11}, d_{22}, ..., d_{m_n,m_n})'$. 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. This proper centering, in turn, allows $\widehat{\delta}_J$ to be both consistent and asymptotically normal under many weak instrument asymptotics. In addition, write $\widehat{\delta}_J - \delta_0 = (X'AX)^{-1} X'A(\varphi_n + Q\alpha + \varepsilon)$, and note that

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

Looking at the equation above, we see that the matrix A is designed to not only partial out the fixed effects, but also to make all the "projection residues" $M^{(Z,Q)}\varphi_n$, $M^{(Z,Q)}\Upsilon_n$, and $M^{(Z,Q)}\Phi_n$ sufficiently small so as not to cause a bias problem even in the presence of many weak instruments. For this purpose, it is important that our specification uses $M^{(Z,Q)}$. This matrix projects into the orthogonal complement of the full set of exogenous variables/approximating functions, $(Z, Q) = (Z_1, Z_2, Q)$, and not $M^{(Z_1,Q)}$, whose use may still leave the projection residue $M^{(Z_1,Q)}\Upsilon_n$ relatively large. In addition, note that $M^{(Z,Q)}$ appears on both sides of the jackknife correction matrix $M^{(Z,Q)} D_{\widehat{\vartheta}} M^{(Z,Q)}$ so that fixed effects and nonlinear exogenous components are taken out on both sides of the (multivariate) bilinear form, not just on one side. FELIM and FEFUL are a bit more complicated than FEJIV, 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.

On the other hand, 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.

A recent paper, Evdokimov and Kolesár (2018), does examine a number of interesting jackknife IV estimators that allow for partialing out of additional covariates. In the following discussion we discuss how these estimators might perform if applied to our setting under many weak instrument asymptotics. 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). Using our notation, the estimator can be written in the form

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

It follows that we can further write the deviation of this estimator from the true value δ_0 as

$$\begin{aligned}\widehat{\delta}_{IJIVE1}-\delta_0 &= \left(X'A_{IJ1}X\right)^{-1}\left(X'A_{IJ1}\varphi_n+X'A_{IJ1}\varepsilon\right) \\ &= \left(X'A_{IJ1}X\right)^{-1}\left(X'A_{IJ1}\varphi_n+\Phi_n'A_{IJ1}\varepsilon+\Upsilon_n'A_{IJ1}\varepsilon+U'A_{IJ1}\varepsilon\right), \qquad (9)\end{aligned}$$

where $A_{IJ1}=M^{(Z_1,Q)}\left[P^{\perp}-D\left(P^{\perp}\right)\right]\left[I_{m_n}-D\left(P^{\perp}\right)\right]^{-1}M^{(Z_1,Q)}$. By straightforward calculation, it is easy to see 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 (9) above, will not be a degenerate U-statistic and will not be properly centered at zero. Another way of looking at this issue 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^{th} element of $\widehat{u}$ will contain not just the i^{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 be properly centered at zero. In some sense, the process of partialing out the covariates has interfered with the process of jackknife recentering in the way this estimator is constructed. We can use a similar argument to also show that the bilinear form for IJIVE2 is not properly centered at zero.

Now consider the UJIVE estimator, which was first introduced in Kolesár (2013) and is further analyzed in the grouped data setting by Evdokimov and Kolesár (2018). This estimator takes the form

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

where $Z = \left[\begin{array}{cc} Z_1 & Z_2 \end{array}\right]$, $\widetilde{P}^{(Z,Q)} = P^{(Z,Q)} - D\left(P^{(Z,Q)}\right)$, and $\widetilde{P}^{(Z_1,Q)} = P^{(Z_1,Q)} - D\left(P^{(Z_1,Q)}\right)$. To discuss this estimator, it is most convenient to consider the $d = 1$ case (i.e., the case where there is only one endogenous regressor). In this case, the diagonal matrix D_μ defined in Assumption 3 reduces to the scalar $\mu_n = \mu_n^{\min}$. Now, we can write the deviation of this estimator from the true value δ_0 as

$$\widehat{\delta}_{UJIVE} - \delta_0 = \left(\frac{X'A_{UJ}X}{\mu_n^2}\right)^{-1}\left(\frac{X'A_{UJ}\varphi_n + X'A_{UJ}Q\alpha + \Phi_n'A_{UJ}\varepsilon + \Upsilon_n'A_{UJ}\varepsilon + U'A_{UJ}\varepsilon}{\mu_n^2}\right),$$

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 $X'A_{UJ}\varphi_n/\mu_n^2$, 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

$$\begin{aligned}\frac{X'A_{UJ}\varphi_n}{\mu_n^2} &= \frac{\Upsilon_n'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}\varphi_n}{\mu_n^2} - \frac{\Upsilon_n'M^{(Z,Q)}D\left(M^{(Z,Q)}\right)^{-1}\varphi_n}{\mu_n^2} \\ &\quad + \frac{\Phi_n'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}\varphi_n}{\mu_n^2} - \frac{\Phi_n'M^{(Z,Q)}D\left(M^{(Z,Q)}\right)^{-1}\varphi_n}{\mu_n^2} \\ &\quad + \frac{U'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}\varphi_n}{\mu_n^2} - \frac{U'M^{(Z,Q)}D\left(M^{(Z,Q)}\right)^{-1}\varphi_n}{\mu_n^2}. \end{aligned} \tag{10}$$

Note that the term on the right-hand side of (10) which can be particularly large in order of magnitude is $\Upsilon_n'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}\varphi_n/\mu_n^2$. In fact, one can show that

$$\frac{\Upsilon_n'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}\varphi_n}{\mu_n^2} = \frac{\mu_n\Gamma_n'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}}{\sqrt{n}\mu_n^2}\frac{\tau_n g_n}{\sqrt{n}}$$

$$= \frac{\tau_n}{\mu_n}\frac{\Gamma_n'M^{(Z_1,Q)}D\left(M^{(Z_1,Q)}\right)^{-1}g_n}{n} = O_{a.s.}\left(\frac{\tau_n}{\mu_n}\right).$$

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 simulation results, given in Section 5, confirm that UJIVE tends to do much less well in terms of bias when there are included exogenous regressors that enter significantly into the structural equation of interest.

It should be noted, however, that, in the context of a linear IV model such as that given by (5) and (6), UJIVE can be shown to be consistent 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 $X'A_{UJ}\varphi_n/\mu_n^2 = \tau_n X'A_{UJ}Z_1\theta^{K_{1,n}}/\left(\mu_n^2\sqrt{n}\right)$, but also because, in a linear model with no included exogenous regressors, $\Upsilon_n' A_{UJ}Q\alpha/\mu_n^2$
$= \Pi^{K_{2,n}\prime}Z_2'\left[M^Q D\left(M^Q\right)^{-1} - M^{(Z_2,Q)}D\left(M^{(Z_2,Q)}\right)^{-1}\right]Q\alpha/\left(\mu_n\sqrt{n}\right) = 0$ so that, without the contaminating effects of the included exogenous regressors, UJIVE does properly partial out the fixed effects.

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. 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[5]. 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

[5] 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.

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 has good 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[6].

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 (2018b)[7]. 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 $\left(M^{(Z,W,Q)} \circ M^{(Z,W,Q)}\right)$ 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. Here, ones performs the minimization problem on $Q(\beta)$ in order to obtain a minimizer $\widetilde{\beta} = \begin{pmatrix} \widetilde{\beta}_1 & \widetilde{\beta}_2' \end{pmatrix}'$, with $\widetilde{\beta}_1$ 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 Supplemental 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 (7) and that also has explicit representation given by equation (8) above.

Assumption 9: Consider the variance-ratio objective function

[6] 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.pdf

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

$Q(\beta) = \left(\beta'\overline{X}'A\overline{X}\beta\right) / \left(\beta'\overline{X}'M^{(Z_1,Q)}\overline{X}\beta\right)$, 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 constant $\underline{C}$ such that

$$\left|\widetilde{\beta}_1\right| \geq \underline{C} > 0 \;\; a.s. \text{ for all } n \text{ sufficiently large.} \tag{11}$$

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 $\left\|\widetilde{\beta}\right\|_2 = 1$.

3 Consistency and Asymptotic Normality of Point Estimators

Theorem 1: Suppose that Assumptions 1-7 are satisfied. 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, as $n \to \infty$, $\left\|D_\mu\left(\overline{\delta}_n - \delta_0\right)/\mu_n^{\min}\right\|_2 \xrightarrow{p} 0$ and $\left\|\overline{\delta}_n - \delta_0\right\|_2 \xrightarrow{p} 0$.

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 1-7 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

$$\Lambda_{I,n} = H_n^{-1}\left(\Sigma_{1,n} + \Sigma_{2,n}\right)H_n^{-1} = H_n^{-1}\Sigma_n H_n^{-1}, \quad (12)$$

$$\Lambda_{II,n} = \frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}} H_n^{-1}\Sigma_{2,n}H_n^{-1}, \quad (13)$$

where $H_n = \Gamma' M^{(Z_1,Q)}\Gamma/n$, $\Sigma_{1,n} = VC\left(\Gamma' M^{(Z_1,Q)}\varepsilon/\sqrt{n}|\mathcal{F}_n^W\right) = \Gamma' M^{(Z_1,Q)} D_{\sigma^2} M^{(Z_1,Q)}\Gamma/n$, and

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

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

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

Theorem 2: Suppose that Assumptions 1-9 are satisfied. In addition, suppose that Case I holds so that $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O(1)$. Then, $\Lambda_{I,n}$ 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: Suppose that Assumptions 1-9 are satisfied, and suppose that Case II holds, so that $\left(\mu_n^{\min}\right)^2/K_{2,n} = o(1)$, but $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. In addition, let $\widetilde{L}_n$ be a $q \times d$ matrix with $1 \leq q \leq d$, and suppose that there exists a positive constant C such that $\left\|\widetilde{L}_n\right\|_2 \leq C < \infty$ and $\lambda_{\min}\left(\widetilde{L}_n \Lambda_{II,n}\widetilde{L}_n'\right) \geq 1/C > 0$ $a.s.n.$ Then,
$\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left(\widetilde{L}_n\Lambda_{II,n}\widetilde{L}_n'\right)^{-1/2}\widetilde{L}_n D_\mu\left(\widehat{\delta}_{L,n} - \delta_0\right) \xrightarrow{d} N(0, I_q)$ and
$\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left(\widetilde{L}_n\Lambda_{II,n}\widetilde{L}_n'\right)^{-1/2}\widetilde{L}_n D_\mu\left(\widehat{\delta}_{F,n} - \delta_0\right) \xrightarrow{d} N(0, I_q)$.

As alluded to earlier, the asymptotic results for FEJIV, FELIM, and FEFUL given in the above theorems can be specialized to obtain results for the linear IV regression case with (possibly) many weak instruments and/or many weak covariates, as specified in equations (5) and (6) above.

4 Covariance Matrix Estimation and Hypothesis Testing

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

$$\widehat{V}_L = \widehat{H}_L^{-1}\widehat{\Sigma}_L\widehat{H}_L^{-1} \text{ and } \widehat{V}_F = \widehat{H}_F^{-1}\widehat{\Sigma}_F\widehat{H}_F^{-1}, \tag{14}$$

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 (14) under both Cases I and II[8].

[8] 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

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

Theorem 4: Suppose that Assumptions 1-9 are satisfied. Then, the following statements are true.

(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 (12).

(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 (13).

Theorem 5 below provides asymptotic normality results for t-statistics associated with the FELIM and FEFUL estimators in the case where $\mu_{1,n} = \cdots = \mu_{d,n} = \mu_n^{\min}$. 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. In this case, we show that, without any additional side conditions that may restrict the form of the linear hypothesis tested, the t-ratio based on our estimators has an asymptotic standard normal distribution under the null hypothesis, as long as $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 = \sqrt{K_{2,n}}/\left(\mu_n\right)^2 \to 0$. Moreover, the results show that, under these same rate conditions, the tests are also consistent, as the test statistics diverge under fixed alternatives.

Theorem 5: Suppose that Assumptions 1-9 are satisfied. Suppose 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 $\mathbb{T}_F \xrightarrow{d} N(0,1)$.

---

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 5 is based on the above formula.

(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 $c'\delta_0 < r$.

Our next result considers cases where we test a null hypothesis involving only one coefficient, such as testing the significance of a particular parameter. We choose to analyze this case because this seems to be the most frequent use of the t-statistic by empirical researchers. In these cases, we establish 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 the degree of instrument weakness across the different first-stage equations. Moreover, our test will be robust to many weak instruments even if the empirical researcher using our test has no knowledge of how the degree of instrument weakness varies across the different first-stage equations. For this result, we introduce a modification of Assumption 3.

Assumption 3*: Suppose that $\Upsilon_n\left(W_{2,(i,t)}\right) = D_\mu \gamma\left(W_{2,(i,t)}\right)/\sqrt{n}$, for $(i,t) = 1, ..., m_n$, where D_μ has the form

$$D_\mu = \begin{pmatrix} \underset{d_1 \times d_1}{D_1} & 0 \\ 0 & \left(\mu_n^{\min}\right) \cdot I_{d_2} \end{pmatrix}, \tag{15}$$

with $D_1 = diag\left(\mu_{1,n}, .., \mu_{d_1,n}\right)$, and where d_1 and d_2 are positive integers, with $d_1 + d_2 = d$. The following conditions are assumed. (i) Either $\mu_{g,n} = \sqrt{n}$ or $\mu_{g,n}/\sqrt{n} \to 0$, for $g \in \{1, ..., d\}$. (ii) Let $\mu_n^{\min} = \min_{1 \leq g \leq d} \mu_{g,n}$,and suppose that $\mu_n^{\min} \to \infty$, as $n \to \infty$, such that$\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \to 0$. (iii) $\left(\mu_n^{\min}\right)/\mu_{g,n} \to 0$, as $n \to \infty$, for $g \in \{1, ..., d_1\}$. (iv) Let H_n be as defined in Assumption 3(iii) above, and suppose that there exists a positive constant C, such that $\lambda_{\max}(H_n) \leq C < \infty$ and $\lambda_{\min}(H_n) \geq 1/C > 0$ $a.s.$, for all n sufficiently large. (v) Let e_g denote a $d \times 1$ elementary vector whose g^{th} element is 1, with all other elements (or components) equal to 0. 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 $d_2 \times d$. Suppose that there exists a positive constant C_* such that $e'_g \overline{H}'_{2\cdot} \overline{H}_{2\cdot} e_g \geq C_* > 0$ $a.s.$for all n sufficiently large and for $g \in \{1, ..., d\}$.

Note that writing the matrix D_μ in the form given by equation (15) may appear to require a particular ordering of the diagonal elements $\mu_{1n}, .., \mu_{d_1,n}, \mu_n^{\min}$ of D_μ, where $\mu_n^{\min}$ is placed in the last d_2 diagonal position. However, it is easily seen that the way D_μ is specified in equation (15) does not really lead to any loss of generality. In fact, a more general D_μ matrix, where not all of the diagonal elements grow at the same rate, as $n \to \infty$, can always be put in the form given in equation (15), 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 (15), 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 ordering. Then, there exists some permutation matrix P such that $D_\mu = PD_\mu^* P'$, where D_μ is the diagonal matrix given in equation (15). 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 (15) will still apply, even if the t-statistic computed by the empirical researcher is based on some other ordering.

Here, we let $D_1 = diag\left(\mu_{1,n}, .., \mu_{d_1,n}\right)$, such that $\left(\mu_n^{\min}\right)/\mu_{g,n} \to 0$, as $n \to \infty$, for $g \in \{1, ..., d_1\}$, where d_1 and d_2 are positive integers with $d_1 + d_2 = d$. This specification excludes the case where $d_1 = 0$, or $d_2 = d$, because this case has already been covered by Theorem 5.

Theorem 6: Suppose that Assumptions 1, 2, 3^*, 4-9 are satisfied; and, in what follows, let e_g denote a $d \times 1$ elementary vector whose g^{th} element is 1, with all other elements (or components) equal to 0, and define the t-statistics $\mathbb{T}_L = \left(e_g'\widehat{\delta}_L - r\right)/\sqrt{e_g'\widehat{V}_L e_g}$ and $\mathbb{T}_F = \left(e_g'\widehat{\delta}_F - r\right)/\sqrt{e_g'\widehat{V}_F e_g}$.

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

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

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

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 any $g \in \{1, ..., d\}$.

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

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

Comparing Assumption 3* with Assumption 3, we see that the one additional side condition required for Theorem 6 is the condition placed on elements of $\overline{H}_{2\cdot}$ in part (v) of Assumption 3*.

To test hypotheses involving only the g^{th} coefficient, this condition will be violated only if the g^{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 (2020), which shows that a condition similar to $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \rightarrow 0$ is both necessary and sufficient for the existence of a consistent test. Interpreted in light of their result, the results presented in Theorems 5 and 6 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 $K_{2,n}/\left(\mu_n^{\min}\right)^2 \rightarrow 0$ (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 and 6 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 $\sqrt{K_{2,n}}/\left(\mu_n^{\min}\right)^2 \rightarrow 0$. Hence, the test statistics proposed in this paper should be useful to empirical researchers interested in testing whether an effect in a particular direction is statistically significant.

5 Monte Carlo Results

In this section, we report some Monte Carlo results based on a setup similar to that of Hausman et al. (2012), but extended to the cluster-sample/panel data setting. In particular, we consider two closely related groups of data-generating processes:

DGP 1:

$$\begin{aligned} y_{(i,t)} &= \underset{1\times 1}{\delta}\underset{1\times 1}{x_{(i,t)}} + \underset{1\times 9}{\gamma'}\underset{9\times 1}{Z_{1,(i,t)}} + \alpha_i + \varepsilon_{(i,t)}, \\ x_{(i,t)} &= \underset{1\times 1}{\pi}\underset{1\times 1}{z_{2,(i,t)}} + \underset{1\times 9}{\Phi'}\underset{9\times 1}{Z_{1,(i,t)}} + \xi_i + u_{(i,t)}. \end{aligned}$$

In all experiments that utilize this DGP, we take $\gamma = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}'$ and $\Phi = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}'$.

DGP 2:

$$
\begin{aligned}
y_{(i,t)} &= \underset{1\times 1}{\delta}\underset{1\times 1}{x_{(i,t)}} + \underset{1\times 1}{\gamma_1}\underset{1\times 1}{z_{1,(i,t)}} + \alpha_i + \varepsilon_{(i,t)}, \\
x_{(i,t)} &= \underset{1\times 1}{\pi}\underset{1\times 1}{z_{2,(i,t)}} + \underset{1\times 1}{\Phi_1}\underset{1\times 1}{z_{1,(i,t)}} + \xi_i + u_{(i,t)}.
\end{aligned}
$$

In all experiments that utilize this DGP, we take $\gamma_1 = 1$ and $\Phi_1 = 1$. Additionally, we set $n = 300$ and $T_i = 3$, for each $i \in \{1, 2, ..., 300\}$, so that $m_n = 900$. We also take $\{z_{1,(i,t)}\}_{(i,t)=1}^{900} \equiv i.i.d.N(0,1)$, $\{z_{2,(i,t)}\}_{(i,t)=1}^{900} \equiv i.i.d.N(0,1)$, and $\{u_{(i,t)}\}_{(i,t)=1}^{900} \equiv i.i.d.N(0,1)$. Moreover, $z_{1,(i,t)}$, $z_{2,(i,t)}$, and $u_{(i,t)}$ are all mutually independent. The $(i,t)^{th}$ observation of the vector of instruments is specified to be $Z_{2,(i,t)} = \begin{pmatrix} z_{2,(i,t)} & z_{2,(i,t)}^2 & z_{2,(i,t)}^3 & z_{2,(i,t)}^4 & z_{2,(i,t)}D_{(i,t),1} & \cdots \end{pmatrix.$ $\begin{matrix} \cdots & z_{2,(i,t)}D_{(i,t),5} \end{pmatrix}'$, while the $(i,t)^{th}$ observation of the vector of included exogenous regressors, or covariates, is given by $Z_{1,(i,t)} = \begin{pmatrix} z_{1,(i,t)} & z_{1,(i,t)}^2 & z_{1,(i,t)}^3 & z_{1,(i,t)}^4 & z_{1,(i,t)}D_{(i,t),1} & \cdots \end{matrix}$ $\begin{matrix} \cdots & z_{1,(i,t)}D_{(i,t),5} \end{pmatrix}'$, where $D_{(i,t),k} \in \{0,1\}$ for $k \in \{1, 2, .., 5\}$ is a binary variable such that $\Pr\left(D_{(i,t),k} = 1\right) = 1/2$, and where $\{D_{(i,t),k}\}$ is independent across both (i,t) and k. The structural disturbance, $\varepsilon_{(i,t)}$, is allowed to exhibit conditional heteroskedasticity in a manner similar to the design given in Hausman et al. (2012). In particular, under DGP1, we take

$$
\varepsilon_{(i,t)} = \rho u_{(i,t)} + \sqrt{\frac{1-\rho^2}{\phi^2 + (0.86)^4}}\left(\phi v_{1,(i,t)} + 0.86 v_{2,(i,t)}\right), \tag{16}
$$

where $v_{1,(i,t)}|Z_{1,(i,t)}, z_{2,(i,t)} \sim N\left(0, \kappa_1\left[1 + \left(\iota_9' Z_{1,(i,t)} + z_{2,(i,t)}\right)^2\right]\right)$ and $v_{2,(i,t)} \sim N\left(0, (0.86)^2\right)$. Both of these distributions are assumed to be independent across the index (i,t). Under DGP2, $\varepsilon_{(i,t)}$ has a similar structure as given in equation (16) above, except that we take $v_{1,(i,t)}|Z_{1,(i,t)}, z_{2,(i,t)} \equiv v_{1,(i,t)}|z_{1,(i,t)}, z_{2,(i,t)} \sim N\left(0, \kappa_2\left[1 + \left(z_{1,(i,t)} + z_{2,(i,t)}\right)^2\right]\right)$. Also, κ_1 and κ_2 are normalization constants chosen so that under both DGP1 and DGP2 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, under both DGP1 and DGP2, we 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, DGP 1

μ^2	$\mathcal{R}^2_{\varepsilon^2 \| z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
24	0	0.1056	0.0429	0.0417	0.3405	0.0012	0.0029	0.0157
	0.1	0.1028	0.0417	0.0400	0.3474	-0.0002	0.0007	0.0122
	0.2	0.1060	0.0423	0.0417	0.3517	0.0022	0.0083	0.0200
32	0	0.0852	0.0321	0.0298	0.2241	-0.0059	0.0009	0.0105
	0.1	0.0830	0.0273	0.0264	0.2376	-0.0131	-0.0010	0.0082
	0.2	0.0842	0.0295	0.0285	0.2453	-0.0094	0.0026	0.0118
40	0	0.0729	0.0247	0.0244	0.1622	-0.0062	-0.0004	0.0079
	0.1	0.0714	0.0245	0.0240	0.1783	-0.0084	-0.0014	0.0058
	0.2	0.0724	0.0270	0.0255	0.1802	-0.0050	0.0032	0.0104
48	0	0.0625	0.0212	0.0200	0.1269	-0.0068	0.0004	0.0065
	0.1	0.0605	0.0182	0.0176	0.1272	-0.0117	-0.0020	0.0042
	0.2	0.0605	0.0190	0.0182	0.1414	-0.0082	0.0009	0.0073

Results based on 10,000 simulations

Table 2: Nine Decile Range 0.05 to 0.95[9], DGP 1

μ^2	$\mathcal{R}^2_{\varepsilon^2 \| z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
24	0	0.5986	0.9415	0.9447	6.1610	1.5655	1.2073	1.0575
	0.1	0.6032	0.9402	0.9428	6.1080	1.5857	1.2502	1.0800
	0.2	0.5968	0.9392	0.9360	6.2265	1.5635	1.1811	1.0528
32	0	0.5386	0.7618	0.7639	5.7060	1.1133	0.9156	0.8485
	0.1	0.5367	0.7662	0.7657	6.2091	1.1017	0.9167	0.8526
	0.2	0.5492	0.7763	0.7716	5.8831	1.1172	0.9282	0.8678
40	0	0.4994	0.6591	0.6570	5.5880	0.8722	0.7608	0.7229
	0.1	0.4961	0.6412	0.6415	5.1816	0.8585	0.7572	0.7170
	0.2	0.4970	0.6528	0.6546	5.5664	0.8550	0.7454	0.7109
48	0	0.4608	0.5790	0.5764	4.7318	0.7231	0.6490	0.6237
	0.1	0.4586	0.5822	0.5822	5.0871	0.7322	0.6516	0.6288
	0.2	0.4700	0.5873	0.5855	5.0922	0.7309	0.6636	0.6389

Results based on 10,000 simulations

[9] By nine decile range we mean the range between the 0.05 and the 0.95 quantiles.

Table 3: 0.05 Rejection Frequencies[10], DGP 1								
μ^2	$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
24	0	0.1861	0.1056	0.0961	0.5167	0.0322	0.0559	0.0588
	0.1	0.1809	0.1013	0.0920	0.5333	0.0307	0.0583	0.0602
	0.2	0.1894	0.1038	0.0940	0.5318	0.0317	0.0579	0.0610
32	0	0.1716	0.1077	0.0962	0.5281	0.0313	0.0486	0.0515
	0.1	0.1678	0.1061	0.0974	0.5307	0.0367	0.0556	0.0587
	0.2	0.1767	0.1124	0.1033	0.5342	0.0373	0.0581	0.0620
40	0	0.1662	0.1143	0.1055	0.5390	0.0371	0.0510	0.0536
	0.1	0.1600	0.1084	0.0987	0.5573	0.0371	0.0519	0.0553
	0.2	0.1643	0.1129	0.1033	0.5469	0.0387	0.0521	0.0555
48	0	0.1551	0.1166	0.1039	0.5749	0.0348	0.0479	0.0508
	0.1	0.1542	0.1126	0.1033	0.5695	0.0399	0.0531	0.0555
	0.2	0.1643	0.1200	0.1102	0.5733	0.0409	0.0567	0.0603

Results based on 10,000 simulations

Table 4: Median Bias, DGP 2								
μ^2	$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
24	0	0.1056	0.0429	0.0417	-0.0095	0.0012	0.0029	0.0157
	0.1	0.1030	0.0403	0.0397	-0.0110	-0.0037	0.0034	0.0155
	0.2	0.1076	0.0457	0.0445	-0.0047	0.0051	0.0135	0.0258
32	0	0.0852	0.0321	0.0298	-0.0120	-0.0059	0.0009	0.0105
	0.1	0.0837	0.0285	0.0281	-0.0137	-0.0102	-0.0007	0.0087
	0.2	0.0869	0.0315	0.0305	-0.0138	-0.0079	0.0058	0.0156
40	0	0.0729	0.0247	0.0244	-0.0135	-0.0062	-0.0004	0.0079
	0.1	0.0715	0.0255	0.0246	-0.0137	-0.0069	-0.0002	0.0073
	0.2	0.0752	0.0277	0.0271	-0.0077	-0.0030	0.0082	0.0148
48	0	0.0625	0.0212	0.0200	-0.0096	-0.0068	0.0004	0.0065
	0.1	0.0603	0.0196	0.0185	-0.0108	-0.0098	-0.0013	0.0048
	0.2	0.0635	0.0208	0.0198	-0.0111	-0.0089	0.0026	0.0085

Results based on 10,000 simulations

[10] 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, DGP 2								
μ^2	$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
24	0	0.5986	0.9415	0.9447	1.5682	1.5655	1.2073	1.0575
	0.1	0.6153	0.9693	0.9704	1.6082	1.6216	1.3049	1.1267
	0.2	0.6399	1.0004	0.9935	1.6862	1.6536	1.3127	1.1508
32	0	0.5386	0.7618	0.7639	1.0812	1.1133	0.9156	0.8485
	0.1	0.5574	0.7975	0.7949	1.1461	1.1600	0.9613	0.8939
	0.2	0.5821	0.8192	0.8103	1.1639	1.1644	1.0028	0.9361
40	0	0.4994	0.6591	0.6570	0.8578	0.8722	0.7608	0.7229
	0.1	0.5119	0.6670	0.6634	0.8851	0.8864	0.7891	0.7481
	0.2	0.5301	0.6920	0.6876	0.9098	0.9150	0.8068	0.7691
48	0	0.4608	0.5790	0.5764	0.7159	0.7231	0.6490	0.6237
	0.1	0.4762	0.6038	0.6019	0.7554	0.7649	0.6822	0.6560
	0.2	0.4995	0.6237	0.6216	0.7731	0.7742	0.7101	0.6838

Results based on 10,000 simulations

Table 6: 0.05 Rejection Frequencies, DGP 2								
μ^2	$\mathcal{R}^2_{\varepsilon^2\|z_1^2}$	2SLS	IJIVE1	IJIVE2	UJIVE	FEJIV	FELIM	FEFUL
24	0	0.1861	0.1056	0.0961	0.2236	0.0322	0.0559	0.0588
	0.1	0.1783	0.1015	0.0924	0.2243	0.0304	0.0627	0.0646
	0.2	0.1786	0.1018	0.0927	0.2306	0.0313	0.0631	0.0657
32	0	0.1716	0.1077	0.0962	0.2563	0.0313	0.0486	0.0515
	0.1	0.1718	0.1098	0.0992	0.2575	0.0356	0.0547	0.0579
	0.2	0.1739	0.1155	0.1039	0.2607	0.0380	0.0600	0.0627
40	0	0.1662	0.1143	0.1055	0.2770	0.0371	0.0510	0.0536
	0.1	0.1600	0.1133	0.1029	0.2767	0.0380	0.0558	0.0590
	0.2	0.1608	0.1167	0.1063	0.2851	0.0389	0.0578	0.0609
48	0	0.1551	0.1166	0.1039	0.2958	0.0348	0.0479	0.0508
	0.1	0.1579	0.1173	0.1057	0.2960	0.0384	0.0524	0.0555
	0.2	0.1614	0.1240	0.1121	0.2988	0.0440	0.0605	0.0616

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 almost uniformly better than 2SLS, IJIVE1, and IJIVE2, although our experiments do show the latter three to be less dispersed than the three estimators studied in this paper. Comparing FELIM and FEFUL in terms of the nine decile range, we see that FEFUL is 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. The 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. Finally, we note that UJIVE did much better under DGP2 than under DGP1. This is due to the fact that UJIVE does not properly partial out the included exogenous regressors; hence, it performs less well under DGP1, where a larger number of included exogenous regressors enter significantly into the structural equation of interest.

6 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 shown to be consistent in 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 to be consistent under many weak instrument asymptotics 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 (2020). 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] Bekker, P. A. (1994): Alternative Approximations to the Distributions of Instrumental Variables Estimators, Econometrica, 62, 657-681.

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

[5] 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.

[6] Cattaneo, M. D., M. Jansson, and W.K. Newey (2018a): Alternative Asymptotics and the Partially Linear Model with Many Regressors, Econometric Theory, 34, 277-301.

[7] Cattaneo, M. D., M. Jansson, and W.K. Newey (2018b): 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] Kolesár, M. (2013): Estimation in Instrumental Variables Models with Treatment Effect Heterogeneity, Unpublished Manuscript, Princeton University.

[20] Mikusheva, A. and L. Sun (2020): Inference with Many Weak Instruments, Unpublished Manuscript, MIT.

[21] 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.

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

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

[24] 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.

[25] 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.

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

[27] 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.

7 Appendix: Proofs of Main Theorems and Other Key Results

This appendix provides the proofs for Theorem 1, Corollary 1, and Theorems 4-6 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.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.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 $H_n = \Gamma' M^{(Z_1,Q)}\Gamma/n = O_p(1)$. 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

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

Moreover, by applying part (a) of Lemma S2-4 and part (a) of Lemma S2-5, we obtain

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

$$= O_p\left(\frac{\tau_n}{\left[\mu_n^{\min}\right] K_{1,n}^{\varrho_g}}\right) + o_p\left(\frac{\left[\mu_n^{\min}\right]\tau_n}{nK_{1,n}^{\varrho_g}}\right) = o_p\left(1\right).$$

Applying part (b) of Lemma S2-4 and part (b) of Lemma S2-5, we get

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

It follows by the triangle inequality and the Slutsky's Theorem that $\left\|D_\mu\left(\overline{\delta}_n - \delta_0\right) / \left(\mu_n^{\min}\right)\right\|_2 = o_p\left(1\right)$, 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 $\overline{\ell}_n = o_p\left(\left[\mu_n^{\min}\right]^2 / n\right) = o_p\left(1\right)$ 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 $\overline{\ell}_n = \widehat{\ell}_{F,n}$ $= \left[\widehat{\ell}_{L,n} - \left(1 - \widehat{\ell}_{L,n}\right)\left(C/m_n\right)\right] / \left[1 - \left(1 - \widehat{\ell}_{L,n}\right)\left(C/m_n\right)\right]$. 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 $S_{L,1} = X'AD\left(J\left[\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right]\right)AX$, $S_{L,2} = \left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)' J\left(A \circ A\right) J\left(\widehat{\varepsilon}_L \iota_d' \circ M^{(Z,Q)}X\right)$, $S_{L,3} = \left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)' J\left(A \circ A\right) J\left(\widehat{\varepsilon}_L \circ \widehat{\varepsilon}_L\right)$, $\underline{S}_{L,4} = \left(\widehat{\varepsilon}_L \iota_d' \circ \widehat{\underline{U}}_L\right)' J\left(A \circ A\right) J\left(\widehat{\varepsilon}_L \iota_d' \circ \widehat{\underline{U}}_L\right)$, $\widehat{H}_L = X'\left[A - \widehat{\ell}_{L,n} M^{(Z_1,Q)}\right]X$, $\Sigma_{1,n} = \Gamma' M^{(Z_1,Q)} D_{\sigma^2} M^{(Z_1,Q)}\Gamma/n$. In addition, also define $\sigma^2_{(i,t)} = E\left[\varepsilon^2_{(i,t)}|\mathcal{F}_n^W\right]$, $\phi_{(i,t)} = E\left[U_{(i,t)}\varepsilon_{(i,t)}|\mathcal{F}_n^W\right]$, $\Psi_{(i,t)} = E\left[U_{(i,t)}U'_{(i,t)}|\mathcal{F}_n^W\right]$, $\underline{\phi}_{(i,t)} = E\left[\underline{U}_{(i,t)}\varepsilon_{(i,t)}|\mathcal{F}_n^W\right]$, and $\underline{\Psi}_{(i,t)} = E\left[\underline{U}_{(i,t)}\underline{U}'_{(i,t)}|\mathcal{F}_n^W\right]$ 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 $\mathcal{F}_n^W = \sigma\left(W_n\right)$.

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,

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

from which we deduce that $D_\mu^{-1} X' A D\left(\varepsilon \circ \varepsilon\right) A X D_\mu^{-1} = O_p\left(1\right)$ using Assumptions 2(i) and 3(iii), Lemma S2-1 part (a), and the assumption that $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O\left(1\right)$ 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 $D_\mu^{-1}\widehat{H}_L D_\mu^{-1} = H_n + o_p\left(1\right)$, where, by Assumption 3(iii), $H_n = \Gamma' M^{(Z_1,Q)}\Gamma/n$ is positive definite $a.s.n.$ In addition, we can apply part (a) of Lemma S2-18 and Slutsky's theorem to deduce that

$$\begin{aligned}
\widehat{V}_{L,1} &= \left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} D_\mu^{-1} S_{L,1} D_\mu^{-1}\left(D_\mu^{-1}\widehat{H}_L D_\mu^{-1}\right)^{-1} \\
&= H_n^{-1}\Sigma_{1,n} H_n^{-1} + H_n^{-1}\left(\sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A_{(i,t),(j,s)}^2 \sigma_{(i,t)}^2 D_\mu^{-1}\Psi_{(j,s)} D_\mu^{-1}\right) H_n^{-1} + o_p\left(1\right). \quad (17)
\end{aligned}$$

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 $K_{2,n}/\left(\mu_n^{\min}\right)^2 = O\left(1\right)$ 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

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

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

$$\widehat{V}_{L,2} = -H_n^{-1} \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} D_\mu^{-1} \left(\rho \sigma^2_{(i,t)} \phi'_{(j,s)} + \sigma^2_{(i,t)} \phi_{(j,s)} \rho' \right) D_\mu^{-1} H_n^{-1} + o_p(1) \tag{18}$$

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

$$\begin{aligned} \widehat{V}_{L,3} &= K_{2,n} H_n^{-1} \left[D_\mu^{-1} \rho + D_\mu^{-1} (\widehat{\rho}_L - \rho) \right] \frac{S_{L,3}}{K_{2,n}} \rho' D_\mu^{-1} H_n^{-1} (1 + o_p(1)) \\ &\quad + K_{2,n} H_n^{-1} \left[D_\mu^{-1} \rho + D_\mu^{-1} (\widehat{\rho}_L - \rho) \right] \frac{S_{L,3}}{K_{2,n}} (\widehat{\rho}_L - \rho)' D_\mu^{-1} H_n^{-1} (1 + o_p(1)) \\ &= H_n^{-1} D_\mu^{-1} \rho \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} \sigma^2_{(i,t)} \sigma^2_{(j,s)} \rho' D_\mu^{-1} H_n^{-1} + o_p(1). \end{aligned} \tag{19}$$

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

$$\begin{aligned} \widehat{V}_{L,4} &= H_n^{-1} D_\mu^{-1} \underline{S}_{L,4} D_\mu^{-1} H_n^{-1} (1 + o_p(1)) \\ &= H_n^{-1} \sum_{\substack{(i,t),(j,s)=1 \\ (i,t)\neq(j,s)}}^{m_n} A^2_{(i,t),(j,s)} D_\mu^{-1} \underline{\phi}_{(i,t)} \underline{\phi}'_{(j,s)} D_\mu^{-1} H_n^{-1} + o_p(1). \end{aligned} \tag{20}$$

It follows from equations (17), (18), (19), and (20) that

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

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 $D_\mu \widehat{V}_F D_\mu = H_n^{-1}\left(\Sigma_{1,n}+\Sigma_{2,n}\right)H_n^{-1}+o_p(1)=\Lambda_{I,n}+o_p(1)$.

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,

$$
\begin{aligned}
\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\widehat{V}_{L,1} &= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\left[\Sigma_{1,n}+\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n}A_{(i,t),(j,s)}^2\sigma_{(i,t)}^2D_\mu^{-1}\Psi_{(j,s)}D_\mu^{-1}\right]H_n^{-1}\left(1+o_p(1)\right)\\
&= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n}A_{(i,t),(j,s)}^2\sigma_{(i,t)}^2D_\mu^{-1}\Psi_{(j,s)}D_\mu^{-1}H_n^{-1}+o_p(1). \qquad (21)
\end{aligned}
$$

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

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

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

$$
\frac{\left(\mu_n^{\min}\right)^2\widehat{V}_{L,2}}{K_{2,n}}=-H_n^{-1}\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n}\frac{\left(\mu_n^{\min}\right)^2A_{(i,t),(j,s)}^2}{K_{2,n}}D_\mu^{-1}\left(\rho\sigma_{(i,t)}^2\phi_{(j,s)}'+\phi_{(j,s)}\sigma_{(i,t)}^2\rho'\right)D_\mu^{-1}H_n^{-1}+o_p(1). \qquad (22)
$$

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,

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

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,

$$\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\widehat{V}_{L,4} = H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}D_\mu^{-1}\underline{S}_{L,4}D_\mu^{-1}H_n^{-1}\left(1+o_p\left(1\right)\right)$$

$$= H_n^{-1}\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}\sum_{\substack{(i,t),(j,s)=1\\(i,t)\neq(j,s)}}^{m_n}A^2_{(i,t),(j,s)}D_\mu^{-1}\underline{\phi}_{(i,t)}\underline{\phi}'_{(j,s)}D_\mu^{-1}H_n^{-1}+o_p\left(1\right). \qquad (24)$$

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

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

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 $\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]D_\mu\widehat{V}_F D_\mu =$
$\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]H_n^{-1}\Sigma_{2,n}H_n^{-1}+o_p\left(1\right)=\Lambda_{II,n}+o_p\left(1\right)$. □

Proof of Theorem 5:

To show part (a), first note that since $\mu_{1,n}=\cdots=\mu_{d,n}=\mu_n^{\min}$ here, we can take $\mu_n^{\min}=\mu_n$. Moreover, by part (d) of Lemma S2-3 and Assumption 3(iii), $\Lambda_{I,n}$ is positive definite *a.s.n.* In addition, specializing the result of part (a) of Theorem 4 to this case, we have $D_\mu\widehat{V}_L D_\mu=\mu_n^2\widehat{V}_L=$

$\Lambda_{I,n} + o_p(1)$, so that $\mu_n^2 \widehat{V}_L$ is positive definite w.p.a.1. Hence, under $H_0 : c'\delta_0 = r$, we can write

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

Now, specializing the result of Theorem 2 to this case, we have $\Lambda_{I,n}^{-1/2} D_\mu \left(\widehat{\delta}_{L,n} - \delta_0\right)$ $= \mu_n \Lambda_{I,n}^{-1/2}\left(\widehat{\delta}_{L,n} - \delta_0\right) \xrightarrow{d} N(0, I_d)$. It follows by the continuous mapping theorem that

$$\mathbb{T}_L = \frac{c'\Lambda_{I,n}^{1/2}\left[\mu_n \Lambda_{I,n}^{-1/2}\left(\widehat{\delta}_{L,n} - \delta_0\right)\right]}{\sqrt{c'\Lambda_{I,n} c}}\left[1 + o_p(1)\right] \xrightarrow{d} N(0,1). \tag{25}$$

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 (25) above, whereas application of part (a) of Theorem 4 and Slutsky's theorem shows that $\mu_n^2 c'\widehat{V}_L c = c'\Lambda_{I,n} c + o_p(1)$, where $c'\Lambda_{I,n} c > 0$ for all $c \neq 0$ *a.s.n.* in light of part (d) of Lemma S2-3 and Assumption 3(iii). In addition, by parts (a) and (c) of Lemma S2-3; Assumption 3(iii); and the fact that, under Case I, $K_{2,n}/\left(\mu_n^{\min}\right)^2 = K_{2,n}/\mu_n^2 = O(1)$; there exists a positive constant $C < \infty$ such that, almost surely for all n sufficiently large,

$$\lambda_{\max}\left(\Lambda_{I,n}\right) \leq \frac{\lambda_{\max}\left[VC\left(\Gamma' M^{(Z_1,Q)}\varepsilon/\sqrt{n}\right)|\mathcal{F}_n^W\right] + \frac{K_{2,n}}{\mu_n^2}\lambda_{\max}\left[VC\left(\underline{U}'A\varepsilon/\sqrt{K_{2,n}}\right)|\mathcal{F}_n^W\right]}{\left[\lambda_{\min}\left(H_n\right)\right]^2} \leq C. \tag{26}$$

It follows that, in this case, $h/\sqrt{c'\widehat{V}_L c} = \mu_n h/\sqrt{c'\mu_n^2 \widehat{V}_L c} = \left(\mu_n h/\sqrt{c'\Lambda_{I,n} c}\right)\left[1 + o_p(1)\right]$. 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), note that, setting $\widetilde{L}_n = c'$ and $D_\mu = \mu_n \cdot I_d$ in Theorem 3, we have $\left(\mu_n/\sqrt{K_{2,n}}\right)\left(c'\Lambda_{II,n} c\right)^{-1/2} c'\left[\mu_n\left(\widehat{\delta}_{L,n} - \delta_0\right)\right] \xrightarrow{d} N(0,1)$. Moreover, part (b) of Theorem 4 implies that $\left(\mu_n^2/K_{2,n}\right) D_\mu \widehat{V}_L D_\mu = \left(\mu_n^4/K_{2,n}\right)\widehat{V}_L = \Lambda_{II,n} + o_p(1)$. It follows that, under $H_0 : c'\delta_0 = r$,

$$\mathbb{T}_L = \frac{\left(\mu_n/\sqrt{K_{2,n}}\right) c'\left[\mu_n\left(\widehat{\delta}_{L,n} - \delta_0\right)\right]}{\sqrt{c'\left(\mu_n^4/K_{2,n}\right)\widehat{V}_L c}} = \frac{\left(\mu_n/\sqrt{K_{2,n}}\right) c'\left[\mu_n\left(\widehat{\delta}_{L,n} - \delta_0\right)\right]}{\sqrt{c'\Lambda_{II,n} c}}\left[1 + o_p(1)\right] \xrightarrow{d} N(0,1). \tag{27}$$

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 part (b) of Theorem 4 and Slutsky's theorem, we have $\left(\mu_n^4/K_{2,n}\right) c'\widehat{V}_L c = c'\Lambda_{II,n} c + o_p(1)$.

Moreover, there exists a positive constant $\underline{C}$ such that $c'\Lambda_{II,n}c = \mu_n^2 c' H_n^{-1}\Sigma_{2,n}H_n^{-1}c/K_{2,n} = c'H_n^{-1}VC\left(\underline{U}'A\varepsilon/\sqrt{K_{2,n}}|\mathcal{F}_n^W\right)H_n^{-1}c \geq \underline{C} > 0$ $a.s.n.$ for all $c \neq 0$, by the almost sure positive definiteness of $VC\left(\underline{U}'A\varepsilon/\sqrt{K_{2,n}}|\mathcal{F}_n^W\right)$ as shown in part (b) of Lemma S2-3. 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

$$\lambda_{\max}\left(\Lambda_{II,n}\right) \leq \frac{\mu_n^2}{K_{2,n}}\frac{1}{\left[\lambda_{\min}\left(H_n\right)\right]^2}\frac{K_{2,n}}{\mu_n^2}\lambda_{\max}\left[VC\left(\frac{\underline{U}'A\varepsilon}{\sqrt{K_{2,n}}}\right)|\mathcal{F}_n^W\right] \leq C < \infty. \tag{28}$$

It follows that, for this case,

$$\frac{h}{\sqrt{c'\widehat{V}_L c}} = \frac{\left(\mu_n^2/\sqrt{K_{2,n}}\right)h}{\sqrt{\left(\mu_n^4/K_{2,n}\right)c'\widehat{V}_L c}} = \frac{\left(\mu_n^2/\sqrt{K_{2,n}}\right)h}{\sqrt{c'\Lambda_{II,n}c}}\left[1 + o_p\left(1\right)\right].$$

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 $\mu_n^2/\sqrt{K_{2,n}} \to \infty$. 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 $O_p\left(1\right)$ as shown in (27), 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 Theorem 6:

To show part (a), note first that, by part (d) of Lemma S2-3 and Assumption 3*(iv), $\Lambda_{I,n}$ is positive definite $a.s.n.$ Hence, under $H_0 : e_g'\delta_0 = r$, we can write

$$\mathbb{T}_L = \frac{e_g'\widehat{\delta}_L - r}{\sqrt{e_g'\widehat{V}_L e_g}} = \frac{e_g'\left(\mu_{g,n}\right)D_\mu^{-1}\Lambda_{I,n}^{1/2}\left[\Lambda_{I,n}^{-1/2}D_\mu\left(\widehat{\delta}_L - \delta_0\right)\right]}{\sqrt{e_g'\left(\mu_{g,n}\right)D_\mu^{-1}D_\mu\widehat{V}_L D_\mu D_\mu^{-1}\left(\mu_{g,n}\right)e_g}} = \frac{e_g'\Lambda_{I,n}^{1/2}\left[\Lambda_{I,n}^{-1/2}D_\mu\left(\widehat{\delta}_L - \delta_0\right)\right]}{\sqrt{e_g' D_\mu\widehat{V}_L D_\mu e_g}} \tag{29}$$

where the last equality follows from the fact that
$e_g'\left(\mu_{g,n}\right)D_\mu^{-1} = \left(\mu_{g,n}\right)e_g'\begin{bmatrix}\left(\mu_{1,n}\right)^{-1}e_1 & \cdots & \left(\mu_{d,n}\right)^{-1}e_d\end{bmatrix} = e_g'$. Now, applying the result of Theorem 2 to this case, we have $\Lambda_{I,n}^{-1/2}D_\mu\left(\widehat{\delta}_L - \delta_0\right) \xrightarrow{d} N\left(0, I_d\right)$. In addition, applying the result of part (a) of Theorem 4 to this case, we have $D_\mu\widehat{V}_L D_\mu = \Lambda_{I,n} + o_p\left(1\right)$. Let $a_n' = e_g'\Lambda_{I,n}^{1/2}/\sqrt{e_g'\Lambda_{I,n}e_g}$, and note that $a_n' a_n = 1$. It follows from these intermediate results and the continuous mapping

theorem that $\mathbb{T}_L = a_n' \Lambda_{I,n}^{-1/2} D_\mu \left(\widehat{\delta}_L - \delta_0\right)[1 + o_p(1)] \xrightarrow{d} N(0,1)$. On the other hand, under H_1, we can take $e_g'\delta_0 = r + h$ for some $h \in \mathbb{R}\backslash\{0\}$. Write

$$\mathbb{T}_L = \frac{e_g'\left(\widehat{\delta}_L - \delta_0\right)}{\sqrt{e_g'\widehat{V}_L e_g}} + \frac{(\mu_{g,n})\, h}{\sqrt{e_g'(\mu_{g,n})\, D_\mu^{-1} D_\mu \widehat{V}_L D_\mu D_\mu^{-1} (\mu_{g,n})\, e_g}} = \frac{e_g'\left(\widehat{\delta}_L - \delta_0\right)}{\sqrt{e_g'\widehat{V}_L e_g}} + \frac{(\mu_{g,n})\, h}{\sqrt{e_g' D_\mu \widehat{V}_L D_\mu e_g}},$$

where the last line again follows from the identity $e_g'(\mu_{g,n})\, D_\mu^{-1} = e_g'$, as shown previously. Now, the first term on the right-hand side of the last equality above has previously been shown to converge to a $N(0,1)$ distribution so that, in particular, $e_g'\left(\widehat{\delta}_L - \delta_0\right)/\sqrt{e_g'\widehat{V}_L e_g} = O_p(1)$. Moreover, application of part (a) of Theorem 4 and Slutsky's theorem shows that $e_g' D_\mu \widehat{V}_L D_\mu e_g = e_g'\Lambda_{I,n} e_g + o_p(1)$, where $e_g'\Lambda_{I,n} e_g > 0$ *a.s.n.* since $\Lambda_{I,n}$ is positive definite *a.s.n.* by part (d) of Lemma S2-3 and Assumption 3*(iv). In addition, by parts (a) and (c) of Lemma S2-3, Assumption 3*(iv), and the fact that $K_{2,n}/\left(\mu_n^{\min}\right)^2 = K_{2,n}/\mu_n^2 = O(1)$ under Case I, there exists a positive constant C such that, almost surely for all n sufficiently large, $\lambda_{\max}(\Lambda_{I,n}) \leq C < \infty$, as can be shown by following an argument similar to that given previously in obtaining expression (26) in the proof of Theorem 5. Hence, in this case, $h/\sqrt{e_g'\widehat{V}_L e_g} = (\mu_{g,n})\, h/\sqrt{e_g' D_\mu \widehat{V}_L D_\mu e_g} = \left((\mu_{g,n})\, h/\sqrt{e_g'\Lambda_{I,n} e_g}\right)[1 + o_p(1)]$. Given that $\mu_{g,n} \to \infty$ as $n \to \infty$, w.p.a.1, $h/\sqrt{e_g'\widehat{V}_L e_g} \to +\infty$ if $h > 0$, whereas $h/\sqrt{e_g'\widehat{V}_L e_g} \to -\infty$ if $h < 0$, from which the stated result follows. 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), note first that

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

Moreover, note that there exists a positive constant C such that $e_g'\Lambda_{II,n} e_g \geq C > 0$ w.p.a.1 as $n \to \infty$ since

$$\begin{aligned} e_g'\Lambda_{II,n} e_g &= e_g' H_n^{-1}\left(\mu_n^{\min}\right) D_\mu^{-1} VC\left(\frac{\underline{U}' A\varepsilon}{\sqrt{K_{2,n}}} | \mathcal{F}_n^W\right) D_\mu^{-1}\left(\mu_n^{\min}\right) H_n^{-1} e_g \\ &= e_g' H_n^{-1} D_0 VC\left(\frac{\underline{U}' A\varepsilon}{\sqrt{K_{2,n}}} | \mathcal{F}_n^W\right) D_0 H_n^{-1} e_g [1 + o_p(1)] \end{aligned}$$

and since, by applying the result of part (b) of Lemma S2-3 and Assumption 3*(v), we have $e_g' H_n^{-1} D_0 V\left(\underline{U}' A\varepsilon/\sqrt{K_{2,n}} | \mathcal{F}_n^W\right) D_0 H_n^{-1} e_g \geq \underline{C} e_g' \overline{H}_{2\cdot}' \overline{H}_{2\cdot} e_g \geq \underline{C} C_* = C > 0$ *a.s.n.* for $\overline{H}_{2\cdot}$ as defined in Assumption 3*(v) and for positive constants $\underline{C}$ (defined in Lemma S2-3), C_* (defined in Assumption 3*(v)), and $C = \underline{C} C_*$. Now, setting $\widetilde{L}_n = e_g'$ in Theorem 3, we have

$\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left(e_g'\Lambda_{II,n}e_g\right)^{-1/2}e_g'D_\mu\left(\widehat{\delta}_L-\delta_0\right)\xrightarrow{d}N(0,1)$, and, by applying part (b) of Theorem 4 and Slutsky's theorem, we also obtain $\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]e_g'D_\mu\widehat{V}_LD_\mu e_g=e_g'\Lambda_{II,n}e_g+o_p(1)$. Hence, under $H_0: e_g'\delta_0=r$, we can apply the identity $e_g'\left(\mu_{g,n}\right)D_\mu^{-1}=e_g'$ to obtain

$$\begin{aligned}\mathbb{T}_L &= \frac{\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left[e_g'\left(\mu_{g,n}\right)D_\mu^{-1}D_\mu\left(\widehat{\delta}_L-\delta_0\right)\right]}{\sqrt{\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]e_g'\left(\mu_{g,n}\right)D_\mu^{-1}D_\mu\widehat{V}_LD_\mu D_\mu^{-1}\left(\mu_{g,n}\right)e_g}}=\frac{\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left[e_g'D_\mu\left(\widehat{\delta}_L-\delta_0\right)\right]}{\sqrt{\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]e_g'D_\mu\widehat{V}_LD_\mu e_g}}\\ &= \frac{\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left[e_g'D_\mu\left(\widehat{\delta}_L-\delta_0\right)\right]}{\sqrt{e_g'\Lambda_{II,n}e_g}}\left[1+o_p(1)\right]\xrightarrow{d}N(0,1).\end{aligned}$$

Under H_1, write $e_g'\delta_0=r+h$ for some $h\in\mathbb{R}\backslash\{0\}$. As shown above, $\left[\left(\mu_n^{\min}\right)^2/K_{2,n}\right]e_g'D_\mu\widehat{V}_LD_\mu e_g=e_g'\Lambda_{II,n}e_g+o_p(1)$, where there exists a positive constant C such that $e_g'\Lambda_{II,n}e_g\geq C>0$ w.p.a.1 as $n\to\infty$. In addition, by part (c) of Lemma S2-3 and Assumption 3^*(iv), there exists a positive constant C such that, almost surely for all n sufficiently large, $\lambda_{\max}\left(\Lambda_{II,n}\right)\leq C<\infty$, as can be shown by following an argument similar to that given previously in obtaining expression (28) in the proof of Theorem 5. It follows from these results and from making use of the identity $e_g'\left(\mu_{g,n}\right)D_\mu^{-1}=e_g'$ that

$$\frac{h}{\sqrt{e_g'\widehat{V}_Le_g}}=\frac{\left(\mu_n^{\min}/\sqrt{K_{2,n}}\right)\left(\mu_{g,n}\right)h}{\sqrt{\frac{\left(\mu_n^{\min}\right)^2}{K_{2,n}}e_g'\left(\mu_{g,n}\right)D_\mu^{-1}D_\mu\widehat{V}_LD_\mu D_\mu^{-1}\left(\mu_{g,n}\right)e_g}}=\frac{\mu_n^{\min}}{\sqrt{K_{2,n}}}\frac{\left(\mu_{g,n}\right)h}{\sqrt{e_g'\Lambda_{II,n}e_g}}\left[1+o_p(1)\right].$$

Thus, w.p.a.1, $h/\sqrt{e_g'\widehat{V}_Le_g}\to+\infty$ if $h>0$ whereas $h/\sqrt{e_g'\widehat{V}_Le_g}\to-\infty$ if $h<0$, given that $\left(\mu_n^{\min}\right)^2/\sqrt{K_{2,n}}\to\infty$ and $\mu_n^{\min}/\mu_{g,n}=O(1)$ for any $g\in\{1,...,d\}$. Finally, write

$$\mathbb{T}_L=\frac{e_g'\widehat{\delta}_L-r}{\sqrt{e_g'\widehat{V}_Le_g}}=\frac{e_g'\left(\widehat{\delta}_L-\delta_0\right)}{\sqrt{e_g'\widehat{V}_Le_g}}+\frac{h}{\sqrt{e_g'\widehat{V}_Le_g}}. \tag{30}$$

Since the first term on the right-hand side of equation (30) is $O_p(1)$, as shown above, we deduce that w.p.a.1., $\mathbb{T}_L\to+\infty$ if $h>0$ and $\mathbb{T}_L\to-\infty$ if $h<0$. Finally, the results for $\mathbb{T}_F$ can be shown in the same way, so to avoid redundancy, we omit the proof. □