MANY (WEAK) JUDGES IN JUDGE-LENIENCY DESIGNS

Koen Jochmans*

Toulouse School of Economics

This version: October 10, 2023

Abstract

Judge-lenciency designs are very popular. Evaluating whether conventional inference procedures apply to it is not immediate. We frame such designs as an inference problem from grouped data in a setting with a growing number of groups and limited variation between groups. Such an asymptotic approximation is well suited for the data sets encountered in practice. The two-stage least-squares estimator should never be used. The jackknife instrumental-variable estimator can present a reliable tool for inference, provided that a non-standard asymptotic-variance estimator is used along with it. Conventional decision rules to gauge instrument strength are typically not valid in our setting. An alternative such decision rule is provided and is found to perform well.

JEL Classification: C23, C26

Keywords: bias, examiner design, fixed effects, inference, jackknife, weak instruments

In an influential study, Kling (2006) exploited the random assignment of court cases to judges to estimate the effect of the duration of incarceration spells on employment and earnings prospects of individuals upon their release from prison. The underlying idea is that variation in the severity of judges induces exogenous variation in incarceration length that can be leveraged via an instrumental-variable strategy. This so-called judge-leniency

*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: koen.jochmans@tse-fr.eu.

Funded by the European Union (ERC, NETWORK, 101044319) and by the French Government and the French National Research Agency under the Investissements d' Avenir program (ANR-17-EURE-0010). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them

design (or examiner design) extends naturally to a variety of other settings and has become very popular in empirical work.[1]

The leniency of a judge is not unobserved in the data. In a two-stage least-squares procedure it is estimated by (in the context of Kling 2006) the average incarceration length across all cases of a given judge, i.e., as a judge fixed effect. This first-stage estimation introduces bias. Several studies have used versions of a jackknife instrumental-variable estimator. However, the few papers that spell out which asymptotic scheme they use to perform inference tend to work under asymptotics where the number of judges is fixed (a recent example is Frandsen, Lefgren and Leslie 2023). Under such a framework both estimators have the same asymptotic distribution, and so such a limit theory does not provide an argument for the use of the jackknife estimator.

It is fruitful to frame the judge-leniency design as a problem concerning data with a group structure. With n judges and (say) m cases per judge the data can be partitioned into n non-overlapping groups of size m. Below we show that the bias in the two-stage least-squares estimator invalidates inference based on it unless, at a minimum, $n/m \to 0$. This rate is familiar from the panel-data literature, although it usually shows up there in the context of nonlinear problems (Hahn and Newey 2004, Dhaene and Jochmans 2015). Kling's (2006) data contain 4609 cases distributed over 52 judges, yielding an average of 88 cases per judge. Similarly, in the data of Bhuller, Dahl, Loken and Mogstad (2020) each of 500 judges handled an average of 258 cases while in the study of González-Uribe and Reyes (2021) 50 judges were involved in the evaluation of 135 grant applications, yielding an average of less than 3 cases per judge. These examples suggest that working under an approximation based on asymptotics where n/m is bounded away from zero (and, possibly, even diverges to infinity) is more suitable.

Not treating the number of judges as fixed leads to asymptotics with a growing number of instruments (as first considered in Kunitomo 1980, Morimune 1983, and Bekker 1994).

[1] Just a few examples of different applications are in Maestas, Mullen and Strand (2013), Aizer and Doyle Jr. (2015), Doyle Jr, Graves, Gruber and Kleiner (2015), Arnold, Dobbie and Yang (2018), Dobbie, Goldin and Yang (2018), Bhuller, Dahl, Loken and Mogstad (2020), and González-Uribe and Reyes (2021).

The jackknife estimator is known to perform better than two-stage least-squares under such a paradigm (Chao and Swanson 2005). In our setting it can yield correct inference under asymptotics where m remains fixed as n grows. To highlight why this is so we will frame the first stage as an error-component model where judge decisions are composed of a judge-specific component and an idiosyncratic case-specific component. When cases are uncorrelated conditional on the judge effect, the decisions made by a given judge on cases other than the one under consideration constitute valid and relevant instrumental variables. With two cases per judge this yields a just-identified setting. Hence, consistent estimation of the judge effect is not needed. This idea is reminiscent to an identification argument used in the literature on peer effects (Jochmans 2023). The jackknife estimator can be obtained from this by averaging over all available other cases. This interpretation, while simple, does not appear to be well-appreciated in the literature.

In our formulation instrument strength is governed by the cross-sectional variance of the judge-specific component, say σ^2_{nm}. A relevant concern may be that this variance is small relative to the number of judges n. When σ^2_{nm} is allowed to shrink with the sample size, the jackknife estimator continues to deliver asymptotically-valid inference provided that $\sigma^2_{nm}\sqrt{n}m$ diverges as the sample size grows large. However, the convergence rate of the estimator is affected and the usual first-order asymptotic variance is no longer correct, in general. Moreover, unless $\sigma^2_{nm}m$ diverges, the variance needs to be adjusted. Below we provide a simple variance estimator that adapts to the severity of the weak-instrument problem. This variance estimator needs to be constructed with care, as naive plug-in estimators of the variance components will have biases that vanish only as m grows, and so are ill-suited for situations where judges are assigned only few cases. Our variance formula is similar to Chao, Swanson, Hausman, Newey and Woutersen (2012). Adjusting standard errors for the jackknife estimator to handle limited variation in judge leniency is important but does not appear to be common practice in empirical work. As part of our numerical illustrations, we highlight how this can lead to standard errors that dramatically understate the true sampling variability of the jackknife estimator, and so lead to substantially too many false positive test results.

Finally, it is useful to have some measure of instrument strength in the judge-leniency design. Classic rules-of-thumb based on the first-stage F-statistic, as in Staiger and Stock (1997) and Stock and Yogo (2005), cannot be used here unless $\sigma^2_{nm} {}^m\!/\!_n \to \infty$. We redo their exercise in our context and look for a statistic that can be informative about the maximal overrejection of a test based on the jackknife estimator under asymptotics where $\sigma^2_{nm}\sqrt{n}m$ remains bounded. The test statistic is a scaled version of the estimated signal-to-noise ratio in the error-component model. Here, again, the variances of both the signal and the noise need to be estimated with care. Although some of the details are different, our approach turns out to yield the same decision rule as the one recently proposed by Mikusheva and Sun (2022).

1 The judge-leniency design

Consider a setting where $n \times m$ individuals (e.g., cases or applications) are assigned to n groups (e.g., judges or examiners) so that each group contains m individuals. Let y_{gi} and x_{gi} denote observable random variables for individual i in group g, taken to be related via

$$y_{gi} = \beta_0 + \beta x_{gi} + \varepsilon_{gi}. \tag{1}$$

Interest lies in learning (the constant β_0 and) the slope coefficient β when x_{gi} is suspected to be correlated with the latent variable ε_{gi}. To this end the specification is completed with

$$x_{gi} = \alpha_0 + \sigma_{nm}\alpha_g + u_{gi}. \tag{2}$$

Here, α_0 and $\sigma_{nm} > 0$ are unknown coefficients while α_g and u_{gi} are latent variables. We impose the following assumptions throughout to complete the model.

A.1 The α_g are i.i.d. with mean zero and variance one.

A.2 The $(\varepsilon_{gi}, u_{gi})$ are i.i.d. across g and i with mean zero and finite fourth-order moments.

A.3 The α_g are independent of $(\varepsilon_{g'i}, u_{g'i})$ for all g, g' and all i.

These assumptions are meant to capture the essence of the judge-leniency design.[2] They state that x_{gi} is endogenous unless $\sigma_{u\varepsilon} := \mathbb{E}(u_{gi}\varepsilon_{gi})$ is equal to zero but also imply that the presence of group-specific effects α_g introduces exogenous variation that can be leveraged in an instrumental-variable strategy. In **A.1** the group effects have been normalized so that

$$\alpha_0 + \sigma_{nm}\alpha_g \sim \text{i.i.d.}(\alpha_0, \sigma_{nm}^2).$$

This is without loss of generality and convenient for our purposes as the instrument strength is governed by the single parameter σ_{nm}.

2 Estimators

Oracle estimator We begin by considering the infeasible instrumental-variable estimator that uses the unobservable α_g as instrument. This reduces the setting to a just-identified problem on group-averaged data. The oracle estimator of β equals

$$\hat{\beta}_{oracle} := \sum_{g=1}^{n} \alpha_g\, \tilde{y}_g \Big/ \sum_{g=1}^{n} \alpha_g\, \tilde{x}_g,$$

where $\tilde{y}_g := \bar{y}_g - \bar{y}$ for $\bar{y}_g := {}^1\!/_m \sum_{i=1}^m y_{gi}$ and $\bar{y} := {}^1\!/_n \sum_{g=1}^n \bar{y}_g$, and $\tilde{x}_g$ and $\bar{x}$ are defined in the same manner. This estimator can be applied in settings where we observe only a single individual per group.

Re-arranging terms using (1)–(2) and scaling up both numerator and denominator by the square-root of the sample size gives.

$$\hat{\beta}_{oracle} - \beta = \frac{\sqrt{{}^m\!/_n} \sum_{g=1}^n \alpha_g \bar{\varepsilon}_g + O_p({}^1\!/\!\sqrt{n})}{\sqrt{nm}\sigma_{nm} + \sqrt{{}^m\!/_n} \sum_{g=1}^n \alpha_g \bar{u}_g + O_p(\sqrt{m}\sigma_{nm})}. \tag{3}$$

[2] The assumptions are stronger than necessary and can be substantially relaxed, e.g., by allowing for heteroskedasticity in both equations. This would not alter our main points but would require additional technical conditions to validate the use of asymptotics, clouding the exposition. Under **A.1**–**A.3** all asymptotic statements below follow from standard arguments for the i.i.d. case when $n \to \infty$ while m is fixed, and from the limit theorems of Hall and Heyde (1980) on martingale arrays when $n, m \to \infty$ jointly.

Now, as $n \to \infty$ with m fixed or $n, m \to \infty$,

$$\begin{pmatrix} \sqrt{m/n}\sum_{g=1}^{n}\alpha_g\bar{\varepsilon}_g \\ \sqrt{m/n}\sum_{g=1}^{n}\alpha_g\bar{u}_g \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{u\varepsilon} \\ \sigma_{u\varepsilon} & \sigma_u^2 \end{pmatrix} \right),$$

where $\sigma_\varepsilon^2 := \mathbb{E}(\varepsilon_{gi}^2)$ and $\sigma_u^2 := \mathbb{E}(u_{gi}^2)$. We thus have three possible cases depending on the behavior of σ_{nm} as the sample size grows.

1. If $nm\sigma_{nm}^2 \to \infty$ the noise in the denominator is of a smaller order than the signal. Then, $\hat{\beta}_{oracle} \xrightarrow{p} \beta$ and

$$\sqrt{nm}(\sigma_{nm}/\sigma_\varepsilon)(\hat{\beta}_{oracle} - \beta) \xrightarrow{d} N(0,1).$$

This covers the strong-instrument setting where $\sigma_{nm} \to \sigma \in (0, \infty)$ but also allows for σ_{nm} to vanish, provided that it does so at a rate slower than $1/\sqrt{nm}$. In the former case the estimator converges at the parametric rate of $1/\sqrt{nm}$. In the latter case the rate slows down to $1/\sqrt{nm\sigma_{nm}^2}$.

2. If $nm\sigma_{nm}^2 \to c \in (0, \infty)$ we are in a setting that corresponds to the weak-instrument framework of Staiger and Stock (1997). The estimator stays random in the limit and, hence, is inconsistent. Its limit distribution is complicated to state but in our case follows readily from the calculations in Hinkley (1969) (see also Marsaglia 1965).

3. If $nm\sigma_{nm}^2 \to 0$, $\hat{\beta}_{oracle} - \beta$ behaves like the ratio of two zero-mean (but not independent) normal variables. Hence (see, e.g., Geary 1930),

$$\hat{\beta}_{oracle} - \beta \xrightarrow{d} \text{Cauchy}\left(\rho\,\sigma_\varepsilon/\sigma_u, \sqrt{1-\rho^2}\,\sigma_u/\sigma_\varepsilon\right),$$

where $\rho := \sigma_{u\varepsilon}/\sigma_u\sigma_\varepsilon$. This limit distribution is the same as what would be obtained with instruments that are completely irrelevant.

Two-stage least-squares estimator Now turn to the case where the group effects are unobserved. With multiple observations per group one can estimate α_g by the sample average

$$\bar{x}_g = \alpha_0 + \sigma_{nm}\alpha_g + \bar{u}_g =: \hat{\alpha}_g,$$

and construct the feasible two-stage least-squares estimator

$$\hat{\beta}_{2sls} := \sum_{g=1}^{n} \hat{\alpha}_g \, \tilde{y}_g \Bigg/ \sum_{g=1}^{n} \hat{\alpha}_g \, \tilde{x}_g .$$

The introduction of the first stage creates complications compared to the oracle estimator. We have

$$\hat{\beta}_{2sls} - \beta = \frac{\sigma_{nm}\sqrt{m/n}\sum_{g=1}^{n} \alpha_g \bar{\varepsilon}_g + \sqrt{m/n}\sum_{g=1}^{n} \bar{u}_g \bar{\varepsilon}_g + O_p(\sigma_{nm}/\sqrt{n} + 1/\sqrt{nm})}{\sqrt{nm}\sigma_{nm}^2 + 2\sigma_{nm}\sqrt{m/n}\sum_{g=1}^{n} \alpha_g \bar{u}_g + \sqrt{n/m}\sum_{g=1}^{n} \bar{u}_g^2 + O_p(\sigma_{nm}/\sqrt{n} + 1/nm)}. \tag{4}$$

Both numerator and denominator now feature (scaled averages of) V-statistics. These satisfy

$$\sqrt{m/n}\sum_{g=1}^{n} \bar{u}_g \bar{\varepsilon}_g = \sqrt{n/m}\,\sigma_{u\varepsilon} + O_p(1/\sqrt{m}), \qquad \sqrt{m/n}\sum_{g=1}^{n} \bar{u}_g^2 = \sqrt{n/m}\,\sigma_u^2 + O_p(1/\sqrt{m}).$$

The former introduces a bias of order $1/m$ into the two-stage least-squares, implying that it is inconsistent as $n \to \infty$ for fixed m. The issue can be understood through many-instrument asymptotics as in Bekker (1994) and others, but in our setting with grouped data it can simply be viewed as a manifestation of the incidental-parameter problem (Neyman and Scott 1948).

When m grows with n the bias shrinks but can remain important. First, inspection of (4) reveals that, in large samples, the bias term dominates all stochastic terms. Moreover,

$$\sigma_{nm}^2 m \to \infty$$

is required for consistency of the two-stage least-squares estimator. Under this condition,

$$\sqrt{nm}(\sigma_{nm}/\sigma_\varepsilon)(\hat{\beta}_{2sls} - \beta) = \sqrt{m/n}\sum_{g=1}^{n} \alpha_g \bar{\varepsilon}_g + \sqrt{n/\sigma_{nm}^2 m}\,\sigma_{u\varepsilon} + o_p(1),$$

which is asymptotically standard normal provided that

$$\sigma_{nm}^2 \, m/n \to \infty.$$

When $\sigma_{nm} \to \sigma \in (0, \infty)$ this requirement amounts to $n/m \to 0$, which is reminiscent to rates obtained in the literature on nonlinear models for panel data; see, e.g., Hahn and

Newey (2004). Thus, even in this case, the validity of conventional inference procedures based on two-stage least-squares requires m to be much larger than n. Such a situation is very rarely met in practice.

The bias issue could be mitigated by working with the bias-corrected estimator

$$\left(\sum_{g=1}^{n} \hat{\alpha}_g \, \tilde{y}_g - \hat{\sigma}_{u\varepsilon} {}^{n}\!/_{m} \right) \bigg/ \sum_{g=1}^{n} \hat{\alpha}_g \, \tilde{x}_g,$$

where $\hat{\sigma}_{u\varepsilon}$ is an estimator of $\sigma_{u\varepsilon}$. If the latter estimator is constructed using residuals from the two-stage least-squares estimator this will yield an asymptotically-unbiased estimator provided that

$$\sigma_{nm}^2 \, {}^{m^3}\!/_{n} \to \infty.$$

While this is an improvement over two-stage least squares, a better approach to debiased estimation is to rely on cross-fitting, which we turn to next.

Jackknife instrumental-variable estimator To motivate the use of cross-fitting in the judge-lenciency design note that $x_{g1}, \ldots, x_{gm}$ are all noisy signals of α_g. Furthermore, conditional on α_g, these signals are all uncorrelated and unbiased for $\alpha_0 + \sigma_{nm}\alpha_g$ while, unconditionally, their covariance is σ_{nm}^2.[3] It follows that

$$\operatorname{cov}(x_{gj}, y_{gi}) = \operatorname{cov}(x_{gj}, x_{gi}) \, \beta$$

for any $j \neq i$. Thus, each such x_{gj} is a valid and relevant instrumental variable for x_{gi}. Averaging over all possible j gives

$$z_{gi} := {}^{1}\!/_{m-1} \sum_{j \neq i} x_{gj}$$

and the estimator

$$\check{\beta}_{jive} := \sum_{g=1}^{n} \sum_{i=1}^{m} z_{gi} \tilde{y}_{gi} \bigg/ \sum_{g=1}^{n} \sum_{i=1}^{m} z_{gi} \tilde{x}_{gi},$$

[3]Lack of correlation between the x_{gi} conditional on α_g may be evaluated by a test for serial correlation in fixed-effect models for panel data; see Jochmans (2020a,b).

where $\tilde{y}_{gi} \coloneqq y_{gi} - \bar{y}$ and $\tilde{x}_{gi} \coloneqq x_{gi} - \bar{x}$. This is a jackknife instrumental-variable estimator as in Angrist, Imbens and Krueger (1999). Indeed,

$$z_{gi} = \alpha_0 + \sigma_{nm}\alpha_g + {}^1\!/_{m-1} \sum_{j \neq i} u_{gj}$$

can be seen as an estimator of $\alpha_0 + \sigma_{nm}\alpha_g$ that, in contrast to $\hat{\alpha}_g$, is uncorrelated with ε_{gi}.

The use of cross-fitting prevents the noise in the estimated group effects from inducing bias. The numerator of the jackknife estimator satisfies

$${}^1\!/_{nm} \sum_{g=1}^{n} \sum_{i=1}^{m} z_{gi}\tilde{y}_{gi} = {}^{\sigma_{n,m}}\!/_{n} \sum_{g=1}^{n} \alpha_g \bar{\varepsilon}_g + {}^1\!/_{nm(m-1)} \sum_{g=1}^{n} \sum_{i=1}^{m} \sum_{j \neq i} u_{gj}\varepsilon_{gi} + O_p({}^{\sigma_{n,m}}\!/_{n\sqrt{m}} + {}^1\!/_{nm}).$$

It differs from the numerator of two-stage least-squares in that it features the (degenerate) U-statistic

$${}^1\!/_{nm(m-1)} \sum_{g=1}^{n} \sum_{i=1}^{m} \sum_{j \neq i} u_{gj}\varepsilon_{gi} = O_p({}^1\!/_{\sqrt{n}m})$$

instead of the V-statistic ${}^1\!/_{nm^2} \sum_{g=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{m} u_{gj}\varepsilon_{gi} = {}^{\sigma_{u\varepsilon}}\!/_{m} + O_p({}^1\!/_{\sqrt{n}m})$ from before, which was the source of bias.

Re-arranging terms reveals that

$$\hat{\beta}_{jive} - \beta = {}^1\!/_{\sigma_{nm}n} \sum_{g=1}^{n} \alpha_g \bar{\varepsilon}_g + {}^1\!/_{\sigma^2_{nm}nm(m-1)} \sum_{g=1}^{n} \sum_{i=1}^{m} \sum_{j \neq i} u_{gj}\varepsilon_{gi} + O_p({}^1\!/_{\sigma_{n,m}n\sqrt{m}} + {}^1\!/_{\sigma^2_{n,m}nm}).$$

The leading two terms on the right-hand side are of the order ${}^1\!/_{\sqrt{\sigma^2_{nm}nm}}$ and ${}^1\!/_{\sigma^2_{nm}\sqrt{n}m}$, respectively. Therefore, consistency requires that

$$\sigma^2_{nm}\sqrt{n}m \to \infty.$$

Notice that this condition can be satisfied when $n \to \infty$ as m is held fixed. Moreover, in contrast to with two-stage least-squares, consistent estimation is possible as soon as two individuals per group are available.

The asymptotic distribution of the estimator depends on the relative importance of the two leading terms. We have three possible cases.

1. The first term converges slower than the second term when

$$\sigma^2_{nm}m \to \infty.$$

This is the same requirement that was needed for two-stage least squares to be consistent. In this case,

$$\sqrt{nm}({}^{\sigma_{nm}}/_{\sigma_\varepsilon})(\hat{\beta}_{jive} - \beta) \xrightarrow{d} N(0,1).$$

and the jackknife estimator has (to first order) the same limit distribution as the oracle estimator. However, this oracle property is clearly not attainable under asymptotics where m is held fixed.

2. Both terms converge at the same rate when

$$\sigma^2_{nm} m \to c^2 \in (0,\infty).$$

In this case we retain the same convergence rate as in Case 1 but the asymptotic variance is affected. We have

$$\sqrt{nm}({}^{\sigma_{nm}}/_{\sigma_\varepsilon})\sqrt{{}^{\sigma^2_{nm}m}/_{\sigma^2_{nm}m+\sigma^2_u}}(\hat{\beta}_{jive} - \beta) \xrightarrow{d} N(0,1).$$

Note that ${}^{\sigma^2_{nm}m}/_{\sigma^2_{nm}m+\sigma^2_u} \to {}^{c^2}/_{c^2+\sigma^2_u} \in (0,1)$ and so the the asymptotic variance in Case 2 is strictly larger than in Case 1. This case covers asymptotics where $n \to \infty$ while m remains fixed and $\sigma_{nm} \to \sigma \in (0,\infty)$. It also covers drifting sequences of the form $\sigma_{nm} = {}^{c}/_{\sqrt{n}}$ for a constant c when n and m grow large at the same rate. This drifting sequence is interesting as it captures the idea that the group effect α_g may explain little of the cross-sectional variation in the x_{gi}.

3. The first term converges faster than the second term when

$$\sigma^2_{nm} m \to 0.$$

Here, the second term dominates. This affects the convergence rate and the asymptotic variance. We now have

$$\sqrt{n}m({}^{\sigma^2_{nm}}/_{\sigma_u\sigma_\varepsilon})(\hat{\beta}_{jive} - \beta) \xrightarrow{d} N(0,1).$$

This final case is important as it covers situations where $\sigma_{nm} \to 0$ when $n \to \infty$ and m remains fixed. It equally applies to drifting sequences of the form $\sigma_{nm} = {}^{c}/_{\sqrt{n}}$ when n grows faster than m, although the consistency requirement in such a case would still demand that ${}^{n}/_{m^2} \to 0$.

Compared to the oracle estimator, we observe two important differences. First, inference based on the former requires that $\sigma_{nm}^2 nm \to \infty$ while the jackknife requires the stronger condition that $\sigma_{nm}^2 \sqrt{n} m \to \infty$. Second, the feasible estimator will often have an asymptotic variance that is different from the one of the oracle estimator. Both these differences can be attributed to the presence of the U-statistics, which arise because the group effects are unobserved.

3 Inference

Estimators for variance components The above results show that there is no basis to further entertain the two-stage least-squares estimator and so we proceed with the jackknife estimator. For inference estimators of the various components of the asymptotic variances above are needed. Mindful of the potential bias induced by estimating group effects we propose to use

$$\hat{\sigma}_{nm}^2 := {}^1/_{nm} \sum_{g=1}^{n} \sum_{i=1}^{m} z_{gi} \tilde{x}_{gi}, \quad \text{and} \quad \hat{\sigma}_u^2 := {}^1/_{nm(m-1)} \sum_{g=1}^{n} \sum_{i=1}^{m} \sum_{j \neq i} {}^1/_2 (x_{gi} - x_{gj})^2,$$

as estimators of σ_{nm}^2 and σ_u^2, respectively. Both can be seen to be cross-fit type estimators. This avoids bias of order ${}^1/_m$ that would arise from conventional plug-in estimators. This is important because we want to cover cases where m may be held fixed as $n \to \infty$. In fact, both of the proposed estimators are exactly unbiased. Furthermore,

$$\hat{\sigma}_u^2 = {}^1/_{nm} \sum_{g=1}^{n} \sum_{i=1}^{m} u_{gi}^2 - {}^1/_{nm(m-1)} \sum_{g=1}^{n} \sum_{i=1}^{m} \sum_{j \neq i} u_{gi} u_{gj} = \sigma_u^2 + O_p({}^1/_{\sqrt{nm}}) + O_p({}^1/_{\sqrt{n}m}),$$

while

$$\hat{\sigma}_{nm}^2 = \sigma_{nm}^2 + \sigma_{nm}{}^2/_n \sum_{g=1}^{n} \alpha_g \bar{u}_g + {}^1/_{nm(m-1)} \sum_{g=1}^{n} \sum_{i=1}^{m} \sum_{j \neq i} u_{gi} u_{gj} + O_p({}^{\sigma_{nm}^2}/_{\sqrt{n}} + {}^{\sigma_{nm}}/_{n\sqrt{m}} + {}^1/_{nm}),$$

and so both estimators are consistent as $n \to \infty$ whether or not m grows with n. Note that $\hat{\sigma}_{nm}^2$ is nothing else than the denominator of the jackknife instrumental-variable estimator.

To estimate σ_ε^2, finally, we will make use of

$$\hat{\sigma}_\varepsilon^2 := {}^1\!/_{nm} \sum_{g=1}^{n} \sum_{i=1}^{m} \hat{\varepsilon}_{gi}^2,$$

where $\hat{\varepsilon}_{gi} := \tilde{y}_{gi} - \tilde{x}_{gi}\hat{\beta}_{jive}$. Let $\omega := \hat{\beta}_{jive} - \beta = O_p({}^1\!/_{\sigma^2_{nm}\sqrt{nm}})$. A small calculation shows that

$$\hat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 - 2\sigma_{u\varepsilon}\,\omega + (\sigma_{nm}^2 + \sigma_u^2)\,\omega^2 + O_p({}^1\!/_{\sqrt{nm}} + {}^{\omega^2}\!/_{\sqrt{n}} + {}^{\omega^2}\!/_{n\sqrt{m}}).$$

Consistency of the jackknife estimator—i.e., $\omega \xrightarrow{p} 0$ or $\sigma^2_{nm}\sqrt{nm} \to \infty$—therefore implies that $\hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_\varepsilon^2$.

Studentization The jackknife estimator can be studentized in a manner so that it is asymptotically standard normal no matter the case which we are in. With the estimators from the previous paragraph,

$$\hat{t} := \sqrt{nm}({}^{\hat{\sigma}_{nm}}\!/_{\hat{\sigma}_\varepsilon})\sqrt{{}^{\hat{\sigma}^2_{nm}m}\!/_{\hat{\sigma}^2_{nm}m+\hat{\sigma}^2_u}}\,(\hat{\beta}_{jive} - \beta) \xrightarrow{d} N(0,1),$$

independent of the behavior of $\sigma^2_{nm}m$, provided, of course, that $\sigma^2_{nm}\sqrt{nm} \to \infty$. Indeed, the factor, $\sqrt{{}^{\hat{\sigma}^2_{nm}m}\!/_{\hat{\sigma}^2_{nm}m+\hat{\sigma}^2_u}}$ makes the studentization adapt to $\sigma^2_{nm}m$. Therefore, we do not need to know which of the three cases we are in to perform asymptotically-valid inference.

Evaluating instrument strength Under asymptotics where $\sigma^2_{nm}\sqrt{nm} \to c < \infty$ the jackknife estimator is inconsistent. Indeed,

$$\omega \xrightarrow{p} {}^1\!/_{\sqrt{nm}} \sum_{g=1}^{n}\sum_{i=1}^{m}\sum_{j\neq i} u_{gj}\varepsilon_{gi} \Bigg/ \left(c + {}^1\!/_{\sqrt{nm}} \sum_{g=1}^{n}\sum_{i=1}^{m}\sum_{j\neq i} u_{gj}u_{gi}\right) \xrightarrow{d} \omega_1/\omega_2,$$

where

$$\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ c \end{pmatrix}, \begin{pmatrix} \sigma_u^2\sigma_\varepsilon^2 & \sigma_u^2\sigma_{u\varepsilon} \\ \sigma_u^2\sigma_{u\varepsilon} & \sigma_u^2\sigma_u^2 \end{pmatrix} \right).$$

The estimator stays random in the limit, following a distribution that can be obtained from Hinkley (1969) (a Cauchy distribution is obtained in the limit where $c \to 0$). Using the expansions from the previous paragraph a small calculation reveals that the studentized statistic from above remains bounded in probability under such asymptotic sequences. Moreover,

$$\hat{t}^2 \xrightarrow{d} \omega_1^2/(\sigma_u^2\sigma_\varepsilon^2 - 2\sigma_u^2\sigma_{u\varepsilon}\,\omega + \sigma_u^4\,\omega^2) \sim \varpi_1^2/(1 - 2\rho\varpi + \varpi^2),$$

where the last transition follows from normalising the random variables ω_1 and ω_2, and uses $\varpi := {}^{\varpi_1}\!/_{\varpi_2}$ for

$$\begin{pmatrix} \varpi_1 \\ \varpi_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ c_0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),$$

where $c_0 := c/\sigma_u^2$ and, recall, $\rho = {}^{\sigma_{u\varepsilon}}\!/_{\sigma_u\sigma_\varepsilon}$ is the correlation of the disturbances. Following Staiger and Stock (1997) and Stock and Yogo (2005) we can then calculate the asymptotic size of a two-sided t-test with nominal size a_0, i.e.,

$$a_{\rho,c_0}(a_0) := \mathbb{P}(|\hat{t}| > \Phi^{-1}(1 - {}^{a_0}\!/_2)),$$

under sequences where $\sigma_{nm}^2\sqrt{nm} \to c < \infty$ given values for the nuisance parameters ρ and c_0. The maximal rejection probability of the t-test with nominal size a_0, as a function of c_0, then is $\max_{\rho\in[-1,1]} a_{\rho,c_0}(a_0) = a_{1,c_0}(a_0)$. Table 1 reports this worst-case size for various combinations of a_0 and c_0. For example, a test with nominal level $a_0 = .01$ has actual size $a_{1,c_0}(a_0) \leq .05$ for $c_0 \geq 3$.

When $\sigma_{nm}^2\sqrt{nm} \to c < \infty$ the scaled jackknife denominator satisfies $\sqrt{nm}\,\hat{\sigma}_{nm}^2 \xrightarrow{d} \omega_2$. Estimating its standard deviation by $\hat{\sigma}_u^2$ and studentizing by it gives rise to the test statistic

$$\hat{\tau} := {}^{\sqrt{nm}\hat{\sigma}_{nm}^2}\!/_{\hat{\sigma}_u^2} \xrightarrow{d} \varpi_2 \sim N(c_0, 1).$$

Moreover, $\mathbb{P}(\hat{\tau} > w) = 1 - \Phi(w - c_0)$. It follows that, for $w_{c_0}(a_0) := c_0 + \Phi^{-1}(1 - a_0)$, as the sample size grows large, the power of the test that $\mathbb{E}(\varpi_2) \leq c_0$ under the alternative that $\mathbb{E}(\varpi_2) = c_1$, i.e.,

$$b_{a_0}(c_0 - c_1) := \mathbb{P}(\hat{\tau} > w_{c_0}(a_0)) = 1 - \Phi(c_0 - c_1 + \Phi^{-1}(1 - a_0))$$

is monotone decreasing in the difference $c_0 - c_1$, and equals a_0 when $c_0 - c_1 = 0$. The values $w_{c_0}(a_0)$ as a function of (a_0, c_0) are tabulated in Table 2. The power curves for different values of a_0 are provided in Figure 1.

4 Numerical illustrations

Design We simulated unobservables by drawing α_g from a standard-normal distribution and $(\varepsilon_{gi}, u_{gi})$ from a bivariate normal distribution with zero mean, unit variances, and correlation $\rho = {}^1\!/_2$. We then generated x_{gi} from (2) with $\alpha_0 = 0$ and $\sigma^2_{nm} \in \{1, {}^1\!/_n\}$, and finally, y_{gi} from (1) with $\beta_0 = 0$ and $\beta = 1$. For the sample size, all combinations of $n \in \{25, 50, 100, 250, 500, 1000\}$ and $m \in \{5, 25, 50, 100, 250, 500\}$ are considered. All simulation results are based on 100,000 replications.

Two-stage least-squares estimator To begin, Table 3 and Figures 2 and 3 concern the two-stage least-squares estimator. The table reports the (median) bias in the point estimator and the size of a two-sided t-test for the null that $\beta = 1$ (with a nominal size of 5%) for the different sample sizes and instrument strength. Figure 2 plots the cumulative distribution function of the studentized estimator (i.e., the t-statistic) (full black curve) against the standard-normal distribution (dashed black curve) when $\sigma^2_{nm} = 1$. Figure 3 does the same for the case where $\sigma^2_{nm} = {}^1\!/\sqrt{n}$.

Table 3 clearly shows that, when $\sigma^2_{nm} = 1$, the bias is roughly proportional to ${}^1\!/_m$ and largely constant in n. The bias is important relative to the estimator's standard error and so the t-test suffers from substantial over-rejection, except in the cases where $m = 500$ and $n \leq 100$. All of this is in full accordance with the theory. Figure 2 reveals that the distribution of the t-statistic stochastically dominates the standard-normal distribution in all but the most upper-right plots, a consequence of the (positive) bias in the point estimator.

When $\sigma^2_{nm} = {}^1\!/\sqrt{n}$ the bias becomes substantially larger for all samples sizes and now also increases with n. The rejection frequencies further go up and, now, no longer improve as ${}^n\!/_m$ becomes smaller. The plots in Figure 3 show large discrepancies between the estimators distribution and the standard-normal curve. In many cases the difference is so pronounced that the former curve falls completely outside the [-4,4] interval, where 99.99% of the total mass of a standard normal lies (and so is absent from the plots).

Jackknife instrumental-variable estimator Table 3 and Figures 4 and 5 deal with the jackknife estimator. In all designs where $\sigma^2_{nm} = 1$ the cross-fitting cleanly avoids the bias issue that plagues the two-stage least-squares estimator. The bias is small in all designs and the rejection frequency of the t-test is close to its nominal size of 5%. Moreover, as the plots in Figure 4 illustrate, the normal approximation to the jackknife estimator is very accurate throughout.

When $\sigma^2_{nm} = {}^1\!/\!\sqrt{n}$ the jackknife estimator suffers from substantial bias in the shortest data sets, where $m = 5$. This can be explained by the fact that, in this set of designs, $\sigma^2_{nm}\sqrt{n}m = \sqrt{{}^{m^2}\!/n}$. Thus, consistency of the jackknife estimator requires that ${}^{m^2}\!/n$ diverges, and so bias is to be expected even in large samples unless m^2 is not small relative to n. When this is the case, the bias is small and the t-statistic constitutes a reliable tool for inference. The latter is confirmed both by the rejection frequencies reported in the table and by the accuracy of the normal approximations to the jackknife's distribution in Figure 5.

It is important to use the adaptive variance estimator from above to obtain reliable inference. To illustrate this Table 5 reports

$$(\sigma^2_{nm}m+\sigma^2_u)/\sigma^2_{nm}m$$

for the various combinations of (n, m) when $\sigma^2_{nm} = {}^1\!/\!\sqrt{n}$. This is the factor by which the actual (asymptotic) variance of the jackknife estimator in our theory exceeds variance obtained through the conventional first-order approximation. Even in cases where both n and m are large, naive approximations lead to variances that can be three to six times too small.

Evaluating instrument strength Finally, Table 6 reports the value of $\sigma^2_{nm}\sqrt{n}m$ in the designs where $\sigma^2_{nm} = {}^1\!/\!\sqrt{n}$ together with the rejection frequency of the test procedure that rejects the null of weak identification with nominal size of 5%. Here, weak identification is used to mean that a two-sided t-test with nominal size of 5% based on the jackknife has an actual size of more than 10%. This is a test of the null that $c_0 \leq 2.5$ based on the limit distribution of $\hat{\tau}$. Clearly, when n is large and m is small—and so c_0 is very small—this test almost always fails to reject this null. In contrast when m is of a comparable magnitude as n, or larger, the test virtually always rejects. The approach thus seems to perform quite satisfactory here. Of course, when c_0 is only little larger than the cut-off of 2.5 there is some probability of false negatives.

Tables and figures

Figure 1: $b_{a_0}(c_0 - c_1)$ for $a_0 \in \{.01, .05, .10\}$ (full, dashed, dashed-dotted) as a function of $c_0 - c_1$

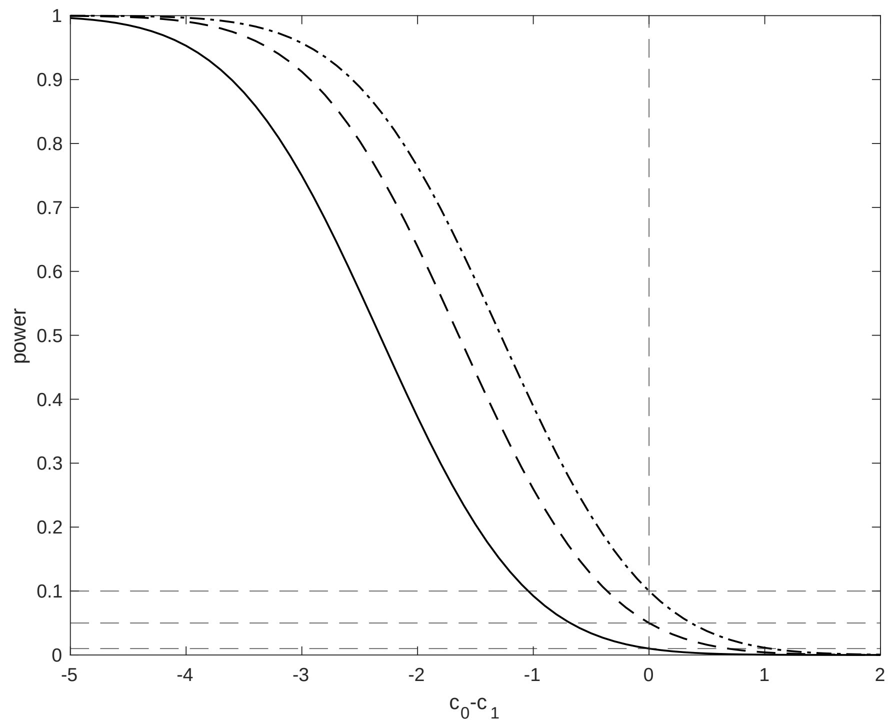


Table 1: $a_{1,c_0}(a_0)$ as a function of a_0, c_0

$c_0 \backslash a_0$	0.010	0.050	0.100	$c_0 \backslash a_0$	0.010	0.050	0.100
0.0	1.000	1.000	1.000	2.3	0,062	0.103	0.134
0.1	0.611	0.657	0.683	2.4	0,059	0.101	0.131
0.2	0.471	0.528	0.562	2.5	0,057	0.098	0.129
0.3	0.378	0.440	0.477	2.6	0,055	0.096	0.127
0.4	0.310	0.374	0.413	2.7	0,054	0.094	0.125
0.5	0.260	0.322	0.362	2.8	0,052	0.092	0.123
0.6	0.221	0.282	0.321	2.9	0,050	0.090	0.121
0.7	0.191	0.249	0.287	3.0	0,049	0.088	0.119
0.8	0.167	0.224	0.261	3.1	0,047	0.087	0.117
0.9	0.148	0.202	0.239	3.2	0,046	0.085	0.116
1.0	0.133	0.185	0.221	3.3	0,045	0.083	0.114
1.1	0.121	0.172	0.206	3.4	0,043	0.082	0.113
1.2	0.111	0.160	0.194	3.5	0,042	0.081	0.111
1.3	0.103	0.150	0.183	3.6	0,041	0.079	0.110
1.4	0.096	0.142	0.174	3.7	0,040	0.078	0.109
1.5	0.090	0.135	0.167	3.8	0,039	0.077	0.108
1.6	0.085	0.130	0.161	3.9	0,038	0.076	0.106
1.7	0.080	0.125	0.156	4.0	0,037	0.075	0.105
1.8	0.076	0.120	0.151	4.1	0,037	0.074	0.104
1.9	0.073	0.116	0.147	4.2	0,036	0.073	0.103
2.0	0.070	0.112	0.143	4.3	0,035	0.072	0.102
2.1	0.067	0.109	0.140	4.4	0,034	0.071	0.101
2.2	0.064	0.106	0.137	4.5	0,034	0.070	0.100

Table 2: $w_{c_0}(a_0)$ as a function of a_0, c_0

$\bar{c}\backslash\alpha_0$	0.010	0.050	0.100	$\bar{c}\backslash\alpha_0$	0.010	0.050	0.100
0.0	2.326	1.645	1.282	2.3	4.626	3.945	3.582
0.1	2.426	1.745	1.382	2.4	4.726	4.045	3.682
0.2	2.526	1.845	1.482	2.5	4.826	4.145	3.782
0.3	2.626	1.945	1.582	2.6	4.926	4.245	3.882
0.4	2.726	2.045	1.682	2.7	5.026	4.345	3.982
0.5	2.826	2.145	1.782	2.8	5.126	4.445	4.082
0.6	2.926	2.245	1.882	2.9	5.226	4.545	4.182
0.7	3.026	2.345	1.982	3.0	5.326	4.645	4.282
0.8	3.126	2.445	2.082	3.1	5.426	4.745	4.382
0.9	3.226	2.545	2.182	3.2	5.526	4.845	4.482
1.0	3.326	2.645	2.282	3.3	5.626	4.945	4.582
1.1	3.426	2.745	2.382	3.4	5.726	5.045	4.682
1.2	3.526	2.845	2.482	3.5	5.826	5.145	4.782
1.3	3.626	2.945	2.582	3.6	5.926	5.245	4.882
1.4	3.726	3.045	2.682	3.7	6.026	5.345	4.982
1.5	3.826	3.145	2.782	3.8	6.126	5.445	5.082
1.6	3.926	3.245	2.882	3.9	6.226	5.545	5.182
1.7	4.026	3.345	2.982	4.0	6.326	5.645	5.282
1.8	4.126	3.445	3.082	4.1	6.426	5.745	5.382
1.9	4.226	3.545	3.182	4.2	6.526	5.845	5.482
2.0	4.326	3.645	3.282	4.3	6.626	5.945	5.582
2.1	4.426	3.745	3.382	4.4	6.726	6.045	5.682
2.2	4.526	3.845	3.482	4.5	6.826	6.145	5.782

Table 3: Two-stage least-squares estimator

$\sigma^2_{nm} = 1$													
Bias							Empirical size (for a nominal size of .05)						
$n \backslash m$	5	25	50	100	250	500	$n \backslash m$	5	25	50	100	250	500
25	0.0829	0.0191	0.0099	0.0050	0.0020	0.0010	25	0.2103	0.0898	0.0727	0.0639	0.0590	0.0570
50	0.0831	0.0192	0.0098	0.0049	0.0020	0.0010	50	0.3324	0.1175	0.0842	0.0672	0.0610	0.0560
100	0.0834	0.0192	0.0097	0.0049	0.0020	0.0010	100	0.5609	0.1736	0.1124	0.0811	0.0639	0.0568
250	0.0833	0.0192	0.0098	0.0050	0.0020	0.0010	250	0.9069	0.3471	0.2022	0.1245	0.0793	0.0639
500	0.0834	0.0192	0.0098	0.0049	0.0020	0.0010	500	0.9961	0.5950	0.3521	0.2010	0.1095	0.0792
1000	0.0833	0.0192	0.0098	0.0050	0.0020	0.0010	1000	1.0000	0.8748	0.6046	0.3519	0.1709	0.1095
$\sigma^2_{nm} = {}^1/\sqrt{n}$													
Bias							Empirical size (for a nominal size of .05)						
$n \backslash m$	5	25	50	100	250	500	$n \backslash m$	5	25	50	100	250	500
25	0.4165	0.2501	0.1675	0.1002	0.0449	0.0235	25	0.7110	0.4955	0.3601	0.2373	0.1378	0.1000
50	0.4542	0.3337	0.2508	0.1669	0.0827	0.0453	50	0.9543	0.8702	0.7570	0.5820	0.3336	0.2051
100	0.4762	0.4002	0.3334	0.2501	0.1431	0.0832	100	0.9995	0.9971	0.9899	0.9595	0.7906	0.5572
250	0.4906	0.4544	0.4167	0.3570	0.2501	0.1668	250	1.0000	1.0000	1.0000	1.0000	0.9999	0.9965
500	0.4951	0.4762	0.4546	0.4167	0.3333	0.2497	500	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1000	0.4976	0.4879	0.4762	0.4546	0.4000	0.3334	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4: Jackknife instrumental-variable estimator

$\sigma^2_{nm} = 1$													
Bias							Empirical size (for a nominal size of .05)						
$n\backslash m$	5	25	50	100	250	500	$n\backslash m$	5	25	50	100	250	500
25	-0.0187	-0.0031	-0.0013	-0.0006	-0.0003	-0.0002	25	0.0561	0.0517	0.0496	0.0492	0.0501	0.0504
50	-0.0080	-0.0013	-0.0006	-0.0004	-0.0001	-0.0001	50	0.0588	0.0516	0.0498	0.0500	0.0508	0.0505
100	-0.0035	-0.0007	-0.0004	-0.0002	-0.0000	-0.0000	100	0.0596	0.0511	0.0505	0.0506	0.0504	0.0498
250	-0.0014	-0.0003	-0.0002	-0.0000	-0.0000	-0.0000	250	0.0621	0.0517	0.0513	0.0499	0.0507	0.0495
500	-0.0006	-0.0002	-0.0000	-0.0001	-0.0000	-0.0000	500	0.0605	0.0515	0.0504	0.0502	0.0498	0.0500
1000	-0.0004	-0.0001	-0.0000	0.0000	-0.0000	0.0000	1000	0.0607	0.0525	0.0496	0.0490	0.0500	0.0495
$\sigma^2_{nm} = 1/\sqrt{n}$													
Bias							Empirical size (for a nominal size of .05)						
$n\backslash m$	5	25	50	100	250	500	$n\backslash m$	5	25	50	100	250	500
25	0.3653	-0.0224	-0.0202	-0.0106	-0.0051	-0.0024	25	0.0756	0.0415	0.0415	0.0455	0.0503	0.0506
50	0.4231	-0.0006	-0.0187	-0.0103	-0.0044	-0.0024	50	0.0777	0.0483	0.0443	0.0483	0.0513	0.0519
100	0.4591	0.0525	-0.0165	-0.0092	-0.0038	-0.0022	100	0.0798	0.0534	0.0491	0.0488	0.0544	0.0531
250	0.4781	0.1729	0.0075	-0.0094	-0.0040	-0.0017	250	0.0797	0.0564	0.0540	0.0503	0.0598	0.0587
500	0.4885	0.2854	0.0678	-0.0077	-0.0032	-0.0024	500	0.0800	0.0589	0.0566	0.0549	0.0597	0.0599
1000	0.4983	0.3732	0.1658	0.0108	-0.0041	-0.0016	1000	0.0803	0.0595	0.0574	0.0578	0.0570	0.0648

Table 5: Variance-inflation factor $^{(\sigma_{nm}^2 m+\sigma_u^2)}/_{\sigma_{nm}^2 m}$ when $\sigma_{nm}^2 = {}^1/_{\sqrt{n}}$

$n\backslash m$	5	25	50	100	250	500
25	6.00	2.00	1.50	1.25	1.10	1.05
50	11.00	3.00	2.00	1.50	1.20	1.10
100	21.00	5.00	3.00	2.00	1.40	1.20
250	51.00	11.00	6.00	3.50	2.00	1.50
500	101.00	21.00	11.00	6.00	3.00	2.00
1000	201.00	41.00	21.00	11.00	5.00	3.00

Table 6: Instrument strength and evaluation when $\sigma_{nm}^2 = {}^1/_{\sqrt{n}}$

$\sigma_{nm}^2\sqrt{n}m$						
$n\backslash m$	5	25	50	100	250	500
25	1.000	5.000	10.000	20.000	50.000	100.000
50	0.707	3.536	7.071	14.142	35.355	70.711
100	0.500	2.500	5.000	10.000	25.000	50.000
250	0.316	1.581	3.162	6.325	15.811	31.623
500	0.224	1.118	2.236	4.472	11.180	22.361
1000	0.158	0.791	1.581	3.162	7.906	15.811
Rejection frequency of null that $c_0 \leq 2.5$ using $\hat{\tau}$						
$n\backslash m$	5	25	50	100	250	500
25	0.0585	0.5286	0.9120	0.9974	1.0000	1.0000
50	0.0352	0.3317	0.8254	0.9963	1.0000	1.0000
100	0.0209	0.1620	0.6133	0.9863	1.0000	1.0000
250	0.0117	0.0536	0.2626	0.8549	1.0000	1.0000
500	0.0095	0.0252	0.1080	0.5555	0.9998	1.0000
1000	0.0081	0.0136	0.0432	0.2553	0.9848	1.0000

Figure 2: Distributions of t-statistics for the two-stage least-squares estimator, $\sigma^2_{nm} = 1$

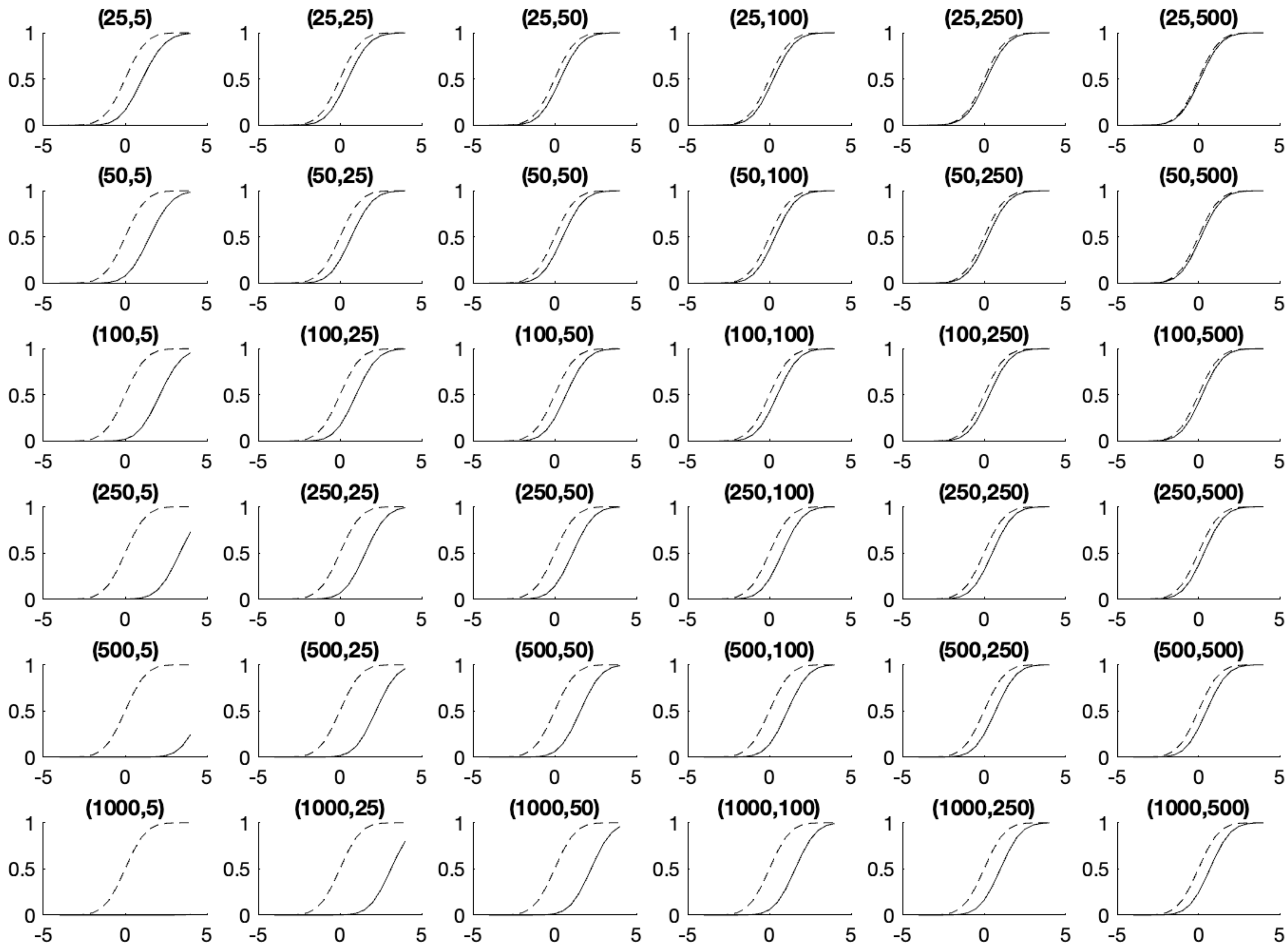


Figure 3: Distributions of t-statistics for the two-stage least-squares estimator, $\sigma^2_{nm} = {}^1/\sqrt{n}$

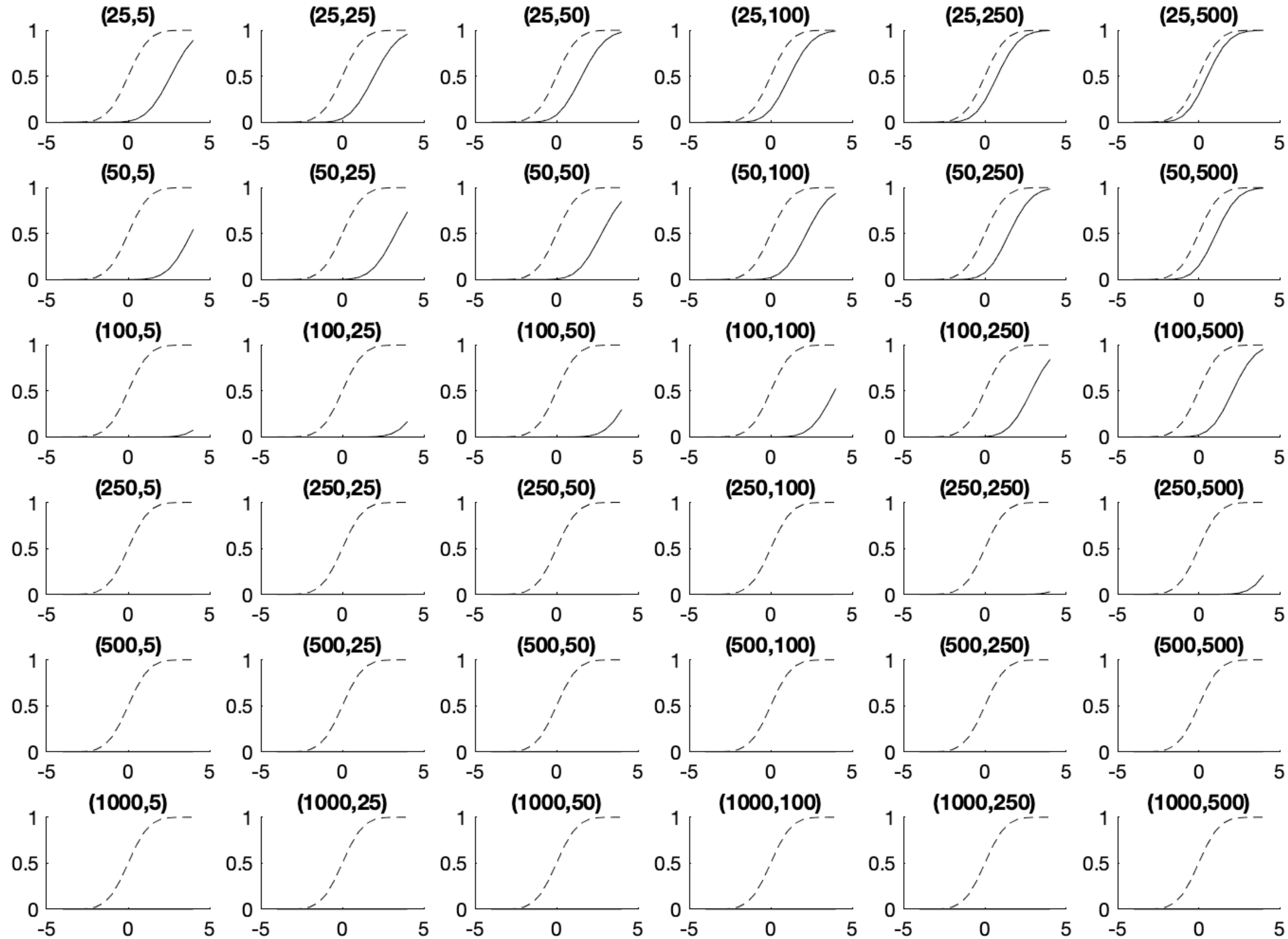


Figure 4: Distributions of t-statistics for the jackknife instrumental-variable estimator, $\sigma^2_{nm} = 1$

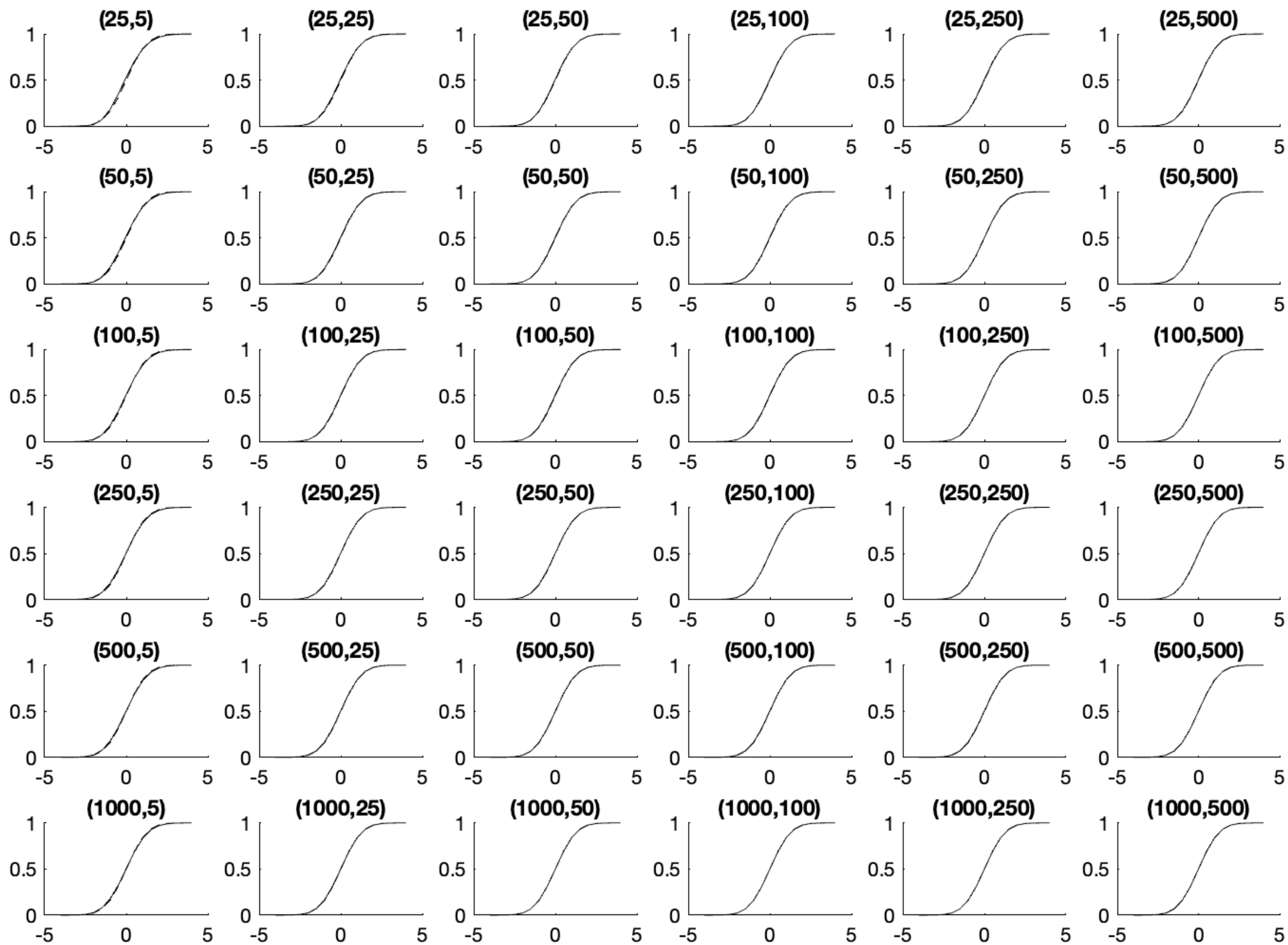
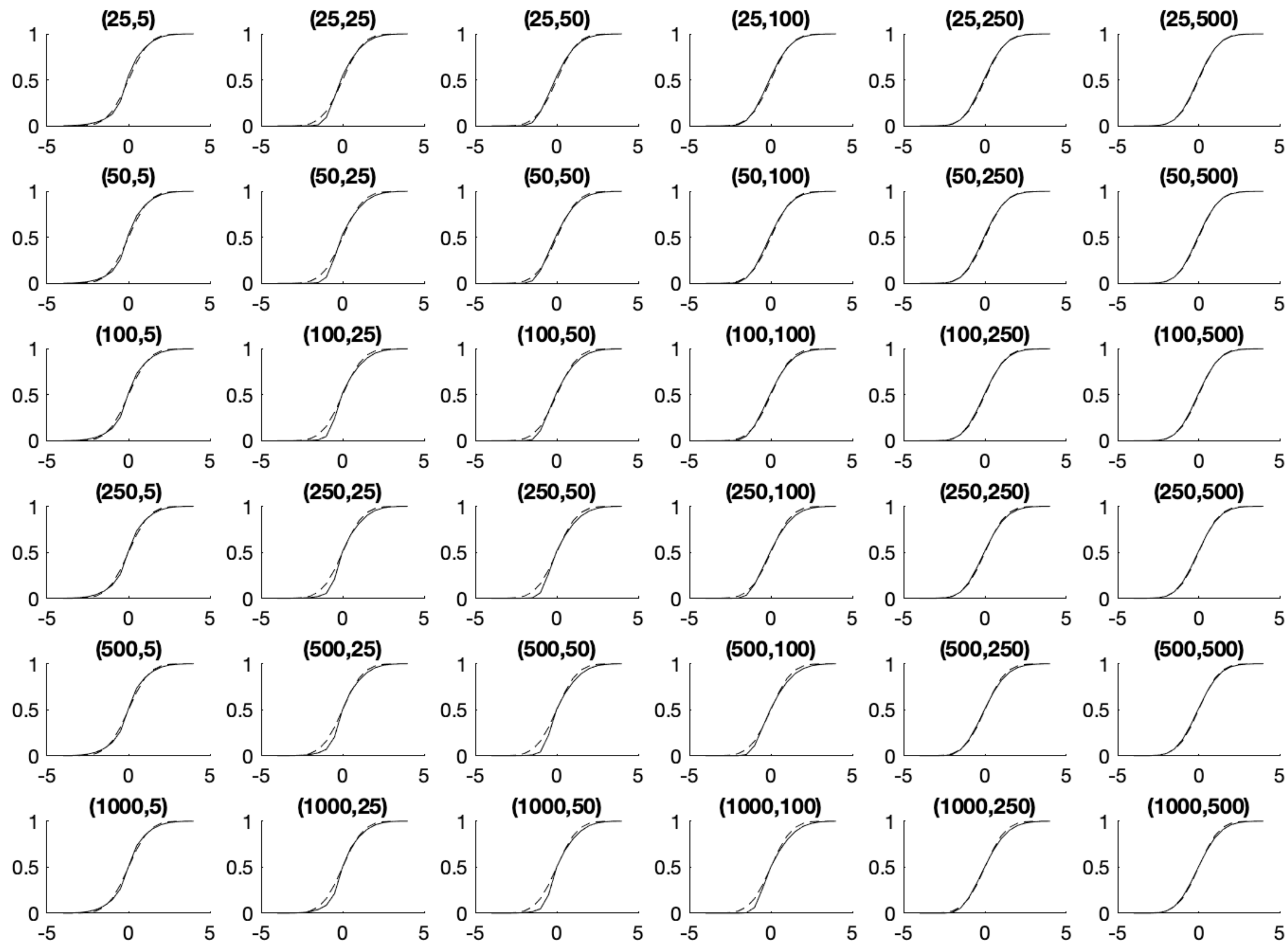


Figure 5: Distributions of t-statistics for the jackknife instrumental-variable estimator, $\sigma^2_{nm} = {}^1/_{\sqrt{n}}$



References

Aizer, A. and J. Doyle Jr. (2015). Juvenile incarceration, human capital, and future crime: Evidence from randonly-assigned judges. *Quarterly Journal of Economics 130*, 759–803.

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

Arnold, D., W. Dobbie, and C. S. Yang (2018). Racial bias in bail decisions. *Quarterly Journal of Economics 133*, 1885–1932.

Bekker, P. A. (1994). Alternative approximations to the distributions of the instrumental variable estimators. *Econometrica 62*, 657–681.

Bhuller, M., G. B. Dahl, K. V. Loken, and M. Mogstad (2020). Incarceration, recidivism, and employment. *Journal of Political Economy 128*, 1269–1324.

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

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012). Asymptotic distribution of JIV in a heteroskedastic IV regression with many weak instruments. *Econometric Theory 28*, 42–86.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*, 991–1030.

Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrail detention on conviction, future crime, and employment. *American Economic Review 108*, 201–240.

Doyle Jr, J., J. A. Graves, J. Gruber, and S. A. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy 123*, 170–214.

Frandsen, B., L. Lefgren, and E. Leslie (2023). Judging judge fixed effects. *American Economic Review 113*, 253–277.

Geary, R. C. (1930). The frequency distribution of the quotient of two normal variates. *Journal of the Royal Statistical Society 93*, 442–446.

González-Uribe, J. and S. Reyes (2021). Identifying and boosting "gazelles": Evidence from business accelerators. *Journal of Financial Economics 139*, 260–287.

Hahn, J. and W. K. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*, 1295–1319.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and its Applications*. Academic Press.

Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika 56*, 635–639.

Jochmans, K. (2020a). A portmanteau test for correlation in short panels. *Econometric Theory 36*, 1159–1166.

Jochmans, K. (2020b). Testing for correlation in error-component models. *Journal of Applied Econometrics 35*, 860–878.

Jochmans, K. (2023). Peer effects and endogenous social interactions. *Journal of Econometrics 235*, 1203–1214.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review 96*, 863–876.

Kunitomo, N. (1980). Asymptotic expansions of the distribution of estimators in a linear functional relationship and simultaneous equations. *Journal of the American Statistical Association 75*, 693–700.

Maestas, N., K. J. Mullen, and A. Strand (2013). Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *American Economic Review 103*, 1797–1929.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association 60*, 193–204.

Mikusheva, A. and L. Sun (2022). Inference with many weak instruments. *Review of Economic Studies 89*, 2663–2686.

Morimune, K. (1983). Approximate distributions of the k-class estimators when the degree of overidentifiability is large compared to the sample size. *Econometrica 51*, 821–841.

Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Stock, J. H. and M. Yogo (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and Inference for Econometric Models*, Chapter 5, pp. 80–108. Cambridge University Press.