Testing for Instrument Validity with Higher-Order Cumulants*

N. Aaron Pancost[†] and Garrett Schaller[‡]

This version: March 23, 2024

Abstract

Instrumental variables estimators are commonly used in economics and finance to establish causal relationships. Although instruments that fail the exclusion restriction do not reliably estimate parameters of interest, testing the exclusion restriction is uncommon, due to the difficulty of finding multiple valid instruments. We derive closed-form instrumental variable estimators that allow for tests of over-identifying restrictions even for the case of a single valid instrument. We also derive estimators that are consistent when instruments and regressors are mis-measured with correlated errors. Monte Carlo simulations suggest that our estimators have power to reject even in relatively small samples. We also apply our estimators to the IV regressions of Mian and Sufi (2014) and cannot reject the null hypothesis that the exclusion restriction holds.

Keywords: instrumental variables, endogeneity, exclusion restriction, measurement error.
JEL Classifications: C26, C36.

*We thank Jonathan Cohn, Cesare Fracassi, Sam Kruger, and Sheridan Titman for helpful comments and conversations, as well as seminar participants at the University of Texas at Austin. Any remaining errors are our own. We are grateful to Changyong Song and Minjoo Kim for outstanding research assistance.

[†]Corresponding author. University of Texas at Austin, McCombs School of Business. Email address: aaron.pancost@mccombs.utexas.edu

[‡]Colorado State University, College of Business. Email address: garrett.schaller@colostate.edu

1 Introduction

A common goal of empirical work is to establish causal relationships among some variables of interest. However, teasing out causal relationships from evidence on correlations alone is difficult, given the wealth of endogeneity problems in economics and finance. Researchers can, and often do, appeal to instrumental variables to address endogeneity, provided the instrument(s) satisfy both a relevance condition and an exclusion restriction. While the relevance condition is both testable and frequently tested, the exclusion restriction is usually treated as a theoretical construct that lies beyond empirical reach. In other words, the relevance condition is tested, while the exclusion restriction is merely argued for.

In this paper, we derive a family of instrumental variable estimators that allow for direct empirical tests of the exclusion restriction, even with a single instrument. Our estimators extend the famous result of Geary (1942) to allow for omitted variable bias. In particular, our estimators use information in the higher order joint moments of the dependent variable, endogenous regressors, and exogenous instruments, generating an infinite number of equations for the parameter of interest with only mild assumptions. We derive asymptotic standard errors and a J test of over-identifying restrictions for our estimators.

Our estimators are based on higher-order cumulants that yield a system of linear equations in the parameters of interest. Thus, unlike many higher-order moment estimators, the estimators we derive are available in closed form. This makes them faster to compute, and avoids the problem of finding a global minimum to an objective based on a system of non-liner equations. Closed form solutions for our estimators also make bootstrapping easy and relatively fast, which we show is an important consideration when using higher-order cumulants with modest sample sizes.

In addition to testing the exclusion restriction, our estimators can recover the causal parameters themselves, even in cases where the instrument is invalid. In particular, if there is correlated measurement error in the endogenous regressors and instruments, then our cumulant equations can still be used to consistently estimate the parameters of interest; the same cannot be said of 2SLS, which fails to converge to the true parameters. As Pancost and Schaller (2022) show, empirical work in economics and finance often overlooks measurement error as a source of endogeneity, despite its quantitative importance. Our benchmark higher-order estimators allow researchers to test for invalidity; if that invalidity is caused by correlated measurement errors, then our higher-order estimators can recover the parameters of interest.

We perform a battery of Monte Carlo exercises to explore the performance of our estimators. We derive a bootstrap procedure that adapts our asymptotic test statistic to small samples, then we apply our bootstrap tests to IV regressions reported in Mian and Sufi (2014). We find that, despite debates on the validity of the Saiz housing-supply instrument (Davidoff, 2016), we cannot reject the hypothesis that the instrument is valid in these regressions.

2 Setting

For ease of exposition, we assume all variables have been de-meaned and describe the univariate case without controls. All proofs in the appendix are multivariate and allow for controls.

2.1 Cumulant equations

Suppose we are interested in estimating the effects of χ_i on y_i, where:

$$y_i = \chi_i \beta_x + \eta_i^y.$$

Instead of observing (y_i, χ_i) directly, we can only observe (y_i, x_i), were x_i is a proxy for χ_i:

$$x_i = \chi_i + u_i^x,$$

and $u_i^x \perp\!\!\!\perp \left(\eta_i^y, \chi_i\right)$ is unobserved, mean-zero measurement error.

Running an OLS regression of y_i on x_i fails to recover the coefficient of interest β_x. In fact, the OLS estimator converges to:

$$\text{plim } \hat{\beta}_{OLS} = \beta_x \left(1 - \frac{\text{var}\left(u_i^x\right)}{\text{var}\left(x_i\right)}\right) + \frac{\text{cov}\left(\chi_i, \eta_i^y\right)}{\text{var}\left(x_i\right)},$$

where $\text{var}\left(u_i^x\right) \neq 0$ is symptomatic of measurement error bias, and $\text{cov}\left(\chi_i, \eta_i^y\right) \neq 0$ omitted variable bias.

Geary (1942) attempts to discern an exact causal relationship by appealing to higher-order moments of (y_i, x_i), and shows that β_x can be recovered using cumulant equations, provided $\chi_i \perp\!\!\!\perp \eta_i^y$. In other words, assuming there are no omitted variables, cumulants can be used to extirpate the measurement error bias.

Proposition 1 (Geary)**.** *Let $\kappa\left(s_y, s_x\right)$ denote the cumulant of order $s_y > 0$ in y_i, and order*

$s_x > 0$ *in* x_i*. If* $\chi_i \perp\!\!\!\perp \eta_i^y$*, then:*

$$\kappa\left(s_y + 1, s_x\right) = \kappa\left(s_y, s_x + 1\right)\beta_x.$$

Because this result holds for all $s_y, s_x > 0$, these cumulant equations provide an infinite number of linear testable restrictions, subject to the assumption that there is no omitted variable bias, i.e., $\chi_i \perp\!\!\!\perp \eta_i^y$.

While cumulants reveal deep patterns in the relationship between y_i and x_i, and having an infinite number of testable equations is incredibly valuable, having to assume away omitted variables is likely to be a sticking point for research in economics and finance. Indeed, it is far more common for researchers to address endogeneity problems with instrumental variables instead of cumulant equations, possibly owing to the former's ability to simultaneously address both measurement error and omitted variable bias. We can introduce an instrument into this framework by supposing that:

$$\eta_i = \zeta_i \beta_z + \varepsilon_i^y,$$
$$\chi_i = \zeta_i \alpha_z + \varepsilon_i^x.$$

The variable ζ_i is a valid instrument provided it satisfies the relevance condition $(\alpha_z \neq 0)$ and exclusion restriction $\left(\beta_z = 0, \zeta_i \perp\!\!\!\perp \varepsilon_i^y\right)$. In fact, instrumental variables will still be able to recover β_x if we only observe a noisy proxy z_i of ζ_i:

$$z_i = \zeta_i + u_i^z,$$

where u_i^z is mean-zero measurement error that is mutually independent of $\left(\zeta_i, \varepsilon_i^y, \varepsilon_i^x, u_i^x\right)$.

Proposition 2 (IV)**.** *If the instrument* z_i *satisfies the relevance condition* $\alpha_z \neq 0$ *and exclusion restriction* $\beta_z = 0$*,* $\zeta_i \perp\!\!\!\perp \varepsilon_i^y$*, then:*

$$plim\ \hat{\beta}_{IV} = \beta_x.$$

Instruments are powerful tools for estimating causal effects, but identification still rests on the instrument's ability to satisfy both the relevance condition and the exclusion restriction, with the latter being far more contentious; in practice, the relevance condition can be tested, while the exclusion restriction is argued for on theoretical grounds.

In this paper, we show that the exclusion restriction *can* be tested by using the deeper information revealed by cumulant equations, provided those cumulant equations incorporate the instrument. Fundamentally, this is because an instrument–even a single instrument–can recover β_x in an infinite number of ways. These possibilities are captured in the cumulants of (y_i, x_i, z_i); hence, we can compute these cumulants, recover an estimate of β_x for each cumulant equation, and compare these estimates to each other.

First, we discuss how to recover β_x by using cumulants.

Proposition 3. *Let $\kappa\left(s_y, s_x, s_z\right)$ denote the cumulant of order $s_y \geq 0$ in y_i, $s_x \geq 0$ in x, and $s_z \geq 1$ in z_i. If the instrument z_i satisfies the relevance condition and exclusion restriction, then:*

$$\kappa\left(s_y, s_x, s_z\right) = \kappa\left(s_y, s_x + 1, s_z\right) \beta_x. \tag{1}$$

Proposition 3 provides a way to generate a system of linear equations in β_x, which can be solved in closed-form and used to construct straightforward estimators (Erickson, Jiang, & Whited, 2014). Despite many researcher's unfamiliarity with these equations, well-known estimators can be written in the form of cumulants. For example, OLS is a 2nd-order cumulant estimator:

$$\hat{\beta}_{OLS} = \frac{\kappa\,(1,1,0)}{\kappa\,(0,2,0)} = \frac{\mathbb{E}\,[y_i x_i]}{\mathbb{E}\left[x_i^2\right]},$$

which fails to consistently estimate β_x in the presence of either measurement error or omitted variable bias.

Two-stage least-squares (2SLS) is another 2nd-oder cumulant estimator:

$$\hat{\beta}_{OLS} = \frac{\kappa\,(1,0,1)}{\kappa\,(0,2,0)} = \frac{\mathbb{E}\,[y_i z_i]}{\mathbb{E}\,[x_i z_i]},$$

which consistently estimates β_x in the presence measurement error and omitted variable bias.

Geary's (1942) leading example is a 3rd-order cumulant estimator:

$$\hat{\beta}_{Geary} = \frac{\kappa\,(2,1,0)}{\kappa\,(1,2,0)} = \frac{\mathbb{E}\left[y_i^2 x_i\right]}{\mathbb{E}\left[y_i x_i^2\right]},$$

which fails to consistently estimate β_x in the presence of omitted variable bias.

Following Proposition 3, we can derive two other 3rd-order estimators:

$$\widehat{\beta}_{011} = \frac{\kappa\,(1,1,1)}{\kappa\,(0,2,1)} = \frac{\mathbb{E}\left[y_i^2 x_i\right]}{\mathbb{E}\left[y_i x_i^2\right]},$$
$$\widehat{\beta}_{101} = \frac{\kappa\,(2,0,1)}{\kappa\,(1,1,1)} = \frac{\mathbb{E}\left[y_i^2 x_i\right]}{\mathbb{E}\left[y_i x_i^2\right]},$$

each of which consistently estimates β_x in the presence measurement error and omitted variable bias. More broadly, there are $\binom{S}{2}$ cumulant estimators of order S that can be used to consistently estimate β_x.

Up to a 3rd-order, cumulants are equal to moments; that is, for $s_y + s_x + s_z \leq 3$, the cumulant of order s_y in y, s_x in x, and s_z in z_i is given by $\kappa\left(s_y, s_x, s_z\right) = \mathbb{E}\left[y_i^{s_y} x_i^{s_x} z_i^{s_z}\right]$. Cumulants of order four and higher instead correspond to combinations of the underlying moments. For example:

$$\kappa\left(2,1,1\right) = \mathbb{E}\left[y_i^2 x_i z_i\right] - \mathbb{E}\left[y_i^2\right]\mathbb{E}\left[x_i z_i\right] - 2\mathbb{E}\left[y_i x_i\right]\mathbb{E}\left[y_i z_i\right].$$

Extracting β_x from arbitrary combinations of moment equations would be, understandably, quite cumbersome; fortunately, the comparatively parsimonious cumulant equation of Proposition 3 extracts this information for us.

Cumulants can also be used to consistently estimate β_x when the instrument z_i is valid for addressing omitted variables bias but invalid for addressing measurement error. As Pancost and Schaller (2022) point out, most instruments in economics and finance are intended to address omitted variables problems, despite good reason to believe that there may be substantial measurement error bias induced by widespread use of proxy variables. If a researcher disregards measurement error, and selects an instrument that addresses omitted variable bias but not measurement error bias, then their resulting IV estimates will fail to recover β_x. However, higher-order cumulant estimators can rectify this problem for $s_y \geq 1$ and $s_z \geq 1$.

Proposition 4. *Let $\kappa\left(s_y, s_x, s_z\right)$ denote the cumulant of order $s_y \geq 1$ in y_i, $s_x \geq 0$ in x, and $s_z \geq 1$ in z_i. If the instrument z_i satisfies the relevance condition and exclusion restriction, but $u_i^x \not\perp\!\!\!\perp u_i^z$, then:*

$$\kappa\left(s_y, s_x, s_z\right) = \kappa\left(s_y, s_x + 1, s_z\right)\beta_x.$$

This shows that instruments that are invalid for measurement error, and hence invalid for 2SLS, can nevertheless be used to estimate β_x by way of cumulant equations. In fact, there are $\binom{S-1}{2}$ robust cumulant estimators of order S that can be used to consistently estimate β_x when the instrument z_i is invalid for measurement error.

Let $\mathbb{K}_y$ denote a vector of left-hand-side cumulants of the form:

$$\mathbb{K}_y = \begin{bmatrix} K\left(s_{1y}+1, s_{1x}, s_{1z}\right) \\ K\left(s_{2y}+1, s_{2x}, s_{2z}\right) \\ \vdots \\ K\left(s_{Qy}+1, s_{Qx}, s_{Qz}\right) \end{bmatrix},$$

where s_{qy} is a non-negative integer, and s_{qx} is a $1 \times J$ vector of non-negative integers, and s_{qz} is a $1 \times L$ vector of non-negative integers for $q = 1, ..., Q$. Let $\mathbb{K}_x$ denote the corresponding matrix of right-hand-side cumulants of the form:

$$\mathbb{K}_x = \begin{bmatrix} K\left(s_{1y}, s_{1x}+e_1, s_{1z}\right) & K\left(s_{1y}, s_{1x}+e_2, s_{1z}\right) & \dots & K\left(s_{1y}, s_{1x}+e_J, s_{1z}\right) \\ K\left(s_{2y}, s_{2x}+e_1, s_{2z}\right) & K\left(s_{2y}, s_{2x}+e_2, s_{2z}\right) & \dots & K\left(s_{2y}, s_{2x}+e_J, s_{2z}\right) \\ \dots & \vdots & \ddots & \vdots \\ K\left(s_{Qy}, s_{Qx}+e_1, s_{Qz}\right) & K\left(s_{Qy}, s_{Qx}+e_2, s_{Qz}\right) & \dots & K\left(s_{Qy}, s_{Qx}+e_J, s_{Qz}\right), \end{bmatrix}$$

where e_j denotes a $1 \times J$ vector of zeros whose jth element is 1, hence $\mathbb{K}_y = \mathbb{K}_x \beta_x$. For $Q > J$, this gives us an over-identified system of equations and hence test for over-identification.

Proposition 5. *The quadratic form of* $\sqrt{n}\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x \widehat{\beta}_x\right)$ *follows a chi-squared distribution:*

$$n\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x \widehat{\beta}_x\right)' \widehat{\Sigma}^{-1} \left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x \widehat{\beta}_x\right) \sim \chi_k^2. \tag{2}$$

The above expression will fail to converge to that chi-square distribution if $z_i \not\perp\!\!\!\perp \varepsilon_i^j$. Thus, Proposition 5 lets us test the exclusion restriction using an arbitrary number of equations.

3 Simulation results

In this section we perform a handful of Monte Carlo exercises to illustrate the estimators and over-identifying tests of Section 2. The benchmark model we simulate is

$$\begin{aligned} y &= \beta_x \chi + \beta_c c + \varepsilon_y, \\ \chi &= \alpha_z \zeta + \alpha_c c + \varepsilon_x \sqrt{1-\rho_\varepsilon^2} + \rho_\varepsilon \varepsilon_y \\ \zeta &= \varepsilon_z \sqrt{1-\rho_z^2} + \rho_z \varepsilon_y \end{aligned} \tag{3}$$

where ε_x, ε_y, and ε_z are all drawn from centered gamma distributions with variance and shape both equal to 1, c is drawn from a standard normal distribution, and the coefficients

$\alpha_z = \alpha_c = \beta_c = \beta_x = 1$. The parameter ρ_ε indexes the degree of omitted variable bias; when $\rho_\varepsilon \neq 0$, OLS estimation of β_x will be biased even though an IV estimator may not be. The parameter ρ_z indexes instrument validity by mechanically inducing correlation between the instrument ζ and the residual ε_y; when $\rho_z \neq 0$, ζ is an invalid instrument.

To model measurement error, we assume that only x and z are observed, where

$$x = \chi + u_x \qquad (4)$$
$$z = \zeta + u_z$$

where u_x and u_z are jointly normally distributed with correlation ρ_u. If $\rho_u \neq 0$, then z is an invalid instrument for x, even though ζ might still be valid if it were observable. We set the variances of u_x and u_z so that the coefficient of determination in both equations is τ^2.

Figure 1 reports the estimator bias and the power of the over-identification test for the benchmark simulations. The top panel plots the median value of $\hat{\beta}_x - \beta_x$, where each point corresponds to 10,000 simulations of equations (3) and (4). In order to roughly match the sample size in our replication of Mian and Sufi (2014) below, for each simulation we draw a sample of 500 observations, and we assign each observation randomly to one of 50 clusters. For this exercise, we set $\rho_\varepsilon = 0.3$, which biases the OLS estimator for all values of ρ_z. When $\rho_z = 0$, the three IV estimators—IV-2 being the usual two-stage least-squares estimator, and IV-3 and Iv-4 our third and fourth-order estimators—are all unbiased.

When $\rho_z \neq 0$, all the estimators in the top panel of Figure 1 are biased; but the IV-3 and IV-4 estimators allow for a test of the over-identifying restrictions. The bottom panel of Figure 1 reports the rejection rate for a 5% test of the over-identifying restrictions. When ρ_z is close to zero, the multiple equations for β_x underlying the IV-3 and IV-4 estimators all point to nearly the same estimate, and thus the rejection rate is low. As ρ_z gets further from zero, not only does the median bias increase, but the multiple equations of the form (1) start to diverge, leading to more rejections of the over-identifying test. From the figure, it appears that the IV-4 test is more powerful than the IV-3 test.

Unfortunately, the asymptotic test derived in Proposition 5 appears to be incorrectly sized. The bottom panel of Figure 1 plots rejection rates for a 5% test; thus when $\rho_z = 0$, the test should reject about 5% of the time, but in fact it rejects much more frequently than that. The problem is worse for the IV-4 than the IV-3.

The size problem evident in the bottom panel of Figure 1 is a result of the small sample size. Proposition 5 is an asymptotic result, that holds as the number of clusters goes to infinity; in Figure 1 we assumed only 50 clusters and 500 observations per sample. The top panel of Figure 2 plots the rejection rate, when $\rho_z = 0$, as the number of clusters increases

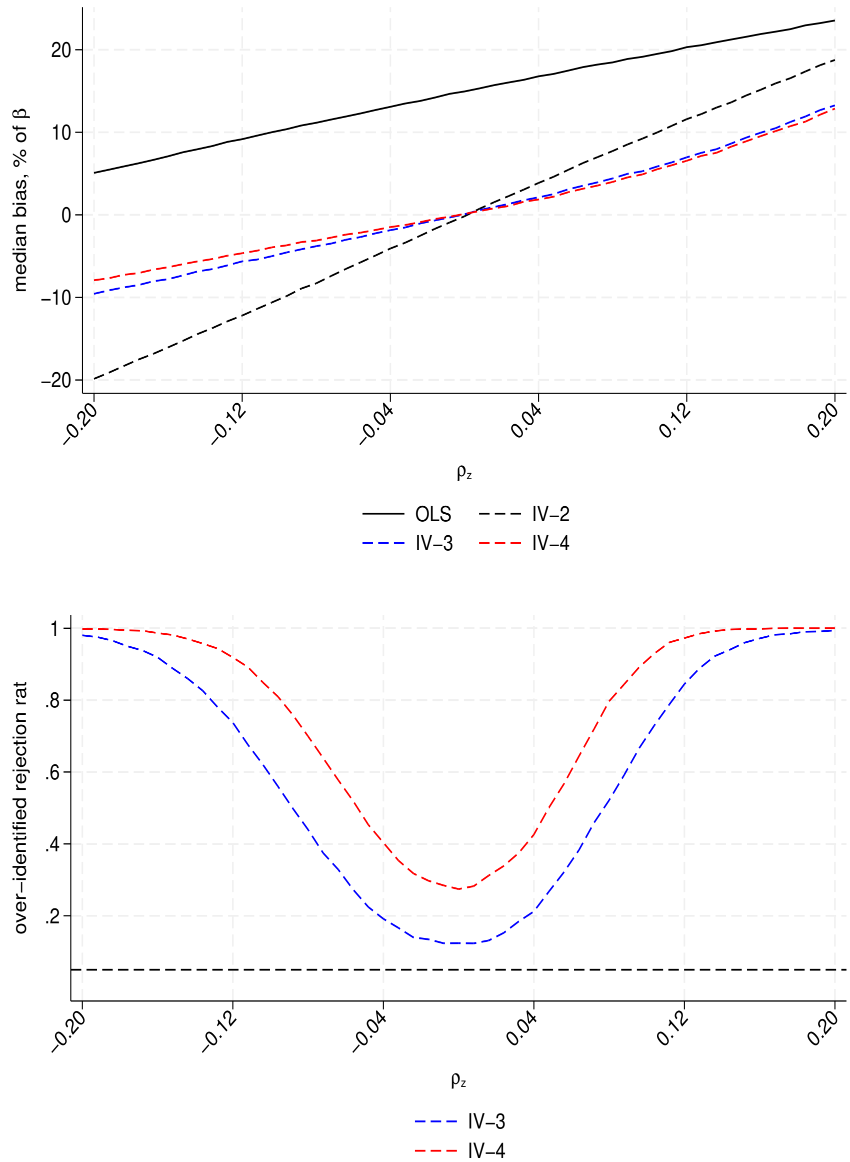

Figure 1. Bias and Rejection Rate for Invalid Instrument
The top panel plots the the median value of $\hat{\beta}_x - \beta$, for various values of the parameter ρ_z, which governs the correlation between the instrument z_i and the second-stage residual ε_i^y. For each value of ρ_z, we simulate 10,000 samples of 500 observations each. The bottom panel plots the rejection rate, at the 5% level, for the test of over-identifying restrictions. In the top panel, the solid black line reports the bias for the OLS estimator, and the dotted black line plots the bias for the IV estimator. In both panels, the dotted blue line denotes the IV-3 estimator, and the dotted red line the IV-4 estimator. The horizontal dotted black line in the bottom panel denotes 5%. In this figure, we set $\rho_\varepsilon = 0.3$, $\rho_u = 0$, and $\tau^2 = 0.3$.

from 50 to 100,000, assuming that the number of observations is ten times the number of clusters. For the IV-3 estimator, the test approaches the proper size as the number cluster

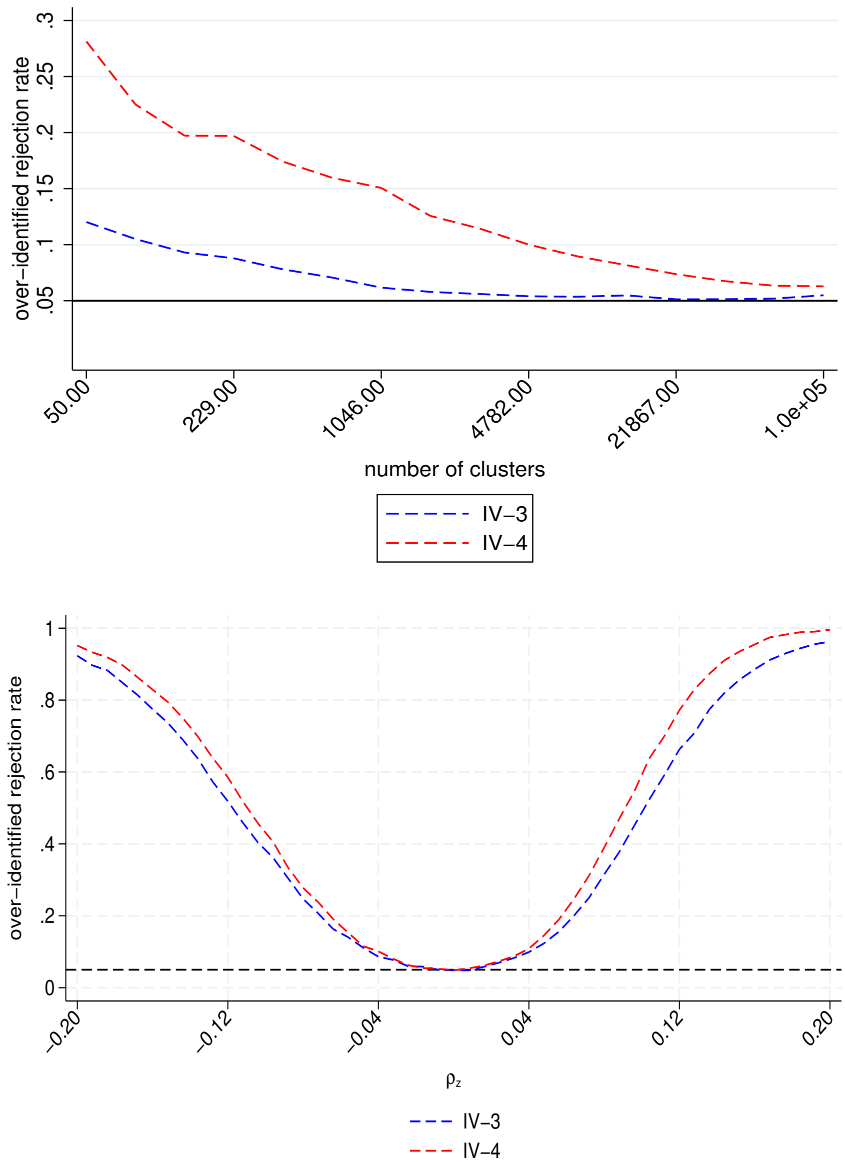

Figure 2. Bootstrapped Rejection Rate for Invalid Instrument
The figure plots the rejection rate for a bootstrap test of overidentifiyng restrictions, for various values of the parameter ρ_z, which governs the correlation between the instrument z_i and the second-stage residual ε_i^y. For each value of ρ_z, we simulate 10,000 samples of 500 observations each; for each sample we run 100 bootstrap replications. The dotted blue line plots the rejection rate for the IV-3 estimator, and the dotted red line plots the rejection rate for IV-4 estimator, both for a 5% test. The horizontal dotted black line in the bottom panel denotes 5%. In this figure, we set $\rho_\varepsilon = 0.3$, $\rho_u = 0$, and $\tau^2 = 0.3$.

reaches about 5,000; the IV-4 over-identification test continues to over-reject until the number of clusters approaches 100,000.

Since sample sizes of 1,000,000 are not feasible for many applications, we propose a bootstrap version of the over-identifying test, rather than the the asymptotic test of Proposition 5. Specifically, we construct bootstrap samples as follows: given an estimated $\hat{\beta}_x$, we estimate the vector of sample residuals. We then sample each cluster with replacement, and within each cluster, for each observation we draw a new residual from that cluster's sample of residuals, again with replacement. For each bootstrap sample we then re-estimate $\hat{\beta}_x$ and the over-identifying J-statistic. We use the former to compute bootstrapped standard errors, and the latter to compute a small-sample critical value to test for over-identification.

The results are plotted in the bottom panel of Figure 2, which reports the rejection rates for the bootstrap test for samples of 500 observations and 50 clusters each. This bootstrap test is now correctly sized, and although the IV-4 appears to be somewhat more powerful than the IV-3 test, it is not significantly so.

Another way that instruments can be invalid is for the measurement errors in the endogenous regressor and the exogenous instrument to be correlated. To explore how our estimators fare in this case, in Figure 3 we set $\rho_z = 0$ but allow $\rho_u \neq 0$. We set the coefficient of determination in equations (4) $\tau^2 = 0.3$, in accordance with the empirical result in Pancost and Schaller (2022) that OLS regression coefficients are, on average, about 3-4 times too small. With this degree of measurement error, OLS coefficients will generally be substantially smaller than IV coefficients, even though in the absence of measurement error they should be larger, since $\rho_\varepsilon > 0$.

The top panel of Figure 3 plots the median bias for the OLS and IV estimators as a function of the correlation between measurement errors. The OLS bias is very large and negative, and does not vary with ρ_u since the OLS estimator ignores the instrument. When $\rho_u = 0$, the instrumental variable estimators are close to unbiased; as ρ_u varies below zero, the IV-2 bias becomes extremely large and positive. For values of $\rho_u > 0$, the IV-2, IV-3, and IV-4 estimators all have a similar negative bias, that approaches the magnitude of the OLS bias. Interestingly, the IV-3 and IV-4 estimators are also biased downwards for values of $\rho_u < 0$. On the other hand, the robust IV-3 and robust Iv-4 estimators are close to unbiased, and their (small) bias does not vary with ρ_u.

The bottom panel of Figure 3 plots the rejection rate of the bootstrap over-identification test for the three higher-order estimators. The robust IV-4 estimator rejects about 5% of the time, as it should, because it is not biased as a result of the correlated measurement errors. The other two estimators, on the other hand, are more likely to reject as ρ_u varies away from 0. It is curious that the rejection rates are so low for higher values of $\rho_u > 0$, but relatively high for similarly-large negative correlations.

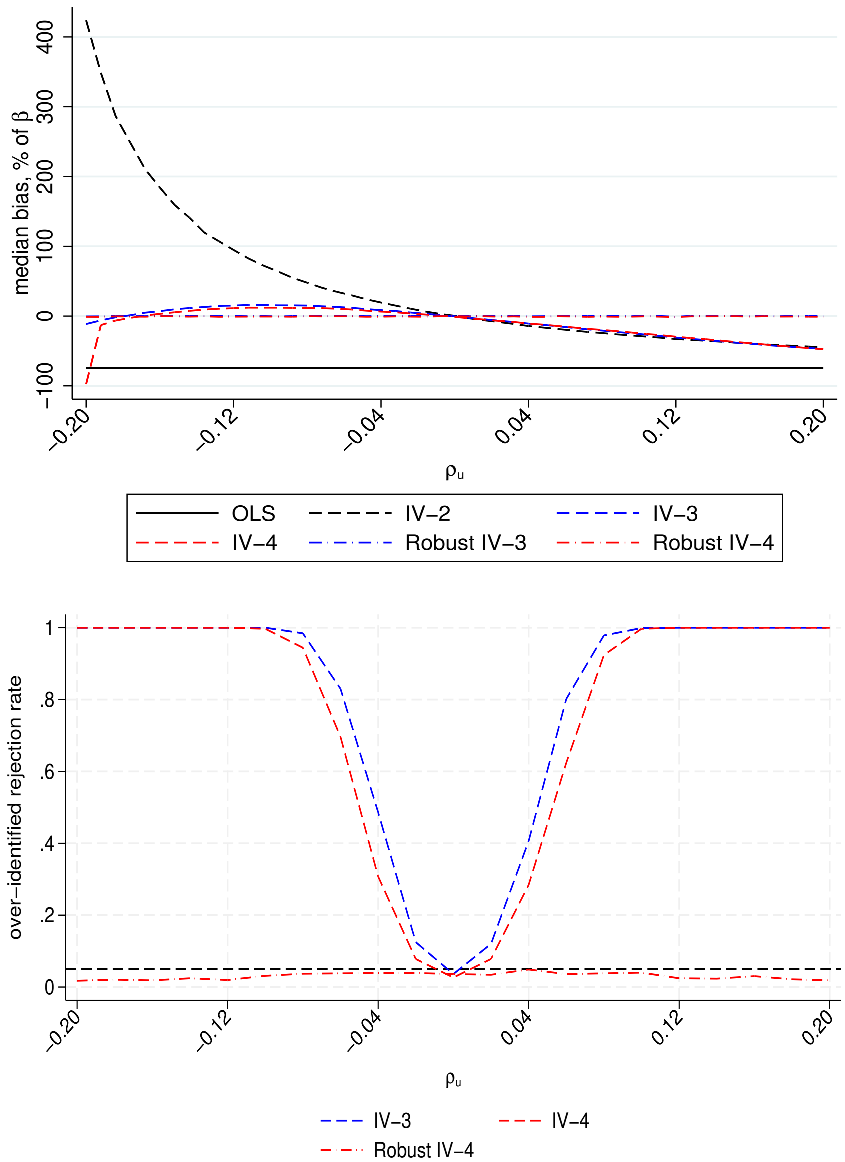

Figure 3. Bias and Rejection Rate for Correlated Measurement Errors
The top panel plots the the median value of $\hat{\beta}_x - \beta$, for various values of the parameter ρ_u, which governs the correlation between the measurement errors in equation (4). For each value of ρ_u, we simulate 10,000 samples of 500 observations each. The bottom panel plots the rejection rate, at the 5% level, for the bootstrap test of over-identifying restrictions. In the top panel, the solid black line reports the bias for the OLS estimator, and the dotted black line plots the bias for the IV estimator. In both panels, the dotted blue line denotes the IV-3 estimator, the dotted red line the IV-4 estimator, and the dash-dotted red line the robust IV-4 estimator. The horizontal dotted black line in the bottom panel denotes 5%. In this figure, we set $\rho_\varepsilon = 0.3$ and $\tau^2 = 0.3$.

4 Replication Results

In this section we apply our estimators to two instrumental variable regressions reported in Mian and Sufi (2014). Mian and Sufi (2014) is an ideal setting because it is well known, Amir Sufi has posted the data and replication code on his website, and (most important of all) the instrument that is uses—the Saiz (2010) housing supply elasticity—is both widely used in empirical work and somewhat controversial. In particular, both Saiz (2010) and Davidoff (2016) argue that housing supply elasticity is not a valid instrument because it is potentially correlated with housing demand.

The two regressions we examine here are in Table 3, Columns 5–6. Both regressions are of the form

$$\Delta \log \mathrm{E}_i^{\mathrm{NT}} = \alpha + \beta_x \Delta HNW_i + \varepsilon_i \tag{5}$$

where i indexes county, $\Delta \log \mathrm{E}_i^{\mathrm{NT}}$ denotes non-tradable employment growth, ΔHNW_i denotes Mian & Sufi's proxy for growth in household net worth, and both growth rates are computed from 2006–2009. The two estimates of equation (5) correspond to two different definitions of "non-tradeable," which are unimportant for our purposes.

Table 1 reports the results, where the estimates from the two definitions of non-tradeable" are reported in separate panels. The first row of each panel reports the two-stage least-squares estimates of β_x, which correspond to the published paper. We denote these estimates as "IV-2" while subsequent rows report various higher-order cumulant estimators. Both the t-statisics and the p-values in Table 1 are bootstrapped.

The most obvious result of Table 1 is that the higher-order cumulant estimates are all much larger than the published estimates. In fact, the IV-3 and IV-4 estimates are about twice as large, while the robust IV-4 estimates are even larger (the robust IV-3 are statistically insignificantly different from zero). Notice also that the published IV estimates of equation (5) are themselves much larger than the OLS estimates in Columns 3 and 4 (not reported here); see Pancost and Schaller (2022) for a discussion and potential explanation.

Despite the difference in magnitudes across the estimates reported in Table 1, in only one case can we reject the over-identifying restriction at conventional levels. The lack of rejection is somewhat surprising given the range of estimated β_x's; although there is a world of *economic* difference between an employment to net worth elasticity of 0.2 and 0.7, in a statistical sense there is not enough power to conclude that the multiple equations for β_x are pointing to different estimates. This is most likely a consequence of the extremely small sample size (540 observations divided into 48 state clusters); indeed, although in most cases we can reject that the coefficients are greater than zero, we cannot say much more than that.

Estimator	Coefficient	t-statistic	Over-identified p-value
Table 3, Column 5 (Retail & World Trade Definition)			
IV-2 (published)	0.37	(2.93)	.
IV-3	0.69	(3.63)	0.059
IV-4	0.67	(2.61)	0.56
Robust IV-3	-2.47	(0.00)	.
Robust IV-4	0.80	(1.23)	0.016
Table 3, Column 6 (Geographic Concentration Definition)			
IV-2 (published)	0.21	(2.50)	.
IV-3	0.40	(2.94)	0.413
IV-4	0.46	(2.00)	0.212
Robust IV-3	0.94	(0.25)	.
Robust IV-4	0.69	(1.67)	0.355

Table 1. Replication of Two IV Regressions from Mian and Sufi (2014)

The table reports multiple estimates of equation (5), from Table 3 of Mian and Sufi (2014). The top panel corresponds to the estimate in Column 5, while the bottom panel corresponds to the estimate in Column 6. The rows labeled "IV-2" report the published estimate in the paper; thwe rows labeled "IV-3" and "IV-4" report our 3rd- and 4th-order cumulant estimators (four and ten equations for β_x, respectively) while the rows labeled "Robust IV-3" and "Robust IV-4" report the 3rd-and 4th-order cumulant estimators that are consistent even in the case of correlated measurement errors (one and four equations for β_x, respectively). The first column reports the coefficient estimate, the second column reports the t-statistic, and the final column reports the p-value of the over-identification test. All regressions have 540 observations and 48 clusters. Both the t-statistics and the p-values are cluster bootstrapped as described in Section 3.

5 Conclusion

In this paper, we derive, test, and apply a family of instrumental variable estimators that yield tests of overidentifying restrictions even in the case of a single instrument. Because we base our estimators on higher-order cumulants, rather than higher-order moments, they are available in closed form, which allows for both rapid estimation and bootstrapping. In addition to deriving over-identifying tests, we show that a subset of our higher-order cumulant equations yield valid estimates for the parameters of interest, even when two-stage least-squares estimators are inconsistent due to correlation between the measurement errors of the regressor and the instrument.

We then apply our estimators to regressions reported in Mian and Sufi (2014), a paper whose instrument's validity is a subject of on-going debate, and find that for most estimators we cannot reject the hypothesis that the instrument is valid. On the other hand, our estimators yield much higher estimates of the coefficient of interest than those that are published.

We plan to apply our estimators to more published instrumental variable regressions across multiple fields. In addition, there are many additional Monte Carlo experiments we hope to run, including allowing for fixed effects, heteroskedasticity that correlates with the regressor and the instrument, varying the skewness of the instrument, and allowing for heterogeneous treatment effects.

References

Davidoff, Thomas. 2016. "Supply Constraints Are Not Valid Instrumental Variables For Home Prices Because They Are Correlated With Many Demand Factors." *Critical Finance Review*, 5(2): 177–206. 3, 13

Geary, R. C. 1942. "Inherent Relations Between Random Variables." *Proceedings of the Royal Irish Academy. Section A: Mathematical and Physical Sciences*, 47: 63–76. 2

Mian, Atif, and Amir Sufi. 2014. "What Explains The 2007–2009 Drop In Employment?" *Econometrica*, 82(6): 2197–2223. 1, 3, 8, 13, 14, 15

Pancost, N. Aaron, and Garrett Schaller. 2022. "Measuring Measurement Error." Working paper, University of Texas at Austin McCombs School of Business. 2, 6, 11, 13

Saiz, Albert. 2010. "The Geographic Determinants Of Housing Supply." *The Quarterly Journal of Economics*, 125(3): 1253–1296. 13

A Proofs

A.1 Proof of Proposition 3 and Proposition 4

Let (y_i, x_i, z_i, w_i) denote a sequence of observable vectors for $i = 1, 2, ..., n$, where y_i is the dependent variable of interest, $x_i = (x_{i1}, x_{i2}, ..., x_{iJ})$ the vector of independent variables of interest, $z_i = (z_{i1}, z_{i2}, ..., z_{iL})$ the vector of instruments with $L \geq J$, and $w_i = (1, w_{i1}, w_{i2}, ..., w_{iM})$ the vector of perfectly-measured i.i.d. controls. These observed vectors are related to each other through the unobserved coefficient matrices β_x, β_z, β_w, α_z and α_w, as well as the i.i.d. sequence of unobservable vectors $(\chi_i, \zeta_i, \varepsilon_i, u_i)$ for $i = 1, 2, ..., n$ as follows:

$$
\begin{aligned}
y_i &= \chi_i \beta_x + \zeta_i \beta_z + w_i \beta_w + \varepsilon_i^y, \\
\chi_i &= \zeta_i \alpha_z + w_i \alpha_w + \varepsilon_i^x, \\
x_i &= \chi_i + u_i^x, \\
z_i &= \zeta_i + u_i^z,
\end{aligned}
$$

where $\varepsilon_i = (\varepsilon_i^y, \varepsilon_i^x)$ and $u_i = (u_i^x, u_i^z)$, $\mathbb{E}\left[\varepsilon_i^y\right] = \mathbb{E}\left[\varepsilon_i^x\right] = \mathbb{E}\left[u_i^x\right] = \mathbb{E}\left[u_i^z\right] = 0$, $\varepsilon_i^y \not\perp\!\!\!\perp \varepsilon_i^x$, and $u_i^x \not\perp\!\!\!\perp u_i^z$. We further assume that the unobserved vector has finite moments of every order. Our goal is to estimate the coefficient vector $\beta_x = (\beta_{x1}, \beta_{x2}, ..., \beta_{xJ})'$, which we assume lies on the interior of a compact set.

To make our estimating equations ultimately more intuitive, we rewrite the observable variables as residuals from linear regressions of (y_i, x_i, z_i) on the vector of control variables w_i. Define:

$$
\begin{aligned}
\dot{y}_i &\equiv y_i - w_i c_y \\
\dot{x}_i &\equiv x_i - w_i c_x, \\
\dot{z}_i &\equiv z_i - w_i c_z,
\end{aligned}
$$

where

$$
\begin{aligned}
c_y &\equiv \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' y_i\right] \\
&= \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' \chi_i\right] \beta_x + \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' \zeta_i\right] \beta_z + \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' w_i\right] \beta_w \\
&= c_x \beta_x + c_z \beta_z + \beta_w, \\
c_x &\equiv \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' x_i\right] \\
&= \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' \zeta_i\right] \alpha_z + \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' w_i\right] \alpha_w \\
&= c_z \alpha_z + \alpha_w, \\
c_z &\equiv \mathbb{E}\left[w_i' w_i\right]^{-1} \mathbb{E}\left[w_i' z_i\right].
\end{aligned}
$$

Our original model is therefore equivalent to:

$$
\begin{aligned}
\dot{y}_i &= \dot{\chi}_i \beta_x + \dot{\zeta}_i \beta_z + \varepsilon_i^y, \\
\dot{\chi}_i &= \dot{\zeta}_i \alpha_z + \varepsilon_i^x, \\
\dot{x}_i &= \dot{\chi}_i + u_i^x, \\
\dot{z}_i &= \dot{\zeta}_i + u_i^z,
\end{aligned}
$$

where $\dot{\chi}_i = \chi_i - w_i c_x$ and $\dot{\zeta}_i = z_i - w_i c_z$. Since w_i contains the constant 1, the variables in this residualized model are all mean-zero.

Let $K\left(t_y, t_x, t_z\right)$ denote the cumulant-generating function of the observable variables $(\dot{y}_i, \dot{x}_i, \dot{z}_i)$, where t_y is a scalar, $t_x = (t_{x1}, t_{x2}, ..., t_{xJ})$, and $t_z = (t_{z1}, t_{z2}, ..., t_{zL})$. We can decompose the cumulant-generating function of the observable variables into cumulant-generating functions of the unobserved variables as follows:

$$
\begin{aligned}
K\left(t_y, t_x, t_z\right) &\equiv \ln\left(\mathbb{E}\left[e^{\dot{y}_i t_y + \dot{x}_i t_x' + \dot{z}_i t_z'}\right]\right) \\
&= \ln\left(\mathbb{E}\left[e^{\sum_l \dot{\zeta}_{il}\left(\left(\sum_j \alpha_{zlj}\beta_{xj} + \beta_{zl}\right)t_y + \sum_j \alpha_{zlj} t_{xj} + t_{zl}\right)}\right] \mathbb{E}\left[e^{\varepsilon_i^y t_y + \sum_J \varepsilon_{ij}^x \beta_{xj} t_{xj}}\right] \mathbb{E}\left[e^{\sum_j u_{ij}^x t_{xj} + \sum_l u_{il}^z t_{zl}}\right]\right) \\
&- K_\zeta\left(t_y\left(\alpha_z \beta_x + \beta_z\right)' + t_x \alpha_z' + t_z\right) + K_\varepsilon\left(t_y, t_y \beta_x' + t_x\right) + K_u\left(t_x, t_z\right),
\end{aligned}
$$

where $K_\zeta\left(s_\zeta\right)$ denotes the cumulant-generating function of $\dot{\zeta}_i$, $K_\varepsilon\left(t_{\varepsilon y}, t_{\varepsilon x}\right)$ the joint cumulant-generating function of $\left(\varepsilon_i^y, \varepsilon_i^x\right)$, and $K_u\left(t_{ux}, t_{uz}\right)$ the joint cumulant-generating function of (u_i^x, u_i^z).

Given the separability of the cumulant-generating function above, we can use the change of variables

$$t_v \equiv t_y \left(\alpha_z \beta_x + \beta_z\right)' + t_x \alpha_z' + t_z$$

to obtain the following result:

$$\kappa\left(s_y, s_x, s_z\right) \equiv \left[\frac{\partial^{s_y+\sum_j s_{xj}+\sum_l s_{zl}} K\left(t_y, t_x, t_z\right)}{\partial t_y^{s_y} \prod_j \partial t_{xj}^{s_{xj}} \prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_y, t_x, t_z\right)=0}$$

$$= \left[\frac{\partial^{s_y+\sum_j s_{xj}+\sum_l s_{zl}} K_\zeta\left(t_y \left(\alpha_z \beta_x + \beta_z\right)' + t_x \alpha_z' + t_z\right)}{\partial t_y^{s_y} \prod_j \partial t_{xj}^{s_{xj}} \prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_y, t_x, t_z\right)=0}$$
$$+ \left[\frac{\partial^{s_y+\sum_j s_{xj}+\sum_l s_{zl}} K_\varepsilon\left(t_y, t_y \beta_x' + t_x\right)}{\partial t_y^{s_y} \prod_j \partial t_{xj}^{s_{xj}} \prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_y, t_x, t_z\right)=0}$$
$$+ \left[\frac{\partial^{s_y+\sum_j s_{xj}+\sum_l s_{zl}} K_u\left(t_x, t_z\right)}{\partial t_y^{s_y} \prod_j \partial t_{xj}^{s_{xj}} \prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_y, t_x, t_z\right)=0}$$

$$= \left[\frac{\partial^{s_y+\sum_j s_{xj}+\sum_l s_{zl}} K_\zeta\left(t_v\right)}{\partial t_v^{s_y+\sum_j s_{xj}+\sum_l s_{zl}}} \frac{\partial^{s_y+\sum_j s_{xj}+\sum_l s_{zl}} t_v}{\partial t_y^{s_y} \prod_j \partial t_{xj}^{s_{xj}} \prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_v, t_y, t_x, t_z\right)=0}$$
$$+ \left[\frac{\partial^{\sum_l s_{zl}}}{\prod_l \partial t_{zl}^{s_{zl}}} \left(\frac{\partial^{s_y+\sum_j s_{xj}} K_\varepsilon\left(t_y, t_y \beta_x' + t_x\right)}{\partial t_y^{s_y} \prod_j \partial t_{xj}^{s_{xj}}}\right)\right]_{\left(t_y, t_x, t_z\right)=0}$$
$$+ \left[\frac{\partial^{s_y}}{\partial t_y^{s_y}} \left(\frac{\partial^{\sum_j s_{xj}+\sum_l s_{zl}} K_u\left(t_x, t_z\right)}{\prod_j \partial t_{xj}^{s_{xj}} \prod_l \partial t_{zl}^{s_{zl}}}\right)\right]_{\left(t_y, t_x, t_z\right)=0}.$$

If $\beta_z = 0$, $s_z > 0$, and either $u_i^x \perp\!\!\!\perp u_i^z$ or $s_y > 0$, we have that:

$$\kappa\left(s_y+1, s_x, s_z\right) \equiv \left[\frac{\partial^{s_y+1+\sum_j s_{xj}+\sum_l s_{zl}} K\left(t_y, t_x, t_z\right)}{\partial t_y^{s_y+1}\prod_j \partial t_{xj}^{s_{xj}}\prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_y,t_x,t_z\right)=0}$$

$$= \left[\frac{\partial^{s_y+1+\sum_j s_{xj}+\sum_l s_{zl}} K_\zeta\left(t_y\left(\alpha_z\beta_x+\beta_z\right)'+t_x\alpha_z'+t_z\right)}{\partial t_y^{s_y+1}\prod_j \partial t_{xj}^{s_{xj}}\prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_y,t_x,t_z\right)=0}$$

$$= \sum_j{}' \beta_{xj'}\left[\frac{\partial^{s_y+\sum_{j\neq j'} s_{xj}+s_{xj'}+1+\sum_l s_{zl}} K_\zeta\left(t_y\left(\alpha_z\beta_x+\beta_z\right)'+t_x\alpha_z'+t_z\right)}{\partial t_y^{s_y+1}\prod_{j\neq j'} \partial t_{xj}^{s_{xj}}\partial t_{xj'}^{s_{xj'}}\prod_l \partial t_{zl}^{s_{zl}}}\right]_{\left(t_y,t_x,t_z\right)=0}$$

$$= \sum_j{}' \beta_{xj'}\kappa\left(s_y, s_x+e_j', s_z\right).$$

Proof of Proposition 5

To derive the distribution of our J-test, we must first derive the distribution of each of our estimators, starting with the residualizing coefficients $c := \text{vec}\left(c_y, c_x, c_z\right)$. By the mean value theorem, Lindeberg-Lévy central limit theorem, and Slutsky's theorem, we have that:

$$\begin{aligned}\sqrt{n}\left(\hat{c}-c\right) &= -I_{J+L+1}\otimes\mathbb{E}\left[w_i'w_i\right]^{-1}\frac{1}{\sqrt{n}} \\ &\quad \sum_{i=1}^{n}\text{vec}\left(w_i'\left(y_i-w_ic_y\right), w_i'\left(x_i-w_ic_x\right), w_i'\left(z_i-w_ic_z\right)\right)+o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i\left(c\right)+o_p(1) \\ &\xrightarrow{d} \mathbb{N}\left(0, \mathbb{E}\left[\psi_i\left(c\right)\psi_i\left(c\right)'\right]\right)\end{aligned}$$

where $\psi_i\left(c\right) = -I_{J+L+1}\otimes\mathbb{E}\left[w_i'w_i\right]^{-1}\text{vec}\left(w_i'\left(y_i-w_ic_y\right), w_i'\left(x_i-w_ic_x\right), w_i'\left(z_i-w_ic_z\right)\right)$.

We now turn to the distribution of the moments used to estimate our cumulant vectors. Let $h_i(c)$ denote an $R\times 1$ vector with elements of the form $\dot{y}_i^{r_y}\prod_j \dot{x}_{ij}^{r_{xj}}\prod_l \dot{z}_{il}^{r_{zl}}$, and define $\hat{h}_n\left(c\right) \equiv \frac{1}{n}\sum_{n=1}^{n} h_i\left(c\right)$. By the mean value theorem, Lindeberg-Lévy central limit theorem, and Slutsky's theorem, we have that:

$$\begin{aligned}\sqrt{n}\left(\hat{h}_n\left(\hat{c}\right)-\mathbb{E}\left[h_i\left(c\right)\right]\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(h_i\left(c\right)-\mathbb{E}\left[h_i\left(c\right)\right]+H\left(c\right)\psi_i\left(c\right)\Big)+o_p(1)\\ &\xrightarrow{d}\mathbb{N}\left(0,\Omega\right)\end{aligned}$$

where $H(c)\equiv\frac{1}{n}\sum_{i=1}^{n}\nabla_c h_i\left(c\right)$ and $\Omega\equiv\text{var}\Big[h_i\left(c\right)-\mathbb{E}\left[h_i\left(c\right)\right]+H\left(c\right)\psi_i\left(c\right)\Big]$.

Next, we turn to the distribution of the estimated over-identified cumulant equation $\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\beta_x$. By the mean value theorem, Lindeberg-Lévy central limit theorem, and Slutsky's theorem, we have that:

$$\begin{aligned}\sqrt{n}\left(\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\beta_x\right) &= \left[I_Q,-I_Q\otimes\beta_x'\right]D\sqrt{n}\left(\hat{h}_n\left(\hat{c}\right)-\mathbb{E}\left[h_i\left(c\right)\right]\right)+o_p(1)\\ &\xrightarrow{d}\mathbb{N}\left(0,\Sigma\right)\end{aligned}$$

where $D\equiv\partial\text{vec}\left(\mathbb{K}_y,\mathbb{K}_x\right)/\partial\mathbb{E}\left[h_i\left(c\right)\right]'$ and $\Sigma\equiv\left[I_Q,-I_Q\otimes\beta_x'\right]D\Omega D\left[I_Q,-I_Q\otimes\beta_x'\right]'$.

By construction, $\widehat{\beta}_x=\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_y$, hence:

$$\begin{aligned}\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\widehat{\beta}_x &= \widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\left(\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_y\right)\\ &= \left(I_Q-\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)\widehat{\mathbb{K}}_y\\ &= \left(I_Q-\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)\left(\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\beta_x+\widehat{\mathbb{K}}_x\beta_x\right)\\ &= \left(I_Q-\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)\left(\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\beta_x\right)\\ &\quad+\left(I_Q-\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)\widehat{\mathbb{K}}_x\beta_x\\ &= \left(I_Q-\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)\left(\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\beta_x\right)+\left(\widehat{\mathbb{K}}_x-\widehat{\mathbb{K}}_x\right)\beta_x\\ &= \left(I_Q-\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)\left(\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\beta_x\right).\end{aligned}$$

Recall that $\text{avar}\left(\widehat{\mathbb{K}}_y-\widehat{\mathbb{K}}_x\beta_x\right)\equiv\Sigma$, and let S be a $Q\times Q$ nonsingular matrix such that $\Sigma=SS'$. If we use an efficient weighting matrix $\widehat{\Sigma}^{-1}\xrightarrow{p}\Sigma^{-1}$, and pre-multiply the first term in the above expression by $\widehat{S}^{-1}$ where $\widehat{S} \xrightarrow{p} S$, then we obtain the expression:

$$
\begin{aligned}
&\widehat{S}^{-1}\left(I_Q - \widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{\Sigma}^{-1}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{\Sigma}^{-1}\right)\\
&= \widehat{S}^{-1}\left(I_Q - \widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\left(\widehat{S}\widehat{S}'\right)^{-1}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\left(\widehat{S}\widehat{S}'\right)^{-1}\right)\\
&= \widehat{S}^{-1}\left(I_Q - \widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\left(\widehat{S}'\right)^{-1}\widehat{S}^{-1}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\left(\widehat{S}'\right)^{-1}\widehat{S}^{-1}\right)\\
&= \left(\widehat{S}^{-1} - \widehat{S}^{-1}\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\left(\widehat{S}'\right)^{-1}\widehat{S}^{-1}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\left(\widehat{S}'\right)^{-1}\widehat{S}^{-1}\right)\\
&= \left(I_Q - \widehat{S}^{-1}\widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\left(\widehat{S}'\right)^{-1}\widehat{S}^{-1}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\left(\widehat{S}'\right)^{-1}\right)\widehat{S}^{-1}\\
&=: \left(I_Q - A\left(A'A\right)^{-1}A\right)\widehat{S}^{-1},
\end{aligned}
$$

where $A \equiv \widehat{S}^{-1}\widehat{\mathbb{K}}_x$. Importantly, $\left(I_Q - A\left(A'A\right)^{-1}A\right)$ is symmetric and idempotent, hence $\mathrm{tr}\left(I_Q - A\left(A'A\right)^{-1}A\right) = Q - \mathrm{rank}\left(A\right) \equiv k$.

Furthermore, note that:

$$
\begin{aligned}
\widehat{S}^{-1}\sqrt{n}\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\beta_x\right) &\xrightarrow{d} \mathbb{N}\left(0, S^{-1}\Sigma\left(S'\right)^{-1}\right)\\
&= \mathbb{N}\left(0, S^{-1}SS'\left(S'\right)^{-1}\right)\\
&= \mathbb{N}\left(0, I_Q\right).
\end{aligned}
$$

Our test statistic for over-identification therefore takes the form:

$$
\begin{aligned}
&n\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\widehat{\beta}_x\right)'\widehat{\Sigma}^{-1}\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\widehat{\beta}_x\right)\\
&= n\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\beta_x\right)'\left(I_Q - \widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)'\widehat{\Sigma}^{-1}\left(I_Q - \widehat{\mathbb{K}}_x\left(\widehat{\mathbb{K}}_x'\widehat{W}\widehat{\mathbb{K}}_x\right)^{-1}\widehat{\mathbb{K}}_x'\widehat{W}\right)\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\beta_x\right)\\
&= n\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\beta_x\right)'\left(\widehat{S}'\right)^{-1}\left(I_Q - A\left(A'A\right)^{-1}A\right)\widehat{S}^{-1}\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\beta_x\right)\\
&= \left(\widehat{S}^{-1}\sqrt{n}\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\beta_x\right)\right)'\left(I_Q - A\left(A'A\right)^{-1}A\right)\left(\widehat{S}^{-1}\sqrt{n}\left(\widehat{\mathbb{K}}_y - \widehat{\mathbb{K}}_x\beta_x\right)\right)\\
&\xrightarrow{d} \chi_k^2.
\end{aligned}
$$