UNBIASED INSTRUMENTAL VARIABLES ESTIMATION UNDER KNOWN FIRST-STAGE SIGN

By

Isaiah Andrews and Timothy B. Armstrong

February 2015

COWLES FOUNDATION DISCUSSION PAPER NO. 1984

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Unbiased Instrumental Variables Estimation Under Known First-Stage Sign*

Isaiah Andrews
Harvard Society of Fellows †

Timothy B. Armstrong
Yale University ‡

January 26, 2015

Abstract

We derive mean-unbiased estimators for the structural parameter in instrumental variables models where the sign of one or more first stage coefficients is known. In the case with a single instrument, the unbiased estimator is unique. For cases with multiple instruments we propose a class of unbiased estimators and show that an estimator within this class is efficient when the instruments are strong while retaining unbiasedness in finite samples. We show numerically that unbiasedness does not come at a cost of increased dispersion: in the single instrument case, the unbiased estimator is less dispersed than the 2SLS estimator. Our finite-sample results apply to normal models with known variance for the reduced-form errors, and imply analogous results under weak instrument asymptotics with an unknown error distribution.

*Preliminary version

†email: iandrews@fas.harvard.edu

‡email: timothy.armstrong@yale.edu

1 Introduction

Researchers often have strong prior beliefs about the sign of the first stage coefficient in instrumental variables models, to the point where the sign can reasonably be treated as known. This paper shows that knowledge of the sign of the first stage coefficient allows us to construct an estimator for the coefficient on the endogenous regressor which is unbiased in finite samples when the reduced form errors are normal with known variance. When the distribution of the reduced form errors is unknown, our results lead to estimators that, in contrast to the usual 2SLS estimator (or, indeed, any other estimator that does not impose a first stage sign restriction), are asymptotically unbiased under weak IV sequences as defined in Staiger & Stock (1997).

The possibility of unbiased estimation stands in sharp contrast to the case where the first stage parameter is unrestricted, where unbiased estimation is impossible (Hirano & Porter 2015). We show that the unbiased estimators introduced in this paper have several desirable properties. In the case with a single instrumental variable, the unbiased estimator is unique, and is less dispersed than the usual two-stage least squares (2SLS) estimator in finite samples. Under standard ("strong instrument") asymptotics, the unbiased estimator has the same asymptotic distribution as the 2SLS estimator. In cases with multiple instrumental variables whose first stage sign is known we propose a class of unbiased estimators, and find a feasible estimator within this class which is asymptotically efficient when instruments are strong. Thus finite sample unbiasedness does not come at the cost of asymptotic efficiency, and in fact reduces finite sample dispersion relative to 2SLS in the case with a single excluded instrument.

The remainder of the paper is organized as follows. The rest of this section discusses the assumption of known first stage sign, introduces the setting and notation, and briefly reviews the related literature. Section 2 introduces the unbiased estimator in the case of a single excluded instrument. Section 3 treats the case with multiple instruments and introduces an estimator that is asymptotically efficient when the instruments are strong while maintaining unbiasedness in finite samples. Section 4 presents simulation results for the case with a single instrument. Proofs and auxiliary results are given in an appendix.

1.1 Knowledge of the First-Stage Sign

The results in this paper rely on knowledge of the first stage sign. This is reasonable in many economic contexts. In their study of schooling and earnings, for instance, Angrist & Krueger (1991) note that compulsory schooling laws in the United States allow those born earlier in the year to drop out after completing fewer years of school than those born later in the year. Arguing that season of birth can reasonably be excluded from a wage equation, they use this fact to motivate season of birth as an instrument for schooling. In this context, a sign restriction on the first stage amounts to an assumption that the mechanism claimed by Angrist & Krueger works in the expected direction: those born earlier in the year tend to drop out earlier. More generally, empirical researchers often have some mechanism in mind for why a model is identified at all (i.e. why the first stage coefficient is nonzero) that leads to a known sign for the direction of this mechanism (i.e. the sign of the first stage coefficient).

In settings with heterogeneous treatment effects, a first stage monotonicity assumption is often used to interpret instrumental variables estimates (see Imbens & Angrist 1994, Heckman et al. 2006). In the language of Imbens & Angrist (1994), the monotonicity assumption requires that either the entire population be composed of "compliers," or that the entire population be composed of "defiers." Once this assumption is made, our assumption that the sign of the first stage coefficient is known amounts to assuming the researcher knows which of these possibilities (compliers or defiers) holds. Indeed, in the examples where they argue that monotonicity is plausible (involving draft lottery numbers in one case and intention to treat in another), Imbens & Angrist (1994) argue that all individuals are "compliers" for a certain definition of the instrument.

It is important to note, however, that knowledge of the first stage sign is not always a reasonable assumption, and thus that the results of this paper are not always applicable.

In settings where the instrumental variables are indicators for groups without a natural ordering, for instance, one typically does not have prior information about signs of the first stage coefficients. To give one example, Aizer & Doyle Jr. (2013) use the fact that judges are randomly assigned to study the effects of prison sentences on recidivism. In this setting, knowledge of the first stage sign would require knowing a priori which judges are more strict.

1.2 Setting

For the remainder of the paper, we suppose that we observe a sample of T observations (Y_t, X_t, Z_t'), $t = 1, ..., T$ where Y_t is an outcome variable, X_t is a scalar endogenous regressor, and Z_t is a $k \times 1$ vector of instruments. Let Y and X be $T \times 1$ vectors with row t equal to Y_t and X_t respectively, and let Z be a $T \times k$ matrix with row t equal to Z_t'. The usual linear IV model, written in reduced-form, is

$$\begin{aligned} Y &= Z\pi\beta + U \\ X &= Z\pi + V \end{aligned} . \tag{1}$$

We treat the instruments Z as fixed and assume that the errors (U, V) are jointly normal with mean zero and known variance-covariance matrix $Var\left((U', V')'\right)$.[1] As is standard (see, for example, D. Andrews et al. (2006)), in contexts with additional exogenous regressors W (for example an intercept), we define Y, X, Z as the residuals after projecting out these exogenous regressors. If we denote the reduced-form and first-stage regression coefficients by ξ_1 and ξ_2, respectively, we can see that

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1} Z'Y \\ (Z'Z)^{-1} Z'X \end{pmatrix} \sim N\left(\begin{pmatrix} \pi\beta \\ \pi \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right) \tag{2}$$

for

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \left((Z'Z)^{-1} Z' \otimes I_2\right) Var\left((U', V')'\right) \left((Z'Z)^{-1} Z' \otimes I_2\right)',$$

[1] Note that we assume a homogenous β, which will generally rule out heterogenous treatment effect models with multiple instruments.

and (ξ_1, ξ_2) are sufficient for (π, β). Thus, going forward we will consider estimation based solely on these statistics. We assume that the sign of each component π_i of π is known, and in particular assume that the parameter space for (β, π) is

$$\Theta = \left\{ (\beta, \pi) : \beta \in B, \pi \in \Pi \subseteq (0, \infty)^k \right\} \tag{3}$$

for some sets B and Π. Note that once we take the sign of π_i to be known, assuming $\pi_i > 0$ is without loss of generality, since this can always be ensured by redefining Z.

In this paper we focus on models with fixed instruments, normal errors, and known error covariance, which allows us to obtain finite-sample results. As usual, these finite-sample results will imply asymptotic results under mild regularity conditions. Even in models with random instruments, non-normal errors, serial correlation, heteroskedasticity, clustering, or any combination of the above, the reduced-form and first stage estimators will be jointly asymptotically normal with consistently estimable covariance matrix Σ under mild regularity conditions. Consequently, the finite-sample results we develop here will imply asymptotic results under both weak and strong instrument asymptotics, where we simply define (ξ_1, ξ_2) as above and replace Σ by an estimator for the variance of ξ to obtain feasible statistics.[2] We omit these derivations here and focus on what we view as the most novel component of the paper: finite-sample mean-unbiased estimation of β in the normal problem (2).

1.3 Related Literature

Our unbiased IV estimators build on results for unbiased estimation of the inverse of a normal mean discussed in Voinov & Nikulin (1993). More broadly, the literature has considered unbiased estimators in numerous other contexts, and we refer the reader to

[2]The feasible analogs of the finite-sample unbiased estimators discussed here will be asymptotically unbiased in general models in the sense of converging in distribution to random variables with mean β. Note, however, that outside the exact normal case it will not in general be true that means of the feasible estimators themselves will converge to β as the sample size increases, since convergence in distribution does not suffice for convergence of moments.

Voinov & Nikulin for details and references. To our knowledge the only other paper to treat finite sample unbiased estimation in IV models is Hirano & Porter (2015), who find that unbiased estimators do not exist when the parameter space is unrestricted. The nonexistence of unbiased estimators has been noted in other nonstandard econometric contexts by Hirano & Porter (2012).

The broader literature on the finite sample properties of IV estimators is huge: see Phillips (1983) and Hillier (2006) for references. While this literature does not study unbiased estimation in finite samples, there has been substantial research on higher order asymptotic bias properties, see e.g. Hahn et al. (2004) and references therin.

Our interest in finite sample results for a normal model with known reduced form variance is motivated by the weak IV literature, where this model arises asymptotically under weak IV sequences as in Staiger & Stock (1997). In contrast to Staiger & Stock, however, our results allow for heteroskedastic, clustered, or serially correlated errors as in Kleibergen (2007). The primary focus of the recent work on weak instruments has, however, been on inference rather than estimation. See Andrews (2014) for references.

Sign restrictions have been used in other settings in the econometrics literature, although the focus is often on inference or on using sign restrictions to improve population bounds, rather than estimation. Recent examples include Moon et al. (2013) and several papers cited therein, which use sign restrictions to partially identify vector autoregression models. Inference for sign restricted parameters has been treated by D. Andrews (2001) and Gouriéroux et al. (1982), among others.

2 Unbiased Estimation with a Single Instrument

To introduce our unbiased estimators, we first focus on the just-identified model with a single instrument, $k = 1$. In this context ξ_1 and ξ_2 are scalars and we write

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

In the just-identified setting, the problem of estimating β reduces to that of estimating

$$\beta = \frac{\pi\beta}{\pi} = \frac{E[\xi_1]}{E[\xi_2]}. \tag{4}$$

The conventional IV estimate $\hat{\beta}_{2SLS} = \frac{\xi_1}{\xi_2}$ is the natural sample-analog of (4). As is well-known, however, this estimator has no integer moments.[3] This lack of unbiasedness reflects the fact that the expectation of the ratio of two random variables is not in general equal to the ratio of their expectations.

The form of (4) nonetheless suggests an approach to deriving an unbiased estimator. Suppose we can construct an estimator $\hat{\tau}$ which (a) is unbiased for $1/\pi$ and (b) depends on the data only through ξ_2. If we then define

$$\hat{\delta}(\xi, \Sigma) = \left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right), \tag{5}$$

we have that $E\left[\hat{\delta}\right] = \pi\beta - \frac{\sigma_{12}}{\sigma_2^2}\pi$, and $\hat{\delta}$ is independent of $\hat{\tau}$. Thus, $E\left[\hat{\tau}\hat{\delta}\right] = E[\hat{\tau}]\,E\left[\hat{\delta}\right] = \beta - \frac{\sigma_{12}}{\sigma_2^2}$, and $\hat{\tau}\hat{\delta} + \frac{\sigma_{12}}{\sigma_2^2}$ will be an unbiased estimator of β. Thus, the problem of unbiased estimation of β reduces to that of unbiased estimation of the inverse of a normal mean.

2.1 Unbiased Estimation of the Inverse of a Normal Mean

A result from Voinov & Nikulin (1993) shows that unbiased estimation of $1/\pi$ is possible if we assume its sign is known. Let Φ and ϕ denote the standard normal cdf and pdf respectively.

Lemma 1. *Define*

$$\hat{\tau}\left(\xi_2, \sigma_2^2\right) = \frac{1}{\sigma_2}\frac{1 - \Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)}.$$

For all $\pi > 0$, $E_\pi\left[\hat{\tau}\left(\xi_2, \sigma_2^2\right)\right] = \frac{1}{\pi}$.

The derivation of $\hat{\tau}\left(\xi_2, \sigma_2^2\right)$ in Voinov & Nikulin (1993) relies on the theory of bilateral Laplace transforms, and offers little by way of intuition. Verifying unbiasedness is

[3] If one instead considers median bias, $\hat{\beta}_{2SLS}$ may be substantially biased for small values of π, though this median bias vanishes rapidly as π increases. See e.g. Angrist & Pischke (2009)

a straightforward calculus exercise, however: for the interested reader, we work through the necessary derivations in the proof of Lemma 1.

From the formula for $\hat{\tau}$, we can see that this estimator has two properties which are arguably desirable for a restricted estimate of $1/\pi$. First, it is positive by definition, thereby incorporating the restriction that $\pi > 0$. Second, in the case where positivity of π is obvious from the data (ξ_2 is very large relative to its variance), it is close to the natural plug-in estimator $1/\xi_2$. The second property is an immediate consequence of a well known approximation to the tail of the normal cdf, which is used extensively in the literature on extreme value limit theorems for normal sequences and processes (see Equation 1.5.4 in Leadbetter et al. 1983, and the remainder of that book for applications). We discuss this further in Section 2.4.

Interestingly, $\hat{\tau}\left(\xi_2, \sigma_2^2\right)$ is equal to Mill's ratio for a $N\left(0, \sigma_2^2\right)$ random variable evaluated at ξ_2. Specifically, if we let $\zeta \sim N\left(0, \sigma_2^2\right)$ be independent of ξ_2,

$$E\left[\zeta | \zeta > \xi_2\right] = \sigma_2 \frac{\phi\left(\xi_2/\sigma_2\right)}{1 - \Phi\left(\xi_2/\sigma_2\right)}, \; \hat{\tau}\left(\xi_2, \sigma_2^2\right) = \frac{1}{E\left[\zeta | \zeta > \xi_2\right]}. \tag{6}$$

The estimator $\hat{\tau}$ is therefore related to a number of important formulas in the econometrics of selection models. For instance, the inverse Mills ratio $\hat{\tau}\left(\xi_2, \sigma_2^2\right)^{-1} = E\left[\zeta | \zeta > \xi_2\right]$ appears in the classic Heckman (1979) selection model.

2.2 Unbiased Estimation of β

Given an unbiased estimator of $1/\pi$ which depends only on ξ_2, we can construct an unbiased estimator of β as suggested above. Moreover, this estimator is unique.

Theorem 1. *Define*

$$\begin{aligned} \hat{\beta}_U\left(\xi, \Sigma\right) &= \hat{\tau}\left(\xi_2, \sigma_2^2\right) \hat{\delta}\left(\xi, \Sigma\right) + \tfrac{\sigma_{12}}{\sigma_2^2} \\ &= \tfrac{1}{\sigma_2} \tfrac{1 - \Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)} \left(\xi_1 - \tfrac{\sigma_{12}}{\sigma_2^2}\xi_2\right) + \tfrac{\sigma_{12}}{\sigma_2^2}. \end{aligned}$$

The estimator $\hat{\beta}_U\left(\xi, \Sigma\right)$ is unbiased for β provided $\pi > 0$.

Moreover, if the parameter space (3) contains an open set then $\hat{\beta}_U\left(\xi, \Sigma\right)$ is the unique non-randomized unbiased estimator for β, in the sense that any other estimator $\hat{\beta}\left(\xi, \Sigma\right)$

satisfying

$$E_{\pi,\beta}\left[\hat{\beta}\left(\xi,\Sigma\right)\right]=\beta\ \forall\beta\in B,\pi\in\Pi$$

also satisfies

$$\hat{\beta}\left(\xi,\Sigma\right)=\hat{\beta}_U\left(\xi,\Sigma\right)\ a.s.\,\forall\beta\in B,\pi\in\Pi.$$

Note that the conventional IV estimator can be written as

$$\hat{\beta}_{2SLS}=\frac{\xi_1}{\xi_2}=\frac{1}{\xi_2}\left(\xi_1-\frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)+\frac{\sigma_{12}}{\sigma_2^2}.$$

Thus, $\hat{\beta}_U$ differs from the conventional IV estimator only in that it replaces the plug-in estimate $1/\xi_2$ for $1/\pi$ by the unbiased estimate $\hat{\tau}$.

2.3 The Role of the Sign Restriction

In the introduction we argued that it is frequently reasonable to assume that the sign of the first-stage relationship is known, and Theorem 1 shows that this restriction suffices to allow mean-unbiased estimation of β in the just-identified model. In fact, a restriction on the parameter space is necessary for an unbiased estimator to exist.

In the just-identified linear IV model with parameter space $\left\{(\beta,\pi)\in\mathbb{R}^2\right\}$, Theorem 2.5 of Hirano & Porter (2015) implies that no mean, median, or quantile unbiased estimator can exist. Given this negative result, the positive conclusion of Theorem 1 may seem surprising. The key point is that by restricting the sign of π to be strictly positive, the parameter space Θ as defined in (3) violates Assumption 2.4 of Hirano & Porter (2015), and so renders their negative result inapplicable. Intuitively, assuming the sign of π is known provides just enough information to allow mean-unbiased estimation of β. For further discussion of this point we refer the interested reader to Appendix B.

2.4 Behavior of $\hat{\beta}_U$ When π is Large

While the finite-sample unbiasedness of $\hat{\beta}_U$ is appealing, it is also natural to consider its performance when the instruments are highly informative. This situation, which we will model by taking π to be large, corresponds to the conventional strong-instrument asymptotics where one fixes the data generating process and takes the sample size to infinity.[4]

As we discussed above, the unbiased and conventional IV estimators differ only in that the former substitutes $\hat{\tau}\left(\xi_2, \sigma_2^2\right)$ for $1/\xi_2$. These two estimators for $1/\pi$ coincide to a high order of approximation for large values of ξ_2. Specifically, as noted in Small (2010) (page 40), for $\xi_2 > 0$ we have

$$\sigma_2 \left| \hat{\tau}\left(\xi_2, \sigma_2^2\right) - \frac{1}{\xi_2} \right| \leq \left| \frac{\sigma_2^3}{\xi_2^3} \right|.$$

Thus, since $\xi_2 \xrightarrow{p} \infty$ as $\pi \to \infty$, the difference between $\hat{\tau}\left(\xi_2, \sigma_2^2\right)$ and $1/\xi_2$ converges rapidly to zero (in probability) as π grows. Consequently, the unbiased estimator $\hat{\beta}_U$ (appropriately normalized) has the same limiting distribution as the conventional IV estimator $\hat{\beta}_{2SLS}$ as we take $\pi \to \infty$.

Theorem 2. *As $\pi \to \infty$, holding β and Σ fixed,*

$$\pi \left(\hat{\beta}_U - \hat{\beta}_{2SLS} \right) \xrightarrow{p} 0.$$

Consequently, $\hat{\beta}_U \xrightarrow{p} \beta$ and

$$\pi \left(\hat{\beta}_U - \beta \right) \xrightarrow{d} N\left(0, \sigma_1^2 - 2\beta\sigma_{12} + \beta^2\sigma_2^2\right).$$

Thus, the unbiased estimator $\hat{\beta}_U$ behaves as the standard IV estimator for large values of π. Consequently, one can show that using this estimator along with conventional standard errors will yield asymptotically valid inference under strong-instrument asymptotics. The details of this analysis are standard and so are omitted.

[4]Formally, in the finite-sample normal IV model (1), strong-instrument asymptotics will correspond to fixing π and taking $T \to \infty$, which under mild conditions on Z and $Var\left((U', V')'\right)$ will result in $\Sigma \to 0$ in (2). However, it is straightforward to show that the behavior of $\hat{\beta}_U$, $\hat{\beta}_{2SLS}$, and many other estimators in this case will be the same as the behavior obtained by holding Σ fixed and taking π to infinity. We focus on the latter case to simplify the exposition.

3 Unbiased Estimation with Multiple Instruments

We now consider the case with multiple instruments, where the model is given by (1) and (2) with k (the dimension of Z_t, π, ξ_1 and ξ_2) greater than 1. As discussed in Section 1.2, we assume that the sign of each element π_i of the first stage vector is known, and we normalize this sign to be positive, giving the parameter space (3).

Using the results in Section 2 one can construct an unbiased estimator for β in many different ways. For any index $i \in \{1, \dots, k\}$, the unbiased estimator based on $(\xi_{1,i}, \xi_{2,i})$ will, of course, still be unbiased for β when $k > 1$. One can also take non-random weighted averages of the unbiased estimators based on different instruments. Using the unbiased estimator based on a fixed linear combination of instruments is another possibility, so long as the linear combination preserves the sign restriction. However, such approaches will not adapt to information from the data about the relative strength of instruments and so will typically be inefficient when the instruments are strong.

By contrast, the usual 2SLS estimator achieves asymptotic efficiency in the strongly identified case (modeled here as taking $\|\pi\| \to \infty$) when errors are homoskedastic. In fact, in this case 2SLS is asymptotically equivalent to an infeasible estimator that uses knowledge of π to choose the optimal combination of instruments. Thus, a reasonable goal is to construct an estimator that (1) is unbiased for fixed π and (2) is asymptotically efficient as $\|\pi\| \to \infty$.[5] In the remainder of this section we first introduce a class of unbiased estimators and then show that a (feasible) estimator in this class attains the desired strong IV efficiency property.

[5]In the heteroskedastic case, the 2SLS estimator will no longer be asymptotically efficient, and a two-step GMM estimator can be used to achieve the efficiency bound. Because it leads to simpler exposition, and because the 2SLS estimator is common in practice, we consider asymptotic equivalence with 2SLS, rather than asymptotic efficiency in the heteroskedastic case, as our goal. As discussed in Section 3.3 below, however, our approach generalizes directly to efficient estimators in non-homoskedastic settings.

3.1 A General Class of Unbiased Estimators

Let

$$\xi(i) = \begin{pmatrix} \xi_{1,i} \\ \xi_{2,i} \end{pmatrix} \text{ and } \Sigma(i) = \begin{pmatrix} \Sigma_{11,ii} & \Sigma_{12,ii} \\ \Sigma_{21,ii} & \Sigma_{12,ii} \end{pmatrix}$$

be the reduced form and first stage estimators based on the ith instrument and their variance matrix, respectively, so that $\hat{\beta}_U(\xi(i), \Sigma(i))$ is the unbiased estimator based on the ith instrument. Given a weight vector $w \in \mathbb{R}^k$ with $\sum_{i=1}^k w_i = 1$, let

$$\hat{\beta}_w(\xi, \Sigma; w) = \sum_{i=1}^{k} w_i \hat{\beta}_U(\xi(i), \Sigma(i)).$$

Clearly, $\hat{\beta}_w$ is unbiased so long as w is nonrandom. Allowing w to depend on the data ξ, however, may introduce bias through the correlation between the weights and the estimators $\hat{\beta}_U(\xi(i), \Sigma(i))$.

To avoid this bias we first consider a randomized unbiased estimator and then take its conditional expectation given the sufficient statistic ξ to eliminate the randomization. Let $\zeta \sim N(0, \Sigma)$ be independent of ξ, and let $\xi^{(a)} = \xi + \zeta$ and $\xi^{(b)} = \xi - \zeta$. Then $\xi^{(a)}$ and $\xi^{(b)}$ are (unconditionally) independent draws with the same marginal distribution as ξ, save that Σ is replaced by 2Σ. If T is even, $Z'Z$ is the same across the first and second halves of the sample, and the errors are iid, then $\xi^{(a)}$ and $\xi^{(b)}$ have the same joint distribution as the reduced form estimators based on the first and second half of the sample. Thus, we can think of these as split-sample reduced-form estimates.

Let $\hat{w} = \hat{w}(\xi^{(b)})$ be a vector of data dependent weights with $\sum_{i=1}^k \hat{w}_i = 1$. By the independence of $\xi^{(a)}$ and $\xi^{(b)}$,

$$E\left[\hat{\beta}_w(\xi^{(a)}, 2\Sigma; \hat{w}(\xi^{(b)}))\right] = \sum_{i=1}^{k} E\left[\hat{w}_i(\xi^{(b)})\right] \cdot E\left[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))\right] = \beta. \qquad (7)$$

To eliminate the noise introduced by generating $\xi^{(a)}$ and $\xi^{(a)}$, define the "Rao-Blackwellized" estimator

$$\hat{\beta}_{RB} = \hat{\beta}_{RB}(\xi, \Sigma; \hat{w}) = E\left[\hat{\beta}_w(\xi^{(a)}, 2\Sigma; \hat{w}(\xi^{(b)})) \middle| \xi\right].$$

Unbiasedness of $\hat{\beta}_{RB}$ follows immediately from (7) and the law of iterated expectations. While $\hat{\beta}_{RB}$ does not, to our knowledge, have a simple closed form, it can be computed by integrating over the distribution of ζ. This can easily be done by simulation, taking the sample average of $\hat{\beta}_w$ over simulated draws of $\xi^{(a)}$ and $\xi^{(b)}$ while holding ξ at its observed value.

3.2 Equivalence with 2SLS under Strong IV Asymptotics

We now propose a set of weights $\hat{w}$ which yield an unbiased estimator asymptotically equivalent to 2SLS. To motivate these weights, note that for $W = Z'Z$ and e_i the ith standard basis vector, the 2SLS estimator can be written as

$$\hat{\beta}_{2SLS} = \frac{\xi_2' W \xi_1}{\xi_2' W \xi_2} = \sum_{i=1}^{k} \frac{\xi_2' W e_i e_i' \xi_2}{\xi_2' W \xi_2} \frac{\xi_{1,i}}{\xi_{2,i}},$$

which is the GMM estimator with weight matrix $W = Z'Z$. Thus, the 2SLS estimator is a weighted average of the 2SLS estimates based on single instruments, where the weight for estimate $\xi_{1,i}/\xi_{2,i}$ based on instrument i is equal to $\frac{\xi_2' W e_i e_i' \xi_2}{\xi_2' W \xi_2}$. This suggests the unbiased Rao-Blackwellized estimator with weights $\hat{w}_i^*(\xi^{(b)}) = \frac{\xi_2^{(b)\prime} W e_i e_i' \xi_2^{(b)}}{\xi_2^{(b)\prime} W \xi_2^{(b)}}$:

$$\hat{\beta}_{RB}^* = \hat{\beta}_{RB}(\xi, \Sigma; \hat{w}) = E\left[\hat{\beta}_w(\xi^{(a)}, 2\Sigma; \hat{w}^*(\xi^{(b)}))\middle|\xi\right]. \tag{8}$$

The following theorem shows that $\hat{\beta}_{RB}^*$ is asymptotically equivalent to $\hat{\beta}_{2SLS}$ in the strongly identified case, and is therefore asymptotically efficient if the errors are iid.

Theorem 3. *Let $\hat{\pi} \to \infty$ with $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$. Then $\|\pi\|(\hat{\beta}_{RB}^* - \hat{\beta}_{2SLS}) \xrightarrow{p} 0$.*

The condition that $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$ amounts to an assumption that the "strength" of all instruments is approximately the same. As discussed below in Section 3.3, this assumption can be relaxed by redefining the instruments.

To understand why Theorem 3 holds, consider the "oracle" weights $w_i^* = \frac{\pi' W e_i e_i' \pi}{\pi' W \pi}$. It is easy to see that $\hat{w}_i^* - w_i^* \xrightarrow{p} 0$ as $\|\pi\| \to \infty$. Consider the oracle unbiased estimator $\hat{\beta}_{RB}^o = \hat{\beta}_{RB}(\xi, \Sigma; w^*)$, and the oracle combination of individual 2SLS estimators

$\hat{\beta}^{o}_{2SLS} = \sum_{i=1}^{k} w_i^* \frac{\xi_{1,i}}{\xi_{2,i}}$. By arguments similar to those used to show that statistical noise in the first stage estimates does not affect the 2SLS asymptotic distribution under strong instrument asymptotics, it can be seen that $\|\pi\|(\hat{\beta}^{o}_{2SLS} - \hat{\beta}_{2SLS}) \xrightarrow{p} 0$ as $\|\pi\| \to \infty$. Further, one can show that $\hat{\beta}^{o}_{RB} = \hat{\beta}_w(\xi, \Sigma; w^*) = \sum_{i=1}^{k} w_i^* \hat{\beta}_U(\xi(i), \Sigma(i))$. Since this is just $\hat{\beta}^{o}_{2SLS}$ with $\hat{\beta}_U(\xi(i), \Sigma(i))$ replacing $\xi_{i,1}/\xi_{i,2}$, it follows by Theorem 2 that $\|\pi\|(\hat{\beta}^{o}_{RB} - \hat{\beta}^{o}_{2SLS}) \xrightarrow{p} 0$. Theorem 3 then follows by showing that $\|\pi\|(\hat{\beta}_{RB} - \hat{\beta}^{o}_{RB}) \xrightarrow{p} 0$, which follows for essentially the same reasons that first stage noise does not affect the asymptotic distribution of the 2SLS estimator but requires some additional argument. We refer the reader to the proof of Theorem 3 in Appendix A for details.

3.3 Extensions

The estimator $\hat{\beta}^*_{RB}$ proposed above may be viewed as deficient because (1) it is asymptotically efficient only under homoskedastic errors and (2) the condition $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$ rules out cases where some instruments are strong while others are weak or "semi-strong." We now discuss extensions of the estimator that address these issues.

First, consider asymptotic efficiency in the heteroskedastic case. In this case the two step GMM estimator given by

$$\hat{\beta}_{GMM,\hat{W}} = \frac{\xi_2' \hat{W} \xi_1}{\xi_2' \hat{W} \xi_2} \text{ where } \hat{W} = \left(\Sigma_{11} - \hat{\beta}_{2SLS}(\Sigma_{12} + \Sigma_{21}) + \hat{\beta}^2_{2SLS}\Sigma_{22}\right)^{-1}, \tag{9}$$

is asymptotically efficient under strong instruments. Here, $\hat{W}$ is an estimate of the inverse of the variance matrix of the moments $\xi_1 - \beta\xi_2$, which the GMM estimator sets close to zero. Let

$$\hat{w}^*_{GMM,i}(\xi^{(b)}) = \frac{\xi_2^{(b)\prime} \hat{W}(\xi^{(b)}) e_i e_i' \xi_2^{(b)}}{\xi_2^{(b)\prime} \hat{W}(\xi^{(b)}) \xi_2^{(b)}} \text{ where } \hat{W}(\xi^{(b)}) = \left(\Sigma_{11} - \hat{\beta}(\xi^{(b)})(\Sigma_{12} + \Sigma_{21}) + \hat{\beta}(\xi^{(b)})^2\Sigma_{22}\right)^{-1} \tag{10}$$

for a preliminary estimator $\hat{\beta}(\xi^{(b)})$ of β based on $\xi^{(b)}$. The Rao-Blackwellized estimator formed by replacing $\hat{w}^*$ with $\hat{w}^*_{GMM}$ in the definition of $\hat{\beta}^*_{RB}$ gives an unbiased estimator that is asymptotically efficient under strong instrument asymptotics with heteroskedastic errors. We refer the reader to Appendix A for details.

Now let us consider the case where, while $\|\pi\| \to \infty$, the elements π_i may increase at different rates. Let M be a $k \times k$ invertible matrix such that all elements are strictly positive, and let

$$\tilde{\xi} = (I_2 \otimes M)\xi, \qquad \tilde{\Sigma} = (I_2 \otimes M)\Sigma(I_2 \otimes M)', \qquad \tilde{W} = M^{-1\prime} W M^{-1}.$$

The GMM estimator based on $\tilde{\xi}$ and $\tilde{W}$ is numerically equivalent to the GMM estimator based on ξ and W (which, for $W = Z'Z$, is the 2SLS estimator). Thus, if we construct the estimator $\hat{\beta}^*_{RB}$ from $\tilde{\xi}$ and $\tilde{W}$ instead of $\tilde{\xi}$ and $\tilde{W}$, we obtain the desired asymptotic equivalence result so long as $\tilde{\pi} = M\pi$ is nonnegative and satisfies $\|\tilde{\pi}\| \to \infty$ and $\|\tilde{\pi}\| / \min_i \tilde{\pi}_i = \mathcal{O}(1)$. Since M contains only positive elements, $\tilde{\pi}$ will be in the positive orthant so long as π is in the positive orthant. Moreover, note that $\min_i \tilde{\pi}_i \geq (\min_{i,j} M_{ij})\|\pi\| = (\min_{i,j} M_{ij}) \frac{\|\pi\|}{\|M\pi\|} \|\tilde{\pi}\| \geq (\min_{i,j} M_{ij}) \left(\inf_{\|u\|=1} \frac{\|u\|}{\|Mu\|} \right) \|\tilde{\pi}\|$, so that the requirement $\|\tilde{\pi}\| / \min_i \tilde{\pi}_i = \mathcal{O}(1)$ now holds automatically.

4 Performance of Single-Instrument Estimators

The estimator $\hat{\beta}_U$ based on a single instrument plays a central role in all of our results, so in this section we examine the performance of this estimator in simulation. For comparison we also discuss results for the two-stage least squares estimator $\hat{\beta}_{2SLS}$. The lack of moments for $\hat{\beta}_{2SLS}$ in the just-identified context renders some comparisons with $\hat{\beta}_U$ infeasible, however, so we also consider the performance of the Fuller (1977) estimator with $c = 1$,

$$\hat{\beta}_{FULL} = \frac{\xi_2 \xi_1 + \sigma_{12}}{\xi_2^2 + \sigma_2^2}$$

which we define as in Mills et al. (2014).[6] Note that in the just-identified case considered here $\hat{\beta}_{FULL}$ also coincides with the bias-corrected 2SLS estimator (again, see Mills et al.).

While the model (2) has five parameters in the single-instrument case, $(\beta, \pi, \sigma_{11}^2, \sigma_{12}, \sigma_{22}^2)$, an equivariance argument implies that for our purposes it suffices to fix $\beta = 0$, $\sigma_{11} =$

[6] In the case where U_t and V_t are correlated or heteroskedastic across t, the definition of $\hat{\beta}_{FULL}$ above is the natural extension of the definition considered in Mills et al. (2014).

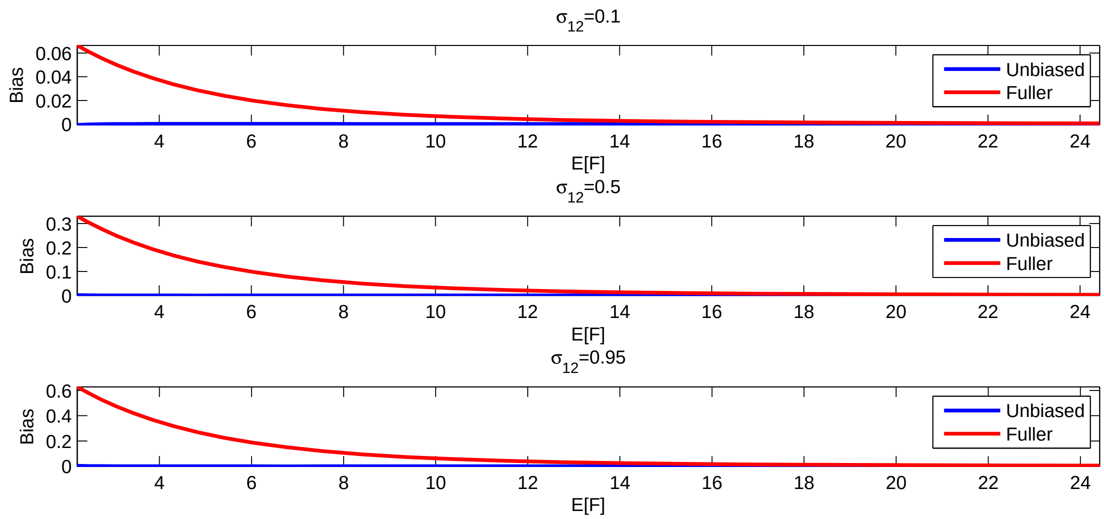

Figure 1: Bias of single-instrument estimators, plotted against mean $E[F]$ of first-stage F-statistic, based on 10 million simulations.

$\sigma_{22} = 1$ and consider the parameter space $(\pi, \sigma_{12}) \in (0, \infty) \times [0, 1)$. See Appendix C for details. Since this parameter space is just two-dimensional, we can fully explore it via simulation.

4.1 Estimator Location

We first compare the bias of $\hat{\beta}_U$ and $\hat{\beta}_{FULL}$ (we omit $\hat{\beta}_{2SLS}$ from this comparison, as it does not have a mean in the just-identified case). We consider $\sigma_{12} \in \{0.1, 0.5, 0.95\}$ and examine a wide range of values for $\pi > 0$.[7]

If, rather than considering mean bias, we instead consider median bias, we find that $\hat{\beta}_U$ and $\hat{\beta}_{2SLS}$ generally exhibit smaller median bias than $\hat{\beta}_{FULL}$. There is no ordering between $\hat{\beta}_U$ and $\hat{\beta}_{2SLS}$ in terms of median bias, however, as the median bias of $\hat{\beta}_U$ is smaller than that of $\hat{\beta}_{2SLS}$ for very small values of π, while the median bias of $\hat{\beta}_{2SLS}$ is

[7]We restrict attention to $\pi > 1$ in these bias plots. Since the first stage F-statistic is $F = \xi_2^2$ in the present context, this corresponds to $E[F] > 2$. The expectation of $\hat{\beta}_U$ ceases to exist at $\pi = 0$, and for π close to zero the heavy tails of $\hat{\beta}_U$ mean that the sample average of $\hat{\beta}_U$ in simulation can differ substantially from zero even with a large number of simulation replications.

smaller for larger values π.

4.2 Estimator Dispersion

The lack of moments for $\hat{\beta}_{2SLS}$ complicates comparisons of dispersion, since we cannot consider mean squared error or mean absolute deviation, and also cannot recenter $\hat{\beta}_{2SLS}$ around its mean. As an alternative, we instead consider the full distribution of the absolute deviation of each estimator from its median. In particular, for the estimators $(\hat{\beta}_U, \hat{\beta}_{2SLS}, \hat{\beta}_{FULL})$ we calculate the zero-median residuals

$$(\varepsilon_U, \varepsilon_{2SLS}, \varepsilon_{FULL}) = \left(\hat{\beta}_U - \text{med}\left(\hat{\beta}_U\right), \hat{\beta}_{2SLS} - \text{med}\left(\hat{\beta}_{2SLS}\right), \hat{\beta}_{FULL} - \text{med}\left(\hat{\beta}_{FULL}\right)\right).$$

Our simulation results suggest a strong stochastic ordering between these residuals (in absolute value). In particular we find that $|\varepsilon_{2SLS}|$ approximately dominates $|\varepsilon_U|$, which in turn approximately dominates $|\varepsilon_{FULL}|$, both in the sense of first order stochastic dominance. In particular, for $\tau \in \{0.001, 0.002, ..., 0.999\}$ the τ-th quantile of $|\varepsilon_{2SLS}|$ in simulation is never more than 10^{-4} smaller than the τ-th quantile of $|\varepsilon_U|$, and the τ-th quantile of $|\varepsilon_U|$ is never more than 10^{-3} smaller than the τ-th quantile of $|\varepsilon_{FULL}|$, both uniformly over τ and (π, σ_{12}).[8] Thus, our simulations demonstrate that $\hat{\beta}_{2SLS}$ is more dispersed around its median than is $\hat{\beta}_U$, which is in turn more dispersed around its median than $\hat{\beta}_{FULL}$. To illustrate this finding, Figure 2 plots the median of $|\varepsilon|$ for the different estimators. While Figure 2 considers only one quantile and three values of σ_{12}, more exhaustive simulation results are discussed in Appendix D.

This numerical result is consistent with analytical results on the tail behavior of the estimators. In particular, $\hat{\beta}_{2SLS}$ has no moments, reflecting thick tails in its sampling distribution, while $\hat{\beta}_{FULL}$ has all moments, reflecting thin tails. As we show in the

[8]By contrast, the τ-th quantile of $|\varepsilon_{2SLS}|$ may exceed corresponding quantile of $|\varepsilon_U|$ by as much as 483, or (in proportional terms) by as much as a factor of 32, while the τ-th quantile of $|\varepsilon_U|$ may exceed the corresponding quantile of $|\varepsilon_{FULL}|$ by as much as 37, or (in proportional terms) by as much as a factor of 170.

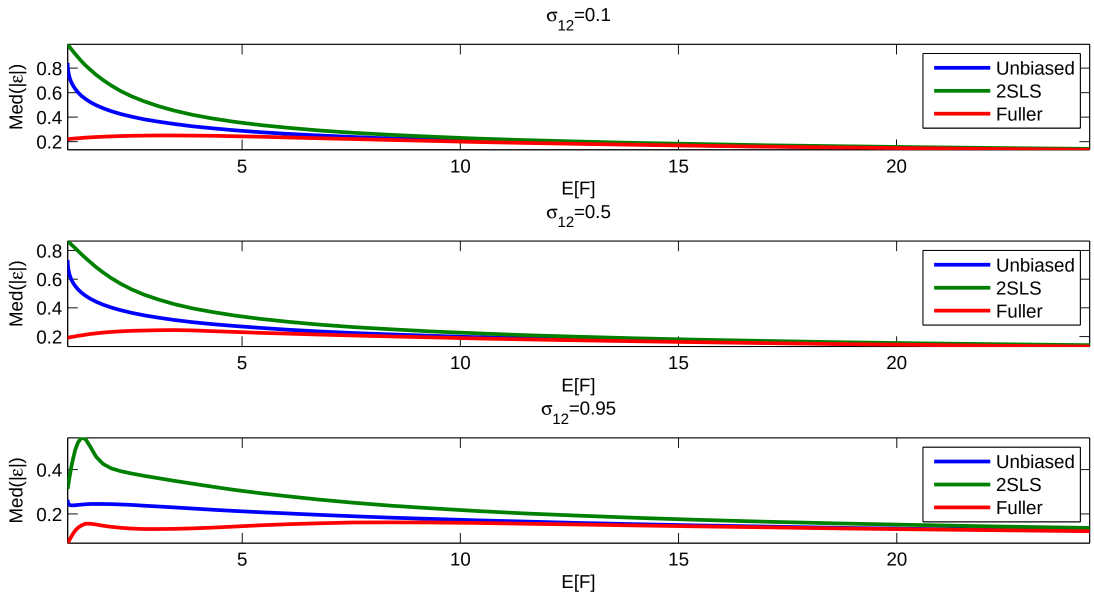

Figure 2: Median of $|\varepsilon| = |\hat{\beta} - \text{med}\left(\hat{\beta}\right)$ for single-instrument IV estimators, plotted against mean $E_\pi [F]$ of first-stage F-statistic, based on 10 million simulations.

appendix, for $\pi > 0$ the unbiased estimator $\hat{\beta}_U$ has a first but not a second moment and so falls between these two extremes.

References

Aizer, A. & Doyle Jr., J. J. (2013), Juvenile incarceration, human capital and future crime: Evidence from randomly-assigned judges, Working Paper 19102, National Bureau of Economic Research.

Andrews, D., Moreira, M. & Stock, J. (2006), 'Optimal two-sided invariant similar tests of instrumental variables regression', *Econometrica* **74**, 715–752.

Andrews, D. W. K. (2001), 'Testing when a parameter is on the boundary of the maintained hypothesis', *Econometrica* **69**(3), 683–734.

Andrews, I. (2014), Conditional linear combination tests for weakly identified models. Unpublished Manuscript.

Angrist, J. D. & Krueger, A. B. (1991), 'Does compulsory school attendance affect schooling and earnings?', *The Quarterly Journal of Economics* **106**(4), 979–1014.

Angrist, J. & Pischke, S. (2009), A note on bias in just identified iv with weak instruments. At http://www.mostlyharmlesseconometrics.com/2009/05/just-identified-iv/.

Baricz, Á. (2008), 'Mills' ratio: Monotonicity patterns and functional inequalities', *Journal of Mathematical Analysis and Applications* **340**(2), 1362–1370.

Fuller, W. (1977), 'Some properties of a modification of the limited information maximum likelihood estimator', *Econometrica* **45**, 939–953.

Gouriéroux, C., Holly, A. & Monfort, A. (1982), 'Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters', *Econometrica* **50**(1), 63–80.

Hahn, J., Hausman, J. & Kuersteiner, G. (2004), 'Estimation with weak instruments: Accuracy of higher order bias and mse approximations', *Econometrics Journal* **7**, 272–306.

Heckman, J. J. (1979), 'Sample selection bias as a specification error', *Econometrica* **47**, 153–161.

Heckman, J. J., Urzua, S. & Vytlacil, E. (2006), 'Understanding instrumental variables in models with essential heterogeneity', *Review of Economics and Statistics* **88**(3), 389–432.

Hillier, G. (2006), 'Yet more on the exact properties of iv estimators', *Econometric Theory* **22**, 913–931.

Hirano, K. & Porter, J. (2015), 'Location properties of point estimators in linear instrumental variables and related models', *Econometric Reviews* . Forthcoming.

Hirano, K. & Porter, J. R. (2012), 'Impossibility results for nondifferentiable functionals', *Econometrica* **80**(4), 1769–1790.

Imbens, G. W. & Angrist, J. D. (1994), 'Identification and estimation of local average treatment effects', *Econometrica* **62**(2), 467–475.

Kleibergen, F. (2007), 'Generalizing weak iv robust statistics towards multiple parameters, unrestricted covariance matrices and identification statistics', *Journal of Econometrics* **133**, 97–126.

Leadbetter, M. R., Lindgren, G. & Rootzen, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, 1 edition edn, Springer, New York.

Lehmann, E. L. & Romano, J. P. (2005), *Testing statistical hypotheses*, Springer.

Mills, B., Moreira, M. & Vilela, L. (2014), 'Estimator tests based on t-statistics for iv regression with weak instruments', *Journal of Econometrics* **182**, 351–363.

Moon, H. R., Schorfheide, F. & Granziera, E. (2013), 'Inference for VARs identified with sign restrictions', *Unpublished Manuscript* .

Phillips, P. (1983), *Handbook of Econometrics*, Amsterdam: North Holland, chapter Exact Small Sample Theory in the Simultaneous Equations Model, pp. 449–516.

Small, C. (2010), *Expansions and Asymptotics for Statistics*, Chapman and Hall.

Staiger, D. & Stock, J. H. (1997), 'Instrumental variables regression with weak instruments', *Econometrica* **65**(3), 557–586.

Voinov, V. & Nikulin, M. (1993), *Unbiased Estimators and Their Applications*, Vol. 1, Springer Science+Business Media.

Appendix

A Proofs

This appendix contains proofs of the results in the main text. The notation is the same as in the main text.

A.1 Single Instrument Case

This section proves the results from Section 2, which treats the single instrument case ($k = 1$). We prove Lemma 1 and Theorems 1 and 2, along with results regarding the lack of second moments of the estimators in these theorems.

We first prove Lemma 1, which shows unbiasedness of $\hat{\tau}$ for $1/\pi$. As discussed in the main text, this result is known in the literature (see, e.g., pp. 181-182 of Voinov & Nikulin 1993). We give a constructive proof based on elementary calculus (Voinov & Nikulin provide a derivation based on the bilateral Laplace transform).

Proof of Lemma 1. Since $\xi_2/\sigma_2 \sim N(\pi/\sigma_2, 1)$, we have

$$E_{\pi,\beta}\hat{\tau}(\xi_2, \sigma_2^2) = \frac{1}{\sigma_2}\int \frac{1-\Phi(x)}{\phi(x)}\phi(x-\pi/\sigma_2)\,dx = \frac{1}{\sigma_2}\int (1-\Phi(x))\exp\left((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2\right)\,dx$$
$$= \frac{1}{\sigma_2}\exp(-(\pi/\sigma_2)^2/2)\left\{[(1-\Phi(x))(\sigma_2/\pi)\exp((\pi/\sigma_2)x)]_{x=-\infty}^{\infty} + \int(\sigma_2/\pi)\exp((\pi/\sigma_2)x)\phi(x)\,dx\right\},$$

using integration by parts to obtain the last equality. Since the first term in brackets in the last line is zero, this is equal to

$$\frac{1}{\sigma_2}\int(\sigma_2/\pi)\exp((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2)\phi(x)\,dx = \frac{1}{\pi}\int\phi(x-\pi/\sigma_2)\,dx = \frac{1}{\pi}.$$

□

We note that $\hat{\tau}$ does not have a second moment.

Lemma 2. *The expectation of $\hat{\tau}(\xi_2, \sigma_2^2)^2$ is undefined for all π.*

Proof. By similar calculations to those in the proof of Lemma 1,

$$E_{\pi,\beta}\hat{\tau}(\xi_2,\sigma_2^2)^2 = \frac{1}{\sigma_2^2}\int \frac{(1-\Phi(x))^2}{\phi(x)} \exp\left((\pi/\sigma_2)x - (\pi/\sigma_2)^2/2\right)\, dx.$$

For $x < 0$, $1-\Phi(x) \geq 1/2$, so the integrand is bounded from below by a constant times $\exp(x^2/2 + (\pi/\sigma_2)x)$, which is bounded away from zero as $x \to -\infty$.

□

Proof of Theorem 1. To establish unbiasedness, note that since ξ_2 and $\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2$ are jointly normal with zero covariance, they are independent. Thus,

$$E_{\pi,\beta}\hat{\beta}_U(\xi,\Sigma) = (E_{\pi,\beta}\hat{\tau})\left[E_{\pi,\beta}\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)\right] + \frac{\sigma_{12}}{\sigma_2^2} = \frac{1}{\pi}\left(\pi\beta - \frac{\sigma_{12}}{\sigma_{22}}\pi\right) + \frac{\sigma_{12}}{\sigma_{22}} = \beta$$

since $E_{\pi,\beta}\hat{\tau} = 1/\pi$ by Lemma 1.

To establish uniqueness, consider any unbiased estimator $\hat{\beta}\,(\xi,\Sigma)$. By unbiasedness

$$E_{\pi,\beta}\left[\hat{\beta}\,(\xi,\Sigma) - \hat{\beta}_U(\xi,\Sigma)\right] = 0 \;\forall \beta \in B, \pi \in \Pi.$$

The parameter space contains an open set by assumption, so by Theorem 4.3.1 of Lehmann & Romano (2005) the family of distributions of ξ under $(\beta,\pi)\in\Theta$ is complete. Thus $\hat{\beta}\,(\xi,\Sigma) - \hat{\beta}_U(\xi,\Sigma) = 0$ almost surely for all $(\beta,\pi)\in\Theta$ by the definition of completeness.

□

We also note that $\hat{\beta}_U$ does not have a second moment.

Lemma 3. *The expectation of $\hat{\beta}_U(\xi,\Sigma)^2$ is undefined for all π,β.*

Proof. If $E_{\pi,\beta}\hat{\beta}_U(\xi,\Sigma)^2$ existed, we would have, by independence of ξ_2 and $\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2$,

$$E_{\pi,\beta}\hat{\beta}_U(\xi,\Sigma)^2 = \left(E_{\pi,\beta}\hat{\tau}^2\right)\left[E_{\pi,\beta}\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)^2\right] + \frac{\sigma_{12}^2}{\sigma_2^4} + 2\frac{\sigma_{12}}{\sigma_2^2}\left(E_{\pi,\beta}\hat{\tau}\right)\left[E_{\pi,\beta}\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right)\right]$$

which must be infinite since all of the terms in the expression are finite except for $E_{\pi,\beta}\hat{\tau}^2$, and the term multiplying $E_{\pi,\beta}\hat{\tau}^2$ is nonzero. □

We now consider the behavior of $\hat{\beta}_U$ relative to the usual 2SLS estimator (which, in the single instrument case considered here, is given by $\hat{\beta}_{2SLS} = \xi_1/\xi_2$) as $\pi \to \infty$.

Proof of Theorem 2. Note that

$$\hat{\beta}_U - \hat{\beta}_{2SLS} = \left(\hat{\tau}(\xi_2, \sigma_2^2) - \frac{1}{\xi_2}\right)\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right) = \left(\xi_2\hat{\tau}(\xi_2, \sigma_2^2) - 1\right)\left(\frac{\xi_1}{\xi_2} - \frac{\sigma_{12}}{\sigma_2^2}\right).$$

As $\pi \to \infty$, $\xi_1/\xi_2 = \hat{\beta}_{2SLS} = \mathcal{O}_P(1)$, so it suffices to show that $\pi\left(\xi_2\hat{\tau}(\xi_2, \sigma_2^2) - 1\right) = o_P(1)$ as $\pi \to \infty$. Note that, by p. 40 of Small (2010),

$$\pi\left|\xi_2\hat{\tau}(\xi_2, \sigma_2^2) - 1\right| = \pi\left|\frac{\xi_2}{\sigma_2}\frac{1 - \Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)} - 1\right| \leq \pi\frac{\sigma_2^2}{\xi_2^2} = \frac{\pi}{\xi_2}\frac{\sigma_2^2}{\xi_2}.$$

This converges in probability to zero since $\pi/\xi_2 \xrightarrow{p} 1$ and $\frac{\sigma_2^2}{\xi_2} \xrightarrow{p} 0$ as $\pi \to \infty$. □

The following lemma regarding the mean absolute deviation of $\hat{\beta}_U$ will be useful in the next section treating the case with multiple instruments.

Lemma 4. *For a constant $K(\beta, \Sigma)$ depending only on Σ and β (but not on π),*

$$\pi E_{\pi,\beta}\left|\hat{\beta}_U(\xi, \Sigma) - \beta\right| \leq K(\beta, \Sigma).$$

Proof. We have

$$\begin{aligned}
\pi\left(\hat{\beta}_U - \beta\right) &= \pi\left[\hat{\tau}\cdot\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right) + \frac{\sigma_{12}}{\sigma_2^2} - \beta\right] = \pi\hat{\tau}\cdot\left(\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2\right) + \pi\frac{\sigma_{12}}{\sigma_2^2} - \pi\beta \\
&= \pi\hat{\tau}\cdot\left(\xi_1 - \beta\pi - \frac{\sigma_{12}}{\sigma_2^2}(\xi_2 - \pi)\right) + \pi\hat{\tau}\beta\pi - \pi\hat{\tau}\frac{\sigma_{12}}{\sigma_2^2}\pi + \pi\frac{\sigma_{12}}{\sigma_2^2} - \pi\beta \\
&= \pi\hat{\tau}\cdot\left(\xi_1 - \beta\pi - \frac{\sigma_{12}}{\sigma_2^2}(\xi_2 - \pi)\right) + \pi(\pi\hat{\tau} - 1)\left(\beta - \frac{\sigma_{12}}{\sigma_2^2}\right).
\end{aligned}$$

Using this and the fact that ξ_2 and $\xi_1 - \frac{\sigma_{12}}{\sigma_2^2}\xi_2$ are independent, it follows that

$$\pi E_{\pi,\beta}\left|\hat{\beta}_U - \beta\right| \leq E_{\pi,\beta}\left|\xi_1 - \beta\pi - \frac{\sigma_{12}}{\sigma_2^2}(\xi_2 - \pi)\right| + \pi E_{\pi,\beta}|\pi\hat{\tau} - 1|\left(\beta - \frac{\sigma_{12}}{\sigma_2^2}\right).$$

where we have used the fact that $E_{\pi,\beta}\pi\hat{\tau} = 1$. The only term in the above expression that depends on π is $\pi E_{\pi,\beta}|\pi\hat{\tau} - 1|$. Note that this is bounded by $\pi E_{\pi,\beta}\pi\hat{\tau} + \pi = 2\pi$ (using unbiasedness and positivity of $\hat{\tau}$), so we can assume an arbitrary lower bound on π when bounding this term.

Letting $\tilde{\pi} = \pi/\sigma_2$, we have $\xi_2/\sigma_2 \sim N(\tilde{\pi}, 1)$, so that

$$\frac{\pi}{\sigma_2} E_{\pi,\beta}|\pi\hat{\tau} - 1| = \frac{\pi}{\sigma_2} E_{\pi,\beta}\left|\frac{\pi}{\sigma_2}\frac{1-\Phi(\xi_2/\sigma_2)}{\phi(\xi_2/\sigma_2)} - 1\right| = \tilde{\pi}\int\left|\tilde{\pi}\frac{1-\Phi(z)}{\phi(z)} - 1\right|\phi(z-\tilde{\pi})\,dz.$$

Let $\varepsilon > 0$ be a constant to be determined later in the proof. By (1.1) in Baricz (2008)

$$\begin{aligned}
&\tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{1-\Phi(z)}{\phi(z)} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz \\
&\leq \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{1}{z} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz + \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{z}{z^2+1} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz.
\end{aligned}$$

The first term is

$$\tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{\tilde{\pi}-z}{\tilde{\pi}z}\right|\phi(z-\tilde{\pi})\,dz \leq \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{\tilde{\pi}-z}{\tilde{\pi}^2\varepsilon}\right|\phi(z-\tilde{\pi})\,dz \leq \frac{1}{\varepsilon}\int|u|\phi(u)\,du.$$

The second term is

$$\begin{aligned}
&\tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{1}{z+1/z} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz = \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\left|\frac{\tilde{\pi}-(z+1/z)}{\tilde{\pi}(z+1/z)}\right|\phi(z-\tilde{\pi})\,dz \\
&\leq \tilde{\pi}^2\int_{z\geq\tilde{\pi}\varepsilon}\frac{|\tilde{\pi}-z| + \frac{1}{\varepsilon\tilde{\pi}}}{\tilde{\pi}^2\varepsilon}\phi(z-\tilde{\pi})\,dz \leq \frac{1}{\varepsilon}\int\left(|u| + \frac{1}{\varepsilon\tilde{\pi}}\right)\phi(u)\,dz
\end{aligned}$$

We also have,

$$\tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}\left|\frac{1-\Phi(z)}{\phi(z)} - \frac{1}{\tilde{\pi}}\right|\phi(z-\tilde{\pi})\,dz \leq \tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}\frac{1-\Phi(z)}{\phi(z)}\phi(z-\tilde{\pi})\,dz + \tilde{\pi}\int_{z<\tilde{\pi}\varepsilon}\phi(z-\tilde{\pi})\,dz.$$

The second term is equal to $\tilde{\pi}\Phi(\tilde{\pi}\varepsilon - \tilde{\pi})$, which is bounded uniformly over $\tilde{\pi}$ for $\varepsilon < 1$.

The first term is

$$\begin{aligned}
&\tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}(1-\Phi(z))\exp\left(\tilde{\pi}z - \frac{1}{2}\tilde{\pi}^2\right)\,dz \\
&= \tilde{\pi}^2\int_{z<\tilde{\pi}\varepsilon}\int_{t\geq z}\phi(t)\exp\left(\tilde{\pi}z - \frac{1}{2}\tilde{\pi}^2\right)\,dtdz \\
&= \tilde{\pi}^2\int_{t\in\mathbb{R}}\int_{z\leq\min\{t,\tilde{\pi}\varepsilon\}}\phi(t)\exp\left(\tilde{\pi}z - \frac{1}{2}\tilde{\pi}^2\right)\,dzdt \\
&= \tilde{\pi}^2\exp\left(-\frac{1}{2}\tilde{\pi}^2\right)\int_{t\in\mathbb{R}}\phi(t)\left[\frac{1}{\tilde{\pi}}\exp\left(\tilde{\pi}z\right)\right]_{z=-\infty}^{\min\{t,\tilde{\pi}\varepsilon\}}\,dt \\
&= \tilde{\pi}\exp\left(-\frac{1}{2}\tilde{\pi}^2\right)\int_{t\in\mathbb{R}}\phi(t)\exp\left(\tilde{\pi}\min\{t,\tilde{\pi}\varepsilon\}\right)\,dt \\
&\leq \tilde{\pi}\exp\left(-\frac{1}{2}\tilde{\pi}^2 + \varepsilon\tilde{\pi}^2\right).
\end{aligned}$$

For $\varepsilon < 1/2$, this is uniformly bounded over all $\tilde{\pi} > 0$.

$\square$

A.2 Multiple Instrument Case

This section proves Theorem 3, and the extension of this theorem discussed in Section 3.3. The result follows from a series of lemmas given below. To accomodate the extension discussed in Section 3.3, we consider a more general setup.

Consider the GMM estimator $\hat{\beta}_{GMM,W} = \frac{\xi_2' \hat{W} \xi_1}{\xi_2' \hat{W} \xi_2}$, where $\hat{W} = \hat{W}(\xi)$ is a data dependent weighting matrix. For Theorem 3.3, $\hat{W}$ is the deterministic matrix $Z'Z$ while, in the extension discussed in Section 3.3, $\hat{W}$ is defined in (9). In both cases, $\hat{W} \xrightarrow{p} W^*$ for some positive definite matrix W^* under the strong instrument asymptotics in the theorem. For this W^*, define the oracle weights

$$w_i^* = \pi_i \frac{\pi' W^* e_i}{\pi' W^* \pi} = \frac{\pi' W^* e_i e_i' \pi}{\pi' W^* \pi}$$

and the oracle estimator

$$\hat{\beta}_{RB}^o = \hat{\beta}_{RB}(\xi, \Sigma; w^*) = \hat{\beta}_w(\xi, \Sigma; w^*) = \sum_{i=1}^k w_i^* \hat{\beta}_U(\xi(i), \Sigma(i)).$$

Define the estimated weights as in (10):

$$\hat{w}_i^* = \hat{w}_i^*(\xi^{(b)}) = \frac{\xi_2^{(b)\prime} \hat{W}(\xi^{(b)}) e_i e_i' \xi_2^{(b)}}{\xi_2^{(b)\prime} \hat{W}(\xi^{(b)}) \xi_2^{(b)}}$$

and the Rao-Blackwellized estimator based on the estimated weights

$$\hat{\beta}_{RB}^* = \hat{\beta}_{RB}(\xi, \Sigma; \hat{w}^*) = E\left[\hat{\beta}_w(\xi^{(a)}, 2\Sigma; \hat{w}^*) \middle| \xi \right] = \sum_{i=1}^k E\left[\hat{w}_i^*(\xi_2^{(b)}) \hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) \middle| \xi \right].$$

Let us also define the oracle linear combination of 2SLS estimators

$$\hat{\beta}_{2SLS}^o = \sum_{i=1}^k w_i^* \frac{\xi_{1,i}}{\xi_{2,i}}.$$

Lemma 5. *Suppose that $\hat{w}$ is deterministic: $\hat{w}(\xi^{(b)}) = w$ for some constant vector w. Then $\hat{\beta}_{RB}(\xi, \Sigma; w) = \hat{\beta}_w(\xi, \Sigma; w)$.*

Proof. We have

$$\hat{\beta}_{RB}(\xi, \Sigma; w) = E\left[\sum_{i=1}^k w_i \hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) \middle| \xi \right] = \sum_{i=1}^k w_i E\left[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) \middle| \xi \right].$$

Since $\xi^{(a)}(i) = \zeta(i) + \xi(i)$ (where $\zeta(i) = (\zeta_i, \zeta_{k+i})'$), $\xi^{(a)}(i)$ is independent of $\{\xi(j)\}_{j\neq i}$ conditional on $\xi(i)$. Thus, $E\left[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))\middle|\xi\right] = E\left[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))\middle|\xi(i)\right]$. Since $E\left[\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))\middle|\xi(i)\right]$ is an unbiased estimator for β that is a deterministic function of $\xi(i)$, it must be equal to $\hat{\beta}_U(\xi(i), \Sigma(i))$, the unique nonrandom unbiased estimator based on $\xi(i)$ (where uniqueness follows by completeness since the parameter space $\{(\beta\pi_i, \pi_i)|\pi_i \in \mathbb{R}_+, \beta \in \mathbb{R}\}$ contains an open rectangle). Plugging this in to the above display gives the result. $\square$

Lemma 6. *Let $\|\pi\| \to \infty$ with $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$. Then $\|\pi\|\left(\hat{\beta}_{GMM,W} - \hat{\beta}^o_{2SLS}\right) \xrightarrow{p} 0$.*

Proof. Note that

$$\hat{\beta}_{GMM,W} - \hat{\beta}^o_{2SLS} = \frac{\xi_2'\hat{W}\xi_1}{\xi_2'\hat{W}\xi_2} - \sum_{i=1}^k w_i^* \frac{\xi_{1,i}}{\xi_{2,i}} = \sum_{i=1}^k \left(\frac{\xi_2'\hat{W}e_ie_i'\xi_2}{\xi_2'\hat{W}\xi_2} - w_i^*\right)\frac{\xi_{1,i}}{\xi_{2,i}}$$
$$= \sum_{i=1}^k \left(\frac{\xi_2'\hat{W}e_ie_i'\xi_2}{\xi_2'\hat{W}\xi_2} - \frac{\pi'W^*e_ie_i'\pi}{\pi'W^*\pi}\right)\frac{\xi_{1,i}}{\xi_{2,i}} = \sum_{i=1}^k \left(\frac{\xi_2'\hat{W}e_ie_i'\xi_2}{\xi_2'\hat{W}\xi_2} - \frac{\pi'W^*e_ie_i'\pi}{\pi'W^*\pi}\right)\left(\frac{\xi_{1,i}}{\xi_{2,i}} - \beta\right),$$

where the last equality follows since $\sum_{i=1}^k \frac{\xi_2'\hat{W}e_ie_i'\xi_2}{\xi_2'\hat{W}\xi_2} = \sum_{i=1}^k \frac{\pi'W^*e_ie_i'\pi}{\pi'W^*\pi} = 1$ with probability one. For each i, $\pi_i(\xi_{1,i}/\xi_{2,i} - \beta) = \mathcal{O}_P(1)$ and $\frac{\xi_2'\hat{W}e_ie_i'\xi_2}{\xi_2'\hat{W}\xi_2} - \frac{\pi'W^*e_ie_i'\pi}{\pi'W^*\pi} \xrightarrow{p} 0$ as the elements of π approach infinity. Combining this with the above display and the fact that $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$ gives the result. $\square$

Lemma 7. *Let $\|\pi\| \to \infty$ with $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$. Then $\|\pi\|\left(\hat{\beta}^o_{2SLS} - \hat{\beta}^o_{RB}\right) \xrightarrow{p} 0$.*

Proof. By Lemma 5,

$$\|\pi\|\left(\hat{\beta}^o_{2SLS} - \hat{\beta}^o_{RB}\right) = \|\pi\|\sum_{i=1}^k w_i^*\left(\frac{\xi_{1,i}}{\xi_{2,i}} - \hat{\beta}_U(\xi(i), \Sigma(i))\right).$$

By Theorem 2, $\pi_i\left(\frac{\xi_{1,i}}{\xi_{2,i}} - \hat{\beta}_U(\xi(i), \Sigma(i))\right) \xrightarrow{p} 0$. Combining this with the boundedness of $\|\pi\|/\min_i \pi_i$ gives the result. $\square$

Lemma 8. *Let $\|\pi\| \to \infty$ with $\|\pi\|/\min_i \pi_i = \mathcal{O}(1)$. Then $\|\pi\|\left(\hat{\beta}^o_{RB} - \hat{\beta}^*_{RB}\right) \xrightarrow{p} 0$.*

Proof. We have

$$\hat{\beta}^{o}_{RB} - \hat{\beta}^{*}_{RB} = \sum_{i=1}^{k} E\left[\left(w_i^* - \hat{w}_i^*(\xi^{(b)})\right)\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i))\middle|\xi\right]$$

$$= \sum_{i=1}^{k} E\left[\left(w_i^* - \hat{w}_i^*(\xi^{(b)})\right)\left(\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) - \beta\right)\middle|\xi\right]$$

using the fact that $\sum_{i=1}^{k} w_i^* = \sum_{i=1}^{k} \hat{w}_i^*(\xi^{(b)}) = 1$ with probability one. Thus,

$$E_{\beta,\pi}\left|\hat{\beta}^{o}_{RB} - \hat{\beta}^{*}_{RB}\right| \leq \sum_{i=1}^{k} E_{\beta,\pi}\left|\left(w_i^* - \hat{w}_i^*(\xi^{(b)})\right)\left(\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) - \beta\right)\right|$$

$$= \sum_{i=1}^{k} E_{\beta,\pi}\left|w_i^* - \hat{w}_i^*(\xi^{(b)})\right| E_{\beta,\pi}\left|\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) - \beta\right|.$$

As $\|\pi\| \to \infty$, $\hat{w}_i^*(\xi^{(b)}) - w_i^* \xrightarrow{p} 0$ so, since $\hat{w}_{1,i}^*(\xi^{(b)})$ is bounded by $\sup_{\|u\|_W=1} u'We_ie_i'u$, $E_{\beta,\pi}\left|w_{1,i}^* - \hat{w}_{1,i}^*(\xi^{(b)})\right| \to 0$. Thus, it suffices to show that $\pi_i E_{\beta,\pi}\left|\hat{\beta}_U(\xi^{(a)}(i), 2\Sigma(i)) - \beta\right| = \mathcal{O}(1)$ for each i. But this follows by Lemma 4, which completes the proof. □

B Relation to Hirano & Porter (2015)

Hirano & Porter (2015) give a negative result establishing the impossibility of unbiased, quantile unbiased, or translation equivariant estimation in a wide variety of models with singularities, including many linear IV models. On initial inspection our derivation of an unbiased estimator for β may appear to contradict the results of Hirano & Porter. In fact, however, one of the key assumptions of Hirano & Porter (2015) no longer applies once we assume that the sign of the first stage is known.

Again consider the linear IV model with a single instrument, where for simplicity we let $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_{12} = 0$. To discuss the the results of Hirano & Porter (2015), it will be helpful to parameterize the model in terms of the reduced-form parameters $(\psi, \pi) = (\pi\beta, \pi)$. For ϕ again the standard normal density, the density of ξ is

$$f(\xi; \psi, \pi) = \phi(\xi_1 - \psi)\,\phi(\xi_2 - \pi).$$

Fix some value ψ^*. For any $\pi \neq 0$ we can define $\beta(\psi, \pi) = \frac{\psi}{\pi}$. If we consider any sequence $\{\pi_j\}_{j=1}^{\infty}$ approaching zero from the right, then $\beta(\psi^*, \pi_j) \to \infty$ if $\psi^* > 0$ and $\beta(\psi^*, \pi_j) \to -\infty$ if $\psi^* < 0$. Thus we can see that β plays the role of the function κ in Hirano & Porter (2015) equation (2.1).

Hirano & Porter (2015) show that if there exists some finite collection of parameter values $(\psi_{l,d}, \pi_{l,d})$ in the parameter space and non-negative constants $c_{l,d}$ such that their Assumption 2.4,

$$f\left(\xi; \psi^*, 0\right) \leq \sum_{l=1}^{s} c_{l,d} f\left(\xi; \psi_{l,d}, \pi_{l,d}\right) \; \forall \xi,$$

holds, then (since one can easily verify their Assumption 2.3 in the present context) there can exist no unbiased estimator of β.

This dominance condition fails in the linear IV model with a sign restriction. For any $(\psi_{l,d}, \pi_{l,d})$ in the parameter space, we have by definition that $\pi_{l,d} > 0$. For any such $\pi_{l,d}$, however, if we fix ξ_1 and take $\xi_2 \to -\infty$,

$$\lim_{\xi_2 \to -\infty} \frac{\phi\left(\xi_2 - \pi_{l,d}\right)}{\phi\left(\xi_2\right)} = \lim_{\xi_2 \to -\infty} \exp\left(-\frac{1}{2}\left(\xi_2 - \pi_{l,d}\right)^2 + \frac{1}{2}\xi_2^2\right) = \lim_{\xi_2 \to -\infty} \exp\left(\xi_2 \pi_{l,d} - \frac{1}{2}\pi_{l,d}^2\right) = 0.$$

Thus, $\lim_{\xi_2 \to -\infty} \frac{f\left(\xi; \psi_{l,d}, \pi_{l,d}\right)}{f\left(\xi; \psi^*, 0\right)} = 0$, and for any fixed ξ_1, $\{c_{l,d}\}_{l=1}^{s}$ and $\{(\psi_{l,d}, \pi_{l,d})\}_{l=1}^{s}$ there exists a ξ_2^* such that $\xi_2 < \xi_2^*$ implies

$$f\left(\xi; \psi^*, 0\right) > \sum_{l=1}^{s} c_{l,d} f\left(\xi; \psi_{l,d}, \pi_{l,d}\right).$$

Thus, Assumption 2.4 in Hirano & Porter (2015) fails in this model, allowing the possibility of an unbiased estimator. Note, however, that if we did not impose $\pi > 0$ then we would satisfy Assumption 2.4, so unbiased estimation of β would again be impossible. Thus, the sign restriction on π plays a central role in the construction of the unbiased estimator $\hat{\beta}_U$.

C Equivariance in the Just-Identified Model

For comparisons between $\left(\hat{\beta}_U, \hat{\beta}_{2SLS}, \hat{\beta}_{FULL}\right)$ in the just-identified case, it sufficies to consider a two-dimensional parameter space. To see that this is the case let $\theta =$

$(\beta, \pi, \sigma_1^2, \sigma_{12}, \sigma_2^2)$ be the vector of model parameters and let g_A, for $A = \begin{bmatrix} a_1 & a_2 \\ 0 & a_3 \end{bmatrix}$, $a_1 \neq 0$, $a_3 > 0$, be the transformation

$$g_A \xi = \tilde{\xi} = A \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} a_1 \xi_1 + a_2 \xi_2 \\ a_3 \xi_2 \end{pmatrix}$$

$$g_A \theta = \tilde{\theta} = \left(\tilde{\beta}, \tilde{\pi}, \tilde{\sigma}_1^2, \tilde{\sigma}_{12}, \tilde{\sigma}_2^2 \right)$$

where

$$\tilde{\beta} = \frac{(a_1 \beta + a_2)}{a_3}$$

$$\tilde{\pi} = a_3 \gamma$$

$$\tilde{\sigma}_1^2 = a_1^2 \sigma_1^2 + a_1 a_2 \sigma_{12} + a_2^2 \sigma_2^2$$

$$\tilde{\sigma}_{12} = a_1 a_3 \sigma_{12} + a_2 a_3 \sigma_2^2$$

and

$$\tilde{\sigma}_2^2 = a_3^2 \sigma_2^2.$$

Define $\mathcal{G}$ as the set of all transformations g_A of the form above. Note that the sign restriction on π is preserved under $g_A \in \mathcal{G}$, and that for each g_A, there exists another transformation $g_A^{-1} \in \mathcal{G}$ such that $g_A g_A^{-1}$ is the identity transformation. We can see that the model (2) is invariant under the transformation g_A. Note further that the estimators $\hat{\beta}_U$, $\hat{\beta}_{2SLS}$, and $\hat{\beta}_{FULL}$ are all equivariant under g_A, in the sense that

$$\hat{\beta}\left(g_A \xi\right) = \frac{a_1 \hat{\beta}\left(\xi\right) + a_2}{a_3}.$$

Thus, for any properties of these estimators (e.g. relative mean and median bias, relative dispersion) which are preserved under the transformations g_A, it suffices to study these properties on the reduced parameter space obtained by equivariance. By choosing A appropriately, we can always obtain

$$\begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_2 \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ \tilde{\pi} \end{pmatrix}, \begin{pmatrix} 1 & \tilde{\sigma}_{12} \\ \sigma_{12} & 1 \end{pmatrix} \right)$$

for $\tilde{\pi} > 0$, $\sigma_{12} \geq 0$ and thus reduce to a two-dimensional parameter (π, σ_{12}) with $\sigma_{12} \in [0, 1)$, $\tilde{\gamma} > 0$.

D Dispersion Simulation Results

We simulated 10^6 draws from the distributions of $\hat{\beta}_U$, $\hat{\beta}_{2SLS}$, and $\hat{\beta}_{FULL}$ on a grid formed by the Cartesian product of $\sigma_{12} \in \left\{0, (0.005)^{\frac{1}{2}}, (0.01)^{\frac{1}{2}}, ..., (0.995)^{\frac{1}{2}}\right\}$ and $\pi \in \left\{(0.01)^2, (0.02)^2, ..., 25\right\}$. We use these grids for σ_{12} and π, rather than a uniformly spaced grid, because preliminary simulations suggested that the behavior or the estimators was particularly sensitive to the parameters for large values of σ_{12} and small values of π.

At each point in the grid we calculate $(\varepsilon_U, \varepsilon_{2SLS}, \varepsilon_{FULL})$, using independent draws to calculate ε_U and the other two estimators, and compute a one-sided Kolmogorov-Smirnov statistic for the hypotheses that (i) $|\varepsilon_{IV}| \geq |\varepsilon_U|$ and (ii) $|\varepsilon_U| \geq |\varepsilon_{FULL}|$, where $A \geq B$ for random variables A and B denotes that A is larger than B in the sense of first-order stochastic dominance. In both cases the maximal value of the Kolmogorov-Smirnov statistic is less than 2×10^{-3}. Conventional Kolmogorov-Smirnov p-values are not valid in the present context (since we use estimated medians to construct ε), but are never below 0.25.

We also compare the τ-quantiles of $(|\varepsilon_U|, |\varepsilon_{2SLS}|, |\varepsilon_{FULL}|)$ for $\tau \in \{0.001, 0.002, ..., 0.999\}$ and for $\hat{F}_A^{-1}$ the estimated quantile function of a random variable A find that

$$\max_{\pi,\sigma_{12}} \max_{\tau} \left(\hat{F}^{-1}_{|\varepsilon_U|}(\tau) - \hat{F}^{-1}_{|\varepsilon_{2SLS}|}(\tau)\right) = 0.000086$$

$$\max_{\pi,\sigma_{12}} \max_{\tau} \left(\hat{F}^{-1}_{|\varepsilon_{FULL}|}(\tau) - \hat{F}^{-1}_{|\varepsilon_U|}(\tau)\right) = 0.00028,$$

or in proportional terms

$$\max_{\pi,\sigma_{12}} \max_{\tau} \left(\frac{\hat{F}^{-1}_{|\varepsilon_U|}(\tau) - \hat{F}^{-1}_{|\varepsilon_{2SLS}|}(\tau)}{\hat{F}^{-1}_{|\varepsilon_U|}(\tau)}\right) = 0.006$$

$$\max_{\pi,\sigma_{12}} \max_{\tau} \left(\frac{\hat{F}^{-1}_{|\varepsilon_{FULL}|}(\tau) - \hat{F}^{-1}_{|\varepsilon_U|}(\tau)}{\hat{F}^{-1}_{|\varepsilon_{FULL}|}(\tau)}\right) = 0.06.$$