Location Properties of Point Estimators in Linear Instrumental Variables and Related Models*

Keisuke Hirano
Department of Economics
University of Arizona
hirano@u.arizona.edu

Jack R. Porter
Department of Economics
University of Wisconsin
jrporter@ssc.wisc.edu

December 23, 2011

Abstract

We examine statistical models, including the workhorse linear instrumental variables model, in which the mapping from the reduced form distribution to the structural parameters of interest is singular. The singularity of this mapping implies certain fundamental restrictions on the finite sample properties of point estimators: they cannot be unbiased, quantile-unbiased, or translation equivariant. The nonexistence of unbiased estimators does not rule out bias reduction of standard estimators, but implies that the bias-variance tradeoff cannot be avoided and needs to be considered carefully. The results can also be extended to weak instrument asymptotics by using the limits of experiments framework.

1 Introduction

Linear simultaneous equations models, especially the linear instrumental variable (IV) model, continue to play a central role in empirical economics. It has long been recognized that these models, despite their relatively simple form, present particular challenges for statistical analysis and inference. Standard point estimators for the IV model, including two stage least squares (TSLS), are biased in general, their exact distributions are nonstandard, and conventional first-order asymptotic theory may provide poor approximations to their finite sample properties when instruments are weak. There is a large literature analyzing the exact distributions of estimators

*We thank the National Science Foundation for research support under grants SES-0962488 (Hirano), SES-0962422 (Porter), and SES-0438123 (Porter).

in the IV model, including Sawa (1969), Mariano and Sawa (1972), Nagar and Ullah (1973), Anderson and Sawa (1973) Ullah and Nagar (1974), Carter (1976), Anderson, Kunitomo, and Sawa (1982), Phillips (1984), Hillier, Kinal, and Srivastava (1984), Phillips (1985), Phillips (1986), and Chamberlain (2007). For a general overview of finite sample techniques and results for econometric models, see Ullah (2004). A number of techniques for bias reduction for 2SLS and other estimators have been proposed, including Nagar (1959), Angrist, Imbens, and Krueger (1999), Blomquist and Dahlberg (1999), Iglesias and Phillips (2008), and Ackerberg and Devereux (2009). In addition, a recent literature has developed alternative asymptotic approximations to estimators and inferential procedures that better approximate their finite sample properties when instruments are weakly correlated with included endogenous explanatory variables, including Staiger and Stock (1997), Stock and Wright (2000), Kleibergen (2002), Moreira (2003), and Andrews, Moreira, and Stock (2006).

In this paper we develop new results on the finite sample properties of point estimators in linear IV and related models. We focus on a key feature of these models: the mapping from the reduced form (observable) distribution to the structural parameters of interest is singular, in the sense that it is unbounded in certain neighborhoods in the parameter space. This is the same setup considered in Gleser and Hwang (1987) and Dufour (1997), who show that valid confidence intervals cannot have finite expected length in such models. In the linear IV case, this singularity arises as the correlation between the instruments and the included endogenous explanatory variables approaches zero, and more generally it can arise when structural models are not identified at certain critical parameter values. We show that when structural models exhibit this type of singularity, there cannot exist *any* unbiased, quantile unbiased, or translation equivariant point estimators. As a result, bias reduction procedures cannot fully remove bias, and researchers applying bias reduction techniques will inevitably face a nontrivial bias-variance tradeoff.

The main intuition for our result is that the distribution of any point estimator, and indeed any statistic, is determined by the distribution of the observable data. As long as the observable data distribution satisfies some mild regularity conditions, the location properties (such as the mean and median) of the point estimator must vary smoothly with the observable data distribution. In particular, these location properties are bounded over bounded neighborhoods of parameter values. But if the *target* of the point estimator is singular, it is unbounded as a function of the observable data distribution, and the location of the point estimator cannot match that of its target. As a result, properties such as unbiasedness and median unbiasedness are not attainable. Our results complement those of Gleser and Hwang (1987) and Dufour (1997) by building a fuller picture of the implications of singularities on statistical estimation and inference.

Our conclusions are similar to the impossibility results in Hirano and Porter (2011) for point estimators in bounds analysis and related problems, in which the quantity of interest is directionally differentiable, but not differentiable in the ordinary sense, at certain points in the parameter space. The case of singular mappings, which we consider here, does not fall under the framework of Hirano and Porter (2011) and requires a different method of proof.

We also extend our finite sample results to the weak instrument asymptotics of Staiger and Stock (1997). Cattaneo, Crump, and Jansson (2011) show that under the Staiger-Stock local asymptotics, there is a certain normal limit experiment that characterizes the set of attainable limit distributions for statistical procedures. This limit experiment exhibits the same type of singularity we have considered in our finite sample theory, so our impossibility results imply that the limit distributions of point estimators under the weak instrument asymptotics cannot generally be unbiased, quantile unbiased, or equivariant.

2 Exact Analysis of Parametric Models with Singularities

In this section, we develop results for parametric models where the parameter of interest has a singularity with respect to the observed-data distribution. We first give some examples that illustrate the issue we wish to examine and the challenges it creates.

Example 2.1 *(Inverse Normal Mean) Suppose $Z \sim N(\mu, \sigma^2)$ with $\sigma^2 > 0$ known, and suppose we wish to estimate*

$$\kappa(\mu) = \frac{1}{\mu}.$$

Let the parameter space for μ be $\mathcal{M}_1 = (0, \infty)$. Then $\kappa(\mu)$ can take values in $(0, \infty)$. As $\mu \downarrow 0$, $\kappa(\mu) \to \infty$. The natural estimator $\hat{\kappa} = Z^{-1}$ is biased.

Example 2.2 *(Linear Instrumental Variables) Consider the linear simultaneous equations model*

$$\begin{aligned} y_{i1} &= x_i' \pi_1 + v_{i1}, \\ y_{i2} &= \beta y_{i1} + \epsilon_i, \end{aligned}$$

where x_i is $k \times 1$, and y_{i1} and y_{i2} are scalar. Write this in reduced form as:

$$\begin{aligned} y_{i1} &= x_i' \pi_1 + v_{i1}, \\ y_{i2} &= x_i' \pi_2 + v_{i2}, \end{aligned}$$

with

$$\pi_2 = \beta \pi_1.$$

Let $v_i = (v_{i1}, v_{i2})'$, with $v_i \sim N(0,\Omega)$ independent of x_i. Suppose we have a random sample of size n from this model. Let X be the $n \times k$ matrix of exogenous regressors, and let Y_1 and Y_2 denote the $n \times 1$ vectors of endogenous variables. We can write the model in stacked form as

$$Y|X \sim N(Q\pi, \Omega \otimes I_n),$$

where

$$\pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}, \quad \textit{and} \quad Q = \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}.$$

This can be viewed as a (conditional) model for Y given X with parameters $\theta = (\pi', vech(\Omega)')'$. In general, the restriction that $\pi_2 = \beta\pi_1$ restricts the set of possible values for θ. For values of θ satisfying this restriction, we can view β as a function of the reduced form parameters θ, and write $\beta = \beta(\theta)$. If π_2 is fixed and $\pi_1 \to \underline{0}$, we have $\beta(\theta) \to \infty$ or $\beta(\theta) \to -\infty$. This corresponds to cases where the instruments x_i are weakly correlated with the endogenous variable y_{i1} (and β is allowed to be large); the failure of identification when $\pi_1 \to \underline{0}$ leads to a singularity in the mapping from reduced-form to structural coefficients.

We can also consider extensions and variations of this model, for example linear simultaneous equations models with exogenous regressors that are present in both structural equations, or models with a different (parametric) specification for the conditional distribution of v_i given x_i. We can still define θ as the collection of reduced form parameters and define the mapping $\beta(\theta)$ analogously to the discussion above.

We develop a general framework that encompasses these examples and other models where identification fails at certain parameter values. Our setup is similar to Gleser and Hwang (1987), who consider the implications of singularities for confidence intervals (see also Dufour (1997)). Here, we will focus on point estimation. Let the data (vector) be $Z \sim P_\theta$, where $\theta \in \Theta \subset \mathbb{R}^k$, and let $\mathcal{Z}$ denote the support of Z, which we assume is common for all values of θ.

We will suppose that the collection of probability measures $\{P_\theta\}$ are dominated by a measure ν, and let $p_\theta(z)$ denote the density of P_θ with respect to ν. In addition, we will assume continuity of the densities in θ:

Assumption 2.3 *$p_\theta(z)$ is continuous in θ almost everywhere ν.*

We are interested in some scalar function of the parameter $\kappa(\theta)$ with $\kappa : \Theta_1 \to \mathbb{R}$ and $\Theta_1 \subset \Theta$. Suppose that for some finite $\theta^* \in \Theta$

$$\kappa(\theta) \to \pm\infty \quad \text{as} \quad \theta \to \theta^*. \tag{2.1}$$

Here, θ^* need not be in Θ_1. For example, in the normal model of Example 2.1, the problematic parameter value is $\mu^* = 0$.

The unboundedness of $\kappa(\theta)$ has a number of important implications for point estimators of κ. Heuristically, any point estimator $\hat{\kappa}$ is a function of the data, so that many characteristics of the distribution of $\hat{\kappa}$ will vary smoothly with the underlying parameter θ, provided that the densities $p_\theta(z)$ are sufficiently well behaved. As a result, properties such as unbiasedness for estimating $\kappa(\theta)$ cannot be maintained as θ approaches θ^*.

Let $\{\theta_j\}_{j=1}^{\infty} \subset \Theta_1$ be a sequence with $\theta_j \to \theta^*$ as $j \to \infty$. We will assume that the densities $p_{\theta_j}(z)$ along this sequence are dominated by a linear combination of the densities at fixed values of θ:

Assumption 2.4 *For some finite positive integer s, there exist parameters $\theta^{d,1}, \ldots, \theta^{d,s} \in \Theta_1$ and constants $c_1, \ldots, c_s$ such that for all $j = 1, \ldots, \infty$, and all $z \in \mathcal{Z}$,*

$$p_{\theta_j}(z) \leq \overline{q}(z) := \sum_{l=1}^{s} c_l p_{\theta^{d,l}}(z).$$

This is a mild condition, and is often easy to verify. For example, it holds in standard exponential families by Lemma 2.1 of Brown (1986).

We consider randomized point estimators $\hat{\kappa}$ for κ. Formally, $\hat{\kappa}$ is a mapping $\hat{\kappa} : \mathcal{Z} \times [0,1] \to \mathbb{R}$ with the interpretation that $\hat{\kappa} = \hat{\kappa}(Z, U)$ where $Z \sim P_\theta$ and $U \sim \text{Unif}[0,1]$, with U independent of Z. This is equivalent to assuming that that $\hat{\kappa}$ is a Markov kernel from $\mathcal{Z}$ to the set of probability measures on $\mathbb{R}$.

We will examine three location properties of estimators of κ. The estimator $\hat{\kappa}$ is *unbiased* for $\kappa(\theta)$ if, for all $\theta \in \Theta_1$,

$$\kappa(\theta) = E_\theta\left[\hat{\kappa}(Z, U)\right].$$

For $\alpha \in (0,1)$, the estimator is *α-quantile unbiased* for $\kappa(\theta)$ if, for all $\theta \in \Theta_1$,

$$\alpha = \Pr_\theta\left[\hat{\kappa} \leq \kappa(\theta)\right].$$

Finally, the estimator is *translation equivariant* if

$$(\hat{\kappa} - \kappa(\theta)) \sim \mathcal{L},$$

where the law $\mathcal{L}$ does not depend on θ for any $\theta \in \Theta_1$. This implies that $(\hat{\kappa} - \kappa(\theta))$ is a pivot, which is useful for statistical inference. We will show that when κ has a singularity, as in (2.1), then there will exist no estimators of κ with any of the above properties.

To illustrate the argument, suppose we have an estimator $\hat{\kappa}$ for κ. Let $\tau(\theta)$ denote the expected value of $\hat{\kappa}$,

$$\tau(\theta) := \iint\limits_{\mathcal{Z}\times[0,1]} \hat{\kappa}(z,u)p_\theta(z)d\nu(z)du,$$

which we will assume exists for all $\theta \in \Theta_1$. By Assumption 2.3, $p_{\theta_j}(z) \to p_{\theta^*}(z)$, pointwise in z, as $\theta_j \to \theta^*$. By Assumption 2.4 and the existence of $\tau(\theta^{d,l})$ for all $l = 1, \ldots, s$, the Dominated Convergence Theorem can be applied to conclude that

$$\tau(\theta_j) = \iint\limits_{\mathcal{Z}\times[0,1]} \hat{\kappa}(z,u)p_{\theta_j}(z)d\nu(z)du \longrightarrow \iint\limits_{\mathcal{Z}\times[0,1]} \hat{\kappa}(z,u)p_{\theta^*}(z)d\nu(z)du < \infty.$$

Hence, for j large enough, $\tau(\theta_j)$ is bounded above by $1 + \iint_{\mathcal{Z}\times[0,1]} \hat{\kappa}(z,u)p_{\theta^*}(z)d\nu(z)du$, which is finite. Since $\tau(\theta_j)$ is bounded, it clearly cannot match the behavior of $\kappa(\theta_j)$ as $j \to \infty$. Hence $\hat{\kappa}$ cannot be an unbiased estimator of κ.

It is known that some estimators in the instrumental variables model can fail to possess moments; see for example Fuller (1977), Kinal (1980), and Davidson and MacKinnon (2007). But similar arguments can be applied for the other location properties we consider, leading to the following result.

Theorem 2.5 *Suppose $\kappa(\theta_j) \to \pm\infty$ as $j \to \infty$, where $\{\theta_j\} \subset \Theta_1$ is a convergent sequence of parameter values. If Assumptions 2.3 and 2.4 hold, then:*

(a) There exists no unbiased estimator for κ.

(b) For any $\alpha \in (0,1)$, there exists no α-quantile unbiased estimator for κ.

(c) There exists no translation equivariant estimator for κ.

Proof: We prove the result for the case where $\kappa(\theta_j) \to \infty$. The case where $\kappa(\theta_j) \to -\infty$ follows by obvious modifications of the arguments.

(a) This was shown above.

(b) Let $\hat{\kappa}$ be a randomized estimator. For any (z,u), $1\{\hat{\kappa}(z,u) \le \kappa(\theta_j)\} \to 1$ as $j \to \infty$. Then, by the Dominated Convergence Theorem,

$$\Pr_{\theta_j}(\hat{\kappa} \le \kappa(\theta_j)) = \iint\limits_{\mathcal{Z}\times[0,1]} 1\{\hat{\kappa}(z,u) \le \kappa(\theta_j)\}p_{\theta_j}(z)d\nu(z)du \longrightarrow \iint\limits_{\mathcal{Z}\times[0,1]} 1\cdot p_{\theta^*}(z)d\nu(z)du = 1.$$

It follows that for $\alpha < 1$, we cannot have $\Pr_{\theta_j}(\hat{\kappa} \le \kappa(\theta_j)) = \alpha$ for j large enough. Hence, $\hat{\kappa}$ cannot be α-quantile unbiased.

(c) Suppose $\hat{\kappa}$ is a translation equivariant estimator for κ. Let $\psi_\theta(s) = E_\theta\left[e^{is(\hat{\kappa}-\kappa(\theta))}\right]$ denote the characteristic function of $\hat{\kappa}-\kappa(\theta)$ under P_θ. Since $\hat{\kappa}$ is translation equivariant, $\psi_\theta(s)$ does not depend on θ and can be denoted simply as $\psi(s)$. Since $|\kappa(\theta_j)| \to \infty$, $\exp(is\kappa(\theta_j))$ does not converge for some choice of s in any neighborhood of zero. On the other hand, for all s, by the Dominated Convergence Theorem,

$$
\begin{aligned}
e^{is\kappa(\theta_j)}\psi_{\theta_j}(s) &= e^{is\kappa(\theta_j)} \iint\limits_{\mathcal{Z}\times[0,1]} e^{is[\hat{\kappa}(z,u)-\kappa(\theta_j)]} p_{\theta_j}(z)\,d\nu(z)\,du \\
&= \iint\limits_{\mathcal{Z}\times[0,1]} \exp(is\hat{\kappa}(z,u))\, p_{\theta_j}(z)\,d\nu(z)\,du \\
&\longrightarrow \iint\limits_{\mathcal{Z}\times[0,1]} \exp(is\hat{\kappa}(z,u))\, p_{\theta^*}(z)\,d\nu(z)\,du,
\end{aligned}
$$

which contradicts the divergence of $e^{is\kappa(\theta_j)}\psi(s)$ as $j \to \infty$. Hence, no such translation equivariant estimator exists.

□

To illustrate the result, let us revisit Example 2.2, the linear instrumental variables model with normal disturbances. Let $Z = (Y_1, Y_2, X)$, and recall that θ is the vector of reduced form parameters. Any point estimator $\hat{\beta}$ is a function of Z and possibly some auxiliary randomization. As noted above, if we permit values of θ such that π_1 can be arbitrarily close to the zero vector while $\pi_2 \neq 0$, then the function $\beta(\theta)$ is unbounded. By Theorem 2.5 we can conclude that the estimator $\hat{\beta}$ cannot be unbiased, translation equivariant, or quantile unbiased. While the bias and other finite-sample properties of particular estimators such as the two-stage least squares estimator have been extensively studied, our result shows that *no* estimator can be unbiased or median unbiased.

When unbiased estimators do not exist, Doss and Sethuraman (1989) show that any sequence of estimators such that the bias is converging toward zero necessarily has variance approaching infinity. Bias reduction may still be desirable, but there is a tradeoff between bias and variance that cannot be avoided.

Theorem 2.5 rules out median unbiased estimation of κ, since this is a special case of quantile unbiasedness. The impossibility of α-quantile unbiased estimation has the further implication that there do not exist similar one-sided confidence intervals for any coverage probability; that is, there exist no confidence intervals of the form $CI = (-\infty, \hat{\kappa}]$ such that

$$Pr_{\theta}(\kappa(\theta) \in CI) = \alpha, \quad \forall \theta \in \Theta_1.$$

3 Exact Analysis of Infinite-Dimensional Models with Singularities

The results of the previous section apply to parametric models, and in practice are easiest to apply when the model is relatively tightly parametrized as in Examples 2.1 and 2.2. In this section we point out how they can be extended to larger classes of models, including infinite-dimensional models, when a suitable parametric model can be embedded into the full model. We adopt an approach used in the semiparametric efficiency bound literature (Newey 1990) where parametric submodels of a larger model space are taken. This argument was also used in Hirano and Porter (2011) to obtain large-sample impossibility results for infinite-dimensional models. But in contrast to those papers, we stay within the finite sample setting and obtain exact results.

The statistical model is a collection of probability measures on $\mathscr{Z}$, denoted by $\mathscr{P}$. The set $\mathscr{P}$ is allowed to be general, and could be infinite-dimensional, corresponding to a nonparametric or semiparametric model. Suppose we are interested in estimating a real-valued functional $\kappa[P]$ of the data-generating probability measure, defined on some subset $\mathscr{P}_1$ of $\mathscr{P}$. As before, we consider randomized point estimators $\hat{\kappa}$ for κ. We can define location properties properties for point estimators analogously to the definitions in Section 2. The estimator $\hat{\kappa}$ is *unbiased* for $\kappa[P]$ if, for all $P \in \mathscr{P}_1$,

$$\kappa[P] = E_P[\hat{\kappa}].$$

For $\alpha \in (0,1)$, the estimator is *α-quantile unbiased* for $\kappa[P]$ if, for all $P \in \mathscr{P}_1$,

$$\alpha = P(\hat{\kappa} \leq \kappa[P]).$$

Finally, the estimator is *translation equivariant* if

$$(\hat{\kappa} - \kappa[P]) \sim \mathscr{L},$$

where the law $\mathscr{L}$ does not depend on P for any $P \in \mathscr{P}_1$.

Within the full collection of probability measures $\mathscr{P}$, we consider finite-dimensional (parametric) submodels. A parametric submodel is a set $\mathscr{P}_f \subset \mathscr{P}$ that has a finite-dimensional parametrization $\mathscr{P}_f = \{P_\theta : \theta \in \Theta\}$ as in Section 2.

Consider a parametric submodel $\mathscr{P}_f$ that overlaps with $\mathscr{P}_1$, i.e. $\mathscr{P}_f \cap \mathscr{P}_1 \neq \emptyset$, and let $\Theta_1 = \{\theta \in \Theta : P_\theta \in \mathscr{P}_f \cap \mathscr{P}_1\}$. We can define a real-valued function κ_f on Θ_1 as $\kappa_f(\theta) = \kappa[P_\theta]$ for $\theta \in \Theta_1$.

As before we are interested in singular functionals κ. Suppose that the parametric submodel *contains* the singularity in the sense that for some convergent sequence $\{\theta_j\}_{j=1}^{\infty} \subset \Theta_1$,

$$\kappa[P_{\theta_j}] \to \pm\infty \tag{3.1}$$

as $j \to \infty$, where $\theta_j \to \theta^*$ for some $\theta^* \in \Theta$. So $\kappa[P]$ has a singularity that is contained in the submodel $\mathscr{P}_f$. If the assumptions of Theorem 2.5 are satisfied on the parametric submodel, then the analysis of Section 2 applies, and there do not exist estimators with any of the three location properties over $\mathscr{P}_f$. *A fortiori*, no estimator can have those properties over the larger set $\mathscr{P}_1$.

Corollary 3.1 *Suppose the data vector $Z \sim P \in \mathscr{P}$. Consider a parametric submodel $\mathscr{P}_f = \{P_\theta : \theta \in \Theta\} \subset \mathscr{P}$, where the measures P_θ are dominated by a measure ν, and p_θ denotes the density of P_θ with respect to ν. Suppose κ satisfies (3.1) and Θ_1 is defined as above. If Assumptions 2.3 and 2.4 are satisfied on $\mathscr{P}_f$, then the conclusion of Theorem 2.5 holds on $\mathscr{P}_1$.*

Corollary 3.1 provides results on the finite-sample properties of estimators in infinite-dimensional models. As an illustration, consider the linear IV model of Example 2.2, but suppose that the distribution of v_i conditional on x_i is only assumed to have mean zero and bounded second moments. This is a semiparametric model where the unknown parameters are π_1, π_2, and the conditional distribution of v_i. The $N(0,\Omega)$ specification forms a parametric submodel of the larger model, so the impossibility of unbiased, quantile-unbiased, and translation equivariant point estimation of the structural coefficient β over the submodel with normal disturbances implies the same for the semiparametric model.

4 Extension to Weak Instrument Asymptotics

Our results so far are exact in nature. Consider the simple instrumental variables model of Example 2.2. Under conventional first-order asymptotics, if $\pi_1 \neq \underline{0}$, usual estimators such as the two-stage least squares estimator are consistent and asymptotically normal for β. However, it is well known that such approximations can be very poor when π_1 is close to $\underline{0}$, or equivalently, when the instruments are weak. See, for example, Nelson and Startz (1990) and Bound, Jaeger, and Baker (1995).

Staiger and Stock (1997) propose an alternative asymptotic approximation to the distribution of statistical procedures when instruments are weak. A large literature has analyzed statistical procedures and proposed new procedures under the Staiger-Stock asymptotics or variants of it, including Stock and Wright (2000), Kleibergen (2002), Moreira (2003), and Andrews, Moreira, and Stock (2006).

The finite sample results of the previous section can be extended to weak instrument asymptotics using Le Cam's limits of experiments theory. Cattaneo, Crump, and Jansson (2011) use the limits of experiments approach to establish a large-sample equivalence between the linear instrumental variables model with weak instruments and a certain shifted normal model, and apply the results to study and develop inference procedures. We will use their results to connect our finite sample theory from the previous section to weak instrument asymptotics and derive implications for point estimators.

To illustrate the arguments, we consider the following simple linear two-equation model: for $i = 1, \dots, n$,

$$
\begin{aligned}
y_{i1} &= x_i' \pi_n + v_{i1}, \\
y_{i2} &= \beta y_{i1} + \epsilon_{i2},
\end{aligned}
$$

where x_i is $k \times 1$, and both y_{i1} and y_{i2} are scalar. This is the model of Example 2.2, but following Staiger and Stock (1997) we capture the weak instrument problem by modeling the first stage coefficient as

$$
\pi_n = h/\sqrt{n},
$$

where the local parameter h can take any value in $\mathbb{R}^k$.

Write the model in reduced form as

$$
\begin{aligned}
y_{i1} &= x_i' \pi_n + v_{i1}, \\
y_{i2} &= x_i' (\beta \pi_n) + v_{i2}.
\end{aligned}
$$

Following Cattaneo, Crump, and Jansson (2011) we assume that the $\{x_i\}$ form a fixed sequence and that $v_i = (v_{i1}, v_{i2})'$ are i.i.d. with mean zero, bounded variance, and Lebesgue density $f(v)$. Let $\mathscr{I}$ be the Fisher information in the location model $W = \gamma + v$, where v has density $f(v)$.

Under the Staiger-Stock weak instrument asymptotics, estimator sequences $\hat{\beta}_n$ of β will typically have unscaled limit distributions:

$$
\hat{\beta}_n \overset{h,\beta}{\rightsquigarrow} \mathscr{L}(h, \beta), \tag{4.1}
$$

where $\overset{h,\beta}{\rightsquigarrow}$ denotes weak convergence (convergence in distribution) under the sequence of probability measures induced by $f(v)$, the sequence $\{x_i\}$, and the parameters $\pi_n = h/\sqrt{n}$ and β. The expression $\mathscr{L}(h, \beta)$ denotes the limit law of $\hat{\beta}$, which could depend on h and β. (The laws will typically be nondegenerate, but this is not required for our conclusions below.) Although $\hat{\beta}_n$ cannot be consistent for β, we can still study its distributional properties through the weak instrument approximations.

Let $Q_{xx} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i x_i'$, which is assumed to exist and be positive definite. Theorem 1 of Cattaneo, Crump, and Jansson (2011) shows that under fairly standard conditions on $\{x_i\}$ and $f(v)$ (their Assumptions 1 and 2), the likelihood ratio process of the model converges in large samples to the likelihood ratio process of the following simple model: we observe a single draw for Z, where

$$Z \sim N\left(\begin{pmatrix} h \\ \beta h \end{pmatrix}, \mathscr{I}^{-1} \otimes Q_{xx}^{-1}\right). \tag{4.2}$$

Here, Z is of dimension $2k$. Therefore, by Le Cam's asymptotic representation theorem (see for example Theorem 9.3 in van der Vaart (1998)), we have the following characterization of the limit laws of point estimators:

Proposition 4.1 *Consider any estimator sequence $\hat{\beta}_n$ satisfying (4.1) for every h, β. Let Assumptions 1 and 2 of Cattaneo, Crump, and Jansson (2011) hold. Then there exists a randomized estimator $T(Z,U)$ in the model (4.2), with $U \sim$ Unif$[0,1]$ and independent of Z, such that the distribution of $T(Z,U)$ under h and β is equal to the law $\mathscr{L}(h,\beta)$, for every h,β.*

This says that any estimator sequence in the weak IV model with limit distributions can be represented by some (randomized) estimator in the normal shift model (4.2). If, for example, $\hat{\beta}_n$ is *asymptotically unbiased* for β in the sense that $\mathscr{L}(h,\beta)$ has mean β for every h,β, then there must exist an *exactly* unbiased estimator of β in the normal shift model. However, it is easy to verify that the normal shift model satisfies the assumptions in Section 2, so Theorem 2.5 rules out unbiased, quantile-unbiased, or translation equivariant estimators. Thus, passing to the limit under weak instruments asymptotics preserves the conclusions of our finite sample analysis.

Corollary 4.2 *In the weak IV model, there exist no estimator sequences satisfying (4.1) with any of the following properties:*

(a) *Asymptotic unbiasedness: $\mathscr{L}(h,\beta)$ has expectation β for all h,β.*

(b) *Asymptotic quantile-unbiasedness: for $\alpha \in (0,1)$, the law $\mathscr{L}(h,\beta)$ has α-quantile β for all h,β.*

(c) *Regularity: the shifted laws $\mathscr{L}(h,\beta) - \beta$ do not depend on h,β.*

5 Conclusion

While econometric models are often constructed by specifying the mapping from structural parameters to the reduced form, it is the inverse of this mapping that is key to understanding identification and the properties of statistical procedures. Our results complement those of Gleser and Hwang (1987) and Dufour (1997) by developing broad implications, not limited to specific statistical procedures, of singularities in the inverse mapping. Properties such as unbiasedness, quantile unbiasedness, and translation equivariance cannot be attained by *any* procedure in this class of models. As a consequence, efforts to more fully quantify, for example, the bias-variance tradeoff in finite samples, would be useful. And, as we have illustrated above, finite sample theory for singular parametric models can be extended via the limits of experiments approach to weak instrument asymptotics. This suggests a renewed role for finite sample theory in the future.

References

ACKERBERG, D. A., AND P. J. DEVEREUX (2009): "Improved JIVE Estimators for Overidentified Linear Models With and Without Heteroskedasticity," *The Review of Economics and Statistics*, 91(2), 351–362.

ANDERSON, T. W., N. KUNITOMO, AND T. SAWA (1982): "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, 50(4), 1009–1027.

ANDERSON, T. W., AND T. SAWA (1973): "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions," *Econometrica*, 41(4), 683–714.

ANDREWS, D. W. K., M. J. MOREIRA, AND J. H. STOCK (2006): "Optimal Two-Sided Invariant Similar Tests for Instrumental Variables Regression," *Econometrica*, 74(3), 715–752.

ANGRIST, J. D., G. W. IMBENS, AND A. B. KRUEGER (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14(1), 57–67.

BLOMQUIST, S., AND M. DAHLBERG (1999): "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments," *Journal of Applied Econometrics*, 14(1), 69–88.

BOUND, J., D. A. JAEGER, AND R. M. BAKER (1995): "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90(430), 443–450.

BROWN, L. D. (1986): *Fundamentals of Statistical Exponential Families*, vol. 9 of *IMS Lecture Notes-Monograph Series*. Institute of Mathematical Statistics, Hayward, CA.

CARTER, R. A. L. (1976): "The Exact Distribution of an Instrumental Variables Estimator," *International Economic Review*, 17(1), 228–233.

CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2011): "Optimal Inference for Instrumental Variables Regression with Non-Gaussian Errors," Working paper, University of Michigan.

CHAMBERLAIN, G. (2007): "Decision Theory Applied to an Instrumental Variables Model," *Econometrica*, 75(3), 609–652.

DAVIDSON, R., AND J. G. MACKINNON (2007): "Moments of IV and JIVE Estimators," *Econometrics Journal*, 10(3), 541–553.

DOSS, H., AND J. SETHURAMAN (1989): "The Price of Bias Reduction when there is no Unbiased Estimate," *The Annals of Statistics*, 17(1), 440–442.

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65(6), 1365–1387.

FULLER, W. A. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45(4), 939–953.

GLESER, L. J., AND J. T. HWANG (1987): "The Nonexistence of $100(1-\alpha)\%$ Confidence Sets of Finite Expected Diameter in Errors-in-Variables and Related Models," *The Annals of Statistics*, 15(4), 1351–1362.

HILLIER, G. H., T. W. KINAL, AND V. K. SRIVASTAVA (1984): "On the Moments of Ordinary Least Squares and Instrumental Variables Estimators in a General Structural Equation," *Econometrica*, 52(1), 185–202.

HIRANO, K., AND J. R. PORTER (2011): "Impossibility Results for Nondifferentiable Functionals," forthcoming, *Econometrica*.

IGLESIAS, E. M., AND G. D. A. PHILLIPS (2008): "Almost Unbiased Estimation in Simultaneous Equations Models with Strong and/or Weak Instruments," Working Paper, Michigan State University.

KINAL, T. W. (1980): "The Existence of Moments of k-Class Estimators," *Econometrica*, 48(1), 241–249.

KLEIBERGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica*, 70(5), 1781–1803.

MARIANO, R. S., AND T. SAWA (1972): "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," *Journal of the American Statistical Association*, 67(337), 159–163.

MOREIRA, M. J. (2003): "A Conditional Likelihood Ratio Test for Structural Models," *Econometrica*, 71(4), 1027–1048.

NAGAR, A. L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27(4), 575–595.

NAGAR, A. L., AND A. ULLAH (1973): "Note on Approximate Skewness and Kurtosis of the Two Stage Least Squares Estimator," *Indian Economic Review*, 7/8, 70–80.

NELSON, C. R., AND R. STARTZ (1990): “Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator,” *Econometrica*, 58(4), 967–976.

NEWEY, W. K. (1990): “Semiparametric Efficiency Bounds,” *Journal of Applied Econometrics*, 5(2), 99–135.

PHILLIPS, P. C. B. (1984): “The Exact Distribution of LIML: I,” *International Economic Review*, 25(1), 249–261.

——— (1985): “The Exact Distribution of LIML: II,” *International Economic Review*, 26(1), 21–36.

——— (1986): “The Distribution of FIML in the Leading Case,” *International Economic Review*, 27(1), 239–243.

SAWA, T. (1969): “The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators,” *Journal of the American Statistical Association*, 64(327), 923–937.

STAIGER, D., AND J. H. STOCK (1997): “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, 65(3), 557–586.

STOCK, J. H., AND J. H. WRIGHT (2000): “GMM with Weak Identification,” *Econometrica*, 68(5), 1055–1096.

ULLAH, A. (2004): *Finite Sample Econometrics*. Oxford University Press, New York.

ULLAH, A., AND A. L. NAGAR (1974): “The Exact Mean of the Two-Stage Least Squares Estimator of the Structural Parameter in an Equation Having Three Endogenous Variables,” *Econometrica*, 42(4), 749–758.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press, New York.