Journal of Econometrics 211 (2019) 294–307

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

On the structure of IV estimands

Isaiah Andrews

Harvard Department of Economics, Littauer Center M18, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Available online 10 December 2018

JEL classification:
C13
C26

Keywords:
Instrumental variables
Misspecification

ABSTRACT

When the overidentifying restrictions of the constant-effect linear instrumental variables model fail, common IV estimators converge to different probability limits. I characterize the estimands of two stage least squares, two step GMM, and limited information maximum likelihood as functions of the single-instrument estimands from the just-identified IV regressions which consider each instrument separately. The limited information maximum likelihood estimand is found to be discontinuous on a set of dimension equal to the number of instruments minus one, and to equal the full parameter space on a set of dimension equal to the number of instruments minus two.

1. Introduction

A wide variety of estimators have been proposed for the constant-effect linear instrumental variables (IV) model, all of which converge to the true parameter value when the model is correctly specified and an instrument relevance condition holds. When the IV model is misspecified, on the other hand, common IV estimators typically converge to different probability limits.

The goal of this paper is to characterize the behavior of commonly-used estimators under model misspecification in linear IV models with a single endogenous regressor. In particular, the paper considers two-stage least squares (TSLS), two-step generalized method of moments (TSGMM), limited information maximum likelihood (LIML), and continuous updating generalized method of moments (CUGMM). The probability limits (estimands) of TSLS, TSGMM, and LIML are characterized as functions of the estimands in the just-identified models that use one instrument at a time, holding other features of the data generating process fixed. More limited results are derived for the CUGMM estimand.

As is well understood, the TSLS estimand is linear in the single-instrument estimands with linear combination weights summing to one. By contrast, the TSGMM estimand is generally nonlinear, though continuous, in the single-instrument estimands. More surprisingly the LIML estimand is highly nonlinear in the single-instrument estimands and is discontinuous on a set of dimension equal to the number of instruments minus one. If the controls include a constant, I show that the LIML estimand is discontinuous if and only if the vector of single-instrument estimands is such that (a) the TSLS estimand coincides with the ordinary least squares (OLS) estimand and (b) the R^2 from the reduced-form regression of the outcome on the instruments is greater than the R^2 from the first-stage regression of the endogenous regressor on the instruments. As the TSLS estimand approaches the OLS estimand from above the LIML estimand diverges to positive infinity, while as the TSLS estimand approaches the OLS estimand from below the LIML estimand diverges to negative infinity. Moreover, when the TSLS and OLS estimands coincide and the reduced-form R^2 is equal to the first-stage R^2, the population LIML objective function does not depend on the structural parameter value considered, so the minimizer is the full parameter space.

Analytical results for the CUGMM estimand are more elusive, but the level sets of this estimand (viewed as a function of the vector of single-instrument estimands) have a structure similar to those of LIML, and I find similar behavior for the LIML and CUGMM estimands in a calibration to data from Yogo (2004).

E-mail address: iandrews@fas.harvard.edu.

https://doi.org/10.1016/j.jeconom.2018.12.017
0304-4076/

The approach taken in this paper is distinct from that in the literature on heterogeneous treatment effects. A large literature originating with Imbens and Angrist (1994) characterizes the probability limits of IV estimators as combinations of heterogenous treatment effects under exogeneity and monotonicity assumptions. By contrast my approach, based on single-instrument IV estimands, is agnostic about the source and form of misspecification and so can accommodate heterogeneous treatment effects, invalidity of the instruments, or misspecification of the linear functional form. Further, my results apply directly to IV applications which are difficult to cast into the treatment effects framework, for example Yogo (2004). At the same time, however, my results only relate IV estimands to the single-instrument estimands and other statistical objects, rather than to the causal or structural parameters of interest. Hence, by remaining agnostic about the source of misspecification my approach accommodates models beyond the scope of the heterogeneous treatment effect literature but obtains correspondingly weaker results.

Two papers from the literature on heterogeneous treatment effects of particular relevance to my results are Kolesar (2013) and Mogstad et al. (2018). Kolesar (2013) shows that the LIML estimand can lie outside the convex hull of the individual treatment effects in a heterogeneous treatment effect model. Kolesar's results do not imply the discontinuity of LIML estimand but do suggest peculiar behavior for this quantity, which my results strongly confirm. Mogstad et al. (2018) derive expressions for a wide variety of estimands in terms of the potential outcomes in the treated and untreated states in a heterogenous treatment effect model with a binary treatment. Their results could be used to link the expressions in the present paper to causal effects in that setting, though further exploration of this possibility is left for future work. Other related work includes Hall and Inoue (2003), who examine the large-sample behavior of GMM estimators under misspecification, and Lee (2017), who proposes an asymptotic variance estimator for TSLS in models with heterogenous treatment effects.

In the next Section I formally introduce the IV model and define the IV estimands. Section 3 then presents analytical results on the structure of IV estimands, while Section 4 illustrates these results in a calibration to data from Yogo (2004). All proofs are given in the Appendix.

2. The linear IV model and estimands

Suppose we observe a sample of T observations (Y_t, X_t, Z_t) drawn from distribution F_T, where Y_t is an outcome variable, X_t is a potentially endogenous regressor, and Z_t is a $k \times 1$ vector of instrumental variables. Let us stack these observations into $T \times 1$ vectors Y and X with row t equal to Y_t and X_t respectively, and a $T \times k$ matrix Z with row t equal to Z_t'. Suppose the data obey the linear model

$$Y = X\beta + \varepsilon,$$

where β is the scalar parameter of interest. Conventional IV methods impose two further restrictions: the instrument relevance condition $E[Z_t X_t] \neq 0$, and the exclusion restriction $E[Z_t \varepsilon_t] = 0$. The model may accommodate additional exogenous regressors W_t as well, but for simplicity I will assume any exogenous variables have already been partialled out.[1]

A wide variety of estimation schemes have been proposed for linear IV. To accommodate different estimators in a unified framework while allowing the possibility of efficient estimation in models with heteroskedastic, serially correlated, or clustered data, here I treat linear IV as a special case of GMM. In particular, for $g_t(\beta) = (Y_t - X_t\beta) Z_t$ note that the usual IV identifying assumptions imply the moment restriction $E[g_t(\beta)] = E[Z_t \varepsilon_t] = 0$. For $g_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\beta) = \frac{1}{T} Z'(Y - X\beta)$ and $\widehat{W}(\beta)$ some (potentially parameter-dependent) symmetric positive-definite $k \times k$ weighting matrix, a general class of GMM estimators is given by

$$\widehat{\beta}_W = \arg\min_{\beta} \widehat{Q}_W(\beta) = \arg\min_{\beta} g_T(\beta)' \widehat{W}(\beta) g_T(\beta)$$

if this argmin exists and is unique. This paper focuses on four GMM estimators in particular: TSLS, LIML, TSGMM, and CUGMM.

The TSLS estimator is the simplest, and takes $\widehat{W}(\beta) = \left(\frac{1}{T}Z'Z\right)^{-1}$. This estimator is asymptotically efficient if the IV model is correctly specified and the errors ε_t are homoskedastic and independent across t, but may otherwise be inefficient. The LIML estimator is likewise efficient under correct specification and homoskedasticity but takes $\widehat{W}(\beta) = \widehat{\sigma}^{-2}(\beta)\left(\frac{1}{T}Z'Z\right)^{-1}$ for

$$\widehat{\sigma}^2(\beta) = \frac{1}{T}(Y - X\beta)' M_Z (Y - X\beta) = \widehat{\sigma}_Y^2 - 2\beta\widehat{\sigma}_{XY} + \beta^2\widehat{\sigma}_X^2.$$

where $M_Z = I_T - Z\left(Z'Z\right)^{-1} Z'$, $\widehat{\sigma}_Y^2 = \frac{1}{T}Y'M_Z Y$, $\widehat{\sigma}_{XY} = \frac{1}{T}Y'M_Z X$, and $\widehat{\sigma}_X^2 = \frac{1}{T}X'M_Z X$. For brevity of notation I write $\widehat{\beta}_{TSLS}$ and $\widehat{Q}_{TSLS}(\beta)$ for the TSLS estimator and objective, respectively, and do the same for the other estimators considered.

TSGMM attempts to improve efficiency by taking into account the asymptotic variance matrix of the moment vector. In particular, for a given first step estimator $\widehat{\beta}_1$ of β, which for concreteness I will take to be the TSLS estimator, TSGMM sets

[1] Thus, for $\tilde{Y}$, $\tilde{X}$, W, and $\tilde{Z}$ matrices collecting observations of the base variables $\tilde{Y}_t$, $\tilde{X}_t$, W_t, and $\tilde{Z}_t$ respectively, I define $Y = M_W\tilde{Y}$ and so on for $M_W = I_T - W\left(W'W\right)^{-1} W'$.

$\widehat{W}(\beta) = \widehat{\Sigma}\left(\widehat{\beta}_1\right)^{-1}$ for $\widehat{\Sigma}(\beta)$ an estimator of the asymptotic variance of $\sqrt{T}g_T(\beta)$. Such variance estimators can typically be decomposed as

$$\widehat{\Sigma}(\beta) = \widehat{\Gamma}_{11} - \beta\left(\widehat{\Gamma}_{12} + \widehat{\Gamma}_{12}'\right) + \beta^2\widehat{\Gamma}_{22}$$

for some $k \times k$ matrix-valued estimators $\left(\widehat{\Gamma}_{11}, \widehat{\Gamma}_{12}, \widehat{\Gamma}_{22}\right)$. I limit attention to variance estimators $\widehat{\Sigma}(\beta)$ of this form. Finally CUGMM bears the same relationship to TSGMM as TSLS does to LIML and takes $\widehat{W}(\beta) = \widehat{\Sigma}(\beta)^{-1}$.

2.1. Linear IV estimands

I am interested in estimands of IV estimators. As a starting point I assume that the terms that enter the GMM objective function all tend to well-defined limits.

Assumption 1. As the sample size T goes to infinity,

$$\left(\frac{1}{T}Z'Y, \frac{1}{T}Z'X\right) \to_p \left(E\left[Z_tY_t\right], E\left[Z_tX_t\right]\right).$$

Moreover,

$$\frac{1}{T}Z'Z \to_p E\left[Z_tZ_t'\right], \tag{1}$$

$$\begin{pmatrix} \widehat{\sigma}_Y^2 & \widehat{\sigma}_{XY} \\ \widehat{\sigma}_{XY} & \widehat{\sigma}_X^2 \end{pmatrix} \to_p \begin{pmatrix} \sigma_Y^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_X^2 \end{pmatrix} \tag{2}$$

$$= \begin{pmatrix} E\left[Y_t^2\right] - E\left[Z_tY_t\right]'E\left[Z_tZ_t'\right]^{-1}E\left[Z_tY_t\right] & E\left[X_tY_t\right] - E\left[Z_tY_t\right]'E\left[Z_tZ_t'\right]^{-1}E\left[Z_tX_t\right] \\ E\left[X_tY_t\right] - E\left[Z_tY_t\right]'E\left[Z_tZ_t'\right]^{-1}E\left[Z_tX_t\right] & E\left[X_t^2\right] - E\left[Z_tX_t\right]'E\left[Z_tZ_t'\right]^{-1}E\left[Z_tX_t\right] \end{pmatrix}$$

and

$$\begin{pmatrix} \widehat{\Gamma}_{11} & \widehat{\Gamma}_{12} \\ \widehat{\Gamma}_{12}' & \widehat{\Gamma}_{22} \end{pmatrix} \to_p \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}' & \Gamma_{22} \end{pmatrix} \tag{3}$$

where the limits in (1)–(3) are all full rank. For convenience I further assume that $\sigma_{XY} \neq 0$.

2.1.1. Parameter space

The IV estimands considered in this paper can be expressed as functions of

$$\psi = \left(E\left[Z_tY_t\right], E\left[Z_tX_t\right], E\left[Z_tZ_t'\right], \sigma_Y^2, \sigma_{XY}, \sigma_X^2, \Gamma_{11}, \Gamma_{12}, \Gamma_{22}\right). \tag{4}$$

It is natural to ask what set of values Ψ for ψ can arise in practice. To explore this question, it is helpful to consider the characterization of the IV model in terms of the reduced-form and first-stage regressions

$$Y_t = Z_t\delta + U_{Y,t}, \tag{5}$$

$$X_t = Z_t\pi + U_{X,t}, \tag{6}$$

where δ and π are the OLS regression coefficients of Y_t and X_t on Z_t, respectively.

To relate this model to ψ, note that $E\left[Z_tY_t\right] = E\left[Z_tZ_t'\right]\delta$ and $E\left[Z_tX_t\right] = E\left[Z_tZ_t'\right]\pi$. Assumption 1 implies that

$$\begin{pmatrix} \sigma_Y^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_X^2 \end{pmatrix} = E\begin{bmatrix} U_{Y,t}^2 & U_{Y,t}U_{X,t} \\ U_{Y,t}U_{X,t} & U_{X,t}^2 \end{bmatrix},$$

so the LIML estimator depends on the second-moment matrix of the residuals from the reduced-form and first-stage regressions. Analogously, depending on the estimator used $(\Gamma_{11}, \Gamma_{12}, \Gamma_{22})$ will typically correspond to the asymptotic variance of either $\left(\frac{1}{\sqrt{T}}Z'Y, \frac{1}{\sqrt{T}}Z'X\right)$ or $\left(\frac{1}{\sqrt{T}}Z'U_Y, \frac{1}{\sqrt{T}}Z'U_X\right)$.

The constant-effect linear instrumental variables model implies that $\delta = \pi\beta$, but this restriction may fail for a variety of reasons. For example, while we treat Z_t as an instrument some elements of Z_t may in fact be exogenous variables which should be included as controls (that is, these elements should be included in W_t). Alternatively, it may be that Z_t is a valid instrument but the true structural relationship is nonlinear, $Y_t = g(X_t) + \varepsilon_t$. In this case

$$\delta = E\left[Z_tZ_t'\right]^{-1}E\left[Z_tg(X_t)\right],$$

so δ will depend on the true functional form $g(X_t)$. Finally, we may have $\delta \neq \pi\beta$ due to heterogenous treatment effects. The exact relationship between the coefficients δ and the underlying heterogeneous effects in such cases is beyond the scope of this paper, but can be derived using the results of Mogstad et al. (2018).

When misspecification is due to misclassification of exogenous variables as instruments we can obtain any value for $\left(\delta, \pi, \sigma_Y^2, \sigma_{XY}, \sigma_X^2, E\left[Z_t Z_t'\right]\right)$ such that $E\left[Z_t Z_t'\right]$ and the variance matrix for $\left(U_{Y,t}, U_{X,t}\right)$ are positive semi-definite.[2] Many cases of functional form misspecification and treatment effect heterogeneity, by contrast, will impose additional restrictions on the possible values of ψ. The possible values of $(\Gamma_{11}, \Gamma_{12}, \Gamma_{22})$ are also restricted, where the exact restrictions depend on the setting and the estimator used. Most of the results of this paper focus on varying $E[Z_t Y_t]$, or equivalently δ, while holding the remaining elements of ψ constant, but in settings that imply additional restrictions one can limit attention to the corresponding parameter space Ψ.

2.1.2. Consistency and estimands

If we define $g(\beta; \psi) = E[Z_t Y_t] - E[Z_t X_t]\beta$ as the probability limit of $g_T(\beta)$ and let $W(\beta; \psi)$ denote the probability limit of $\widehat{W}(\beta)$,[3] then for the four GMM objective functions discussed above, Assumption 1 implies that

$$\widehat{Q}_W(\beta) \to_p Q_W(\beta; \psi) = g(\beta; \psi)' W(\beta; \psi) g(\beta; \psi)$$

for each fixed β. I define IV the estimand $\beta_W(\psi)$ to be the minimizer of $Q_W(\cdot; \psi)$.

Definition 1.

Define the IV estimand for weight matrix $\widehat{W}$ as

$$\beta_W(\psi) = \arg\min_{\beta \in \mathbb{R}} Q_W(\beta; \psi).$$

Note that $\beta_W(\psi)$ may be set-valued if $Q_W(\beta; \psi)$ has multiple minimizers. If $\lim_{\beta \to \pm\infty} Q_W(\beta; \psi) = \inf_\beta Q_W(\beta; \psi)$ then $\{-\infty, +\infty\} \subseteq \beta_W(\psi)$.

To connect $\beta_W(\psi)$ to the asymptotic behavior of $\widehat{\beta}_W$, note that if we define

$$\widehat{\psi} = \left(\frac{1}{T} Z'Y, \frac{1}{T} Z'X, \frac{1}{T} Z'Z, \widehat{\sigma}_Y^2, \widehat{\sigma}_{XY}, \widehat{\sigma}_X^2, \widehat{\Gamma}_{11}, \widehat{\Gamma}_{12}, \widehat{\Gamma}_{22}\right)$$

then for the estimators discussed here $\widehat{\beta}_W = \beta_W\left(\widehat{\psi}\right)$. Thus $\widehat{\beta}_W \to_p \beta_W(\psi)$ provided $\beta_W(\psi)$ is a singleton and ψ is a continuity point of $\beta_W(\cdot)$. Formally:

Lemma 1. *If $\beta_W(\psi)$ is a singleton and $\beta_W(\cdot)$ is continuous on an open neighborhood of ψ, Assumption 1 implies that $\widehat{\beta}_W = \beta_W\left(\widehat{\psi}\right) \to_p \beta_W(\psi)$.*

The focus of this paper is on characterizing the behavior of $\beta_W(\psi)$ as ψ varies. A well-understood pathology of IV estimators arises when the instrument relevance condition fails and the instruments are orthogonal to the endogenous regressor, $E[Z_t X_t] = 0$. For completeness I briefly discuss this case, but will rule it out for the remainder of the paper.

Irrelevant instruments. When $E[Z_t X_t] = 0$ the parameter β is not identified even if we assume the exclusion restriction holds. Correspondingly neither $Q_{TSLS}(\beta)$ nor $Q_{TSGMM}(\beta)$ depends on β,[4] with the result that $\beta_{TSLS}(\psi) = \beta_{TSGMM}(\psi) = \mathbb{R} \cup \{-\infty, +\infty\}$. Under the IV exclusion restriction, $E[Z_t X_t] = 0$ implies $E[Z_t Y_t] = 0$ since this is the only way the IV moment condition $E[Z_t \varepsilon_t] = 0$ can hold. In this case we can see that $\beta_{LIML}(\psi) = \beta_{CUGMM}(\psi) = \mathbb{R} \cup \{-\infty, +\infty\}$ as well. If, on the other hand, the IV model is misspecified so $E[Z_t X_t] = 0$ but $E[Z_t Y_t] \neq 0$, one can show that $\beta_{LIML}(\psi) = \beta_{CUGMM}(\psi) = \{-\infty, +\infty\}$. To avoid these pathologies, for the remainder of the paper I maintain the instrument relevance assumption. Further, for convenience I assume that each instrument is non-orthogonal to X_t.[5]

Assumption 2. For each element $Z_{i,t}$ of Z_t, $E\left[Z_{i,t} X_t\right] \neq 0$.

2.2. Single-instrument IV estimands

To study the behavior of IV estimands $\beta_W(\psi)$ under model misspecification it is helpful to have a concise representation for the degree and form of misspecification. One convenient such representation is based on the single-instrument IV estimands.

2 In particular, if Z_t is an exogenous control that we have misclassified as an instrument, for $\beta = 0$ we can obtain any values of $\left(\delta, \pi, \sigma_Y^2, \sigma_{XY}, \sigma_X^2\right)$ in (5) and (6).

3 If $E[Z_t X_t] = 0$ the weighting matrix $\widehat{W}$ used by TSGMM typically will not tend to a fixed probability limit, so it will not in general be the case that $\widehat{Q}_{TSGMM}(\beta)$ converges. Assumption 2 rules out this case.

4 Here I define $\beta_1(\psi)$ to be an arbitrary finite singleton so that $Q_{TSGMM}(\beta)$ is well-defined.

5 While this is stronger than $E[Z_t X_t] \neq 0$, given an instrument vector Z_t^* such that $E\left[Z_t^* X_t\right] \neq 0$ we can always define a rotation of the instruments $Z_t = OZ_t^*$ such that Assumption 2 holds.

Note that in a just-identified IV model with $k = 1$, provided the instrument relevance condition holds all of the estimands discussed above reduce to $\beta_W(\psi) = E[Z_t Y_t]/E[Z_t X_t]$. Even when $k > 1$ we can obtain a just-identified system by restricting attention to the ith instrument $Z_{i,t}$, yielding IV estimand

$$b_i = \frac{E\left[Z_{i,t} Y_t\right]}{E\left[Z_{i,t} X_t\right]}.$$

The IV exclusion restriction

$$E[Z_t Y_t] - \beta E[Z_t X_t] = 0$$

holds for some value β if and only if $b_i = \beta$ for all i. Hence, the IV model's over-identifying restrictions hold if and only if $b_i = b_j$ for all i and j. Denote the set of $b = (b_1, \ldots, b_k)'$ such that the IV over-identifying restrictions hold by

$$\mathcal{B} = \left\{b : b_i = b_j \text{ for all } i \text{ and } j\right\}. \tag{7}$$

Note that since $b = D(E[Z_t X_t])^{-1} E[Z_t Y_t]$ for $D(V)$ which maps the $k \times 1$ vector V to a $k \times k$ diagonal matrix with the elements of V along the diagonal and zeros elsewhere, we can write

$$Q_W(\beta; \psi) = (b - \iota\beta)' \Omega(\beta)(b - \iota\beta)$$

for $\Omega(\beta) = D(E[Z_t X_t])' W(\beta) D(E[Z_t X_t])$ and ι the $k \times 1$ vector of ones. Since correct specification of the IV model restricts only the vector b, to understand the effect of misspecification on IV estimands I will consider behavior as b varies, holding the other elements of ψ (i.e. $E[Z_t X_t], E\left[Z_t Z_t'\right], \sigma_Y^2, \sigma_{XY}, \sigma_X^2, \Gamma_{11}, \Gamma_{12}, \Gamma_{22}$) fixed. See Section 2.1.1 for further discussion. To emphasize this focus on the behavior of Q_W and β_W in b, I abbreviate $Q_W(\beta; \psi) = Q_W(\beta; b)$ and $\beta_W(\psi) = \beta_W(b)$ for the remainder of the paper.

3. The structure of IV estimands

As noted in the previous section, all IV estimands coincide in just-identified models, provided the instrument relevance condition holds. In over-identified models, by contrast, each instrument implies a corresponding IV estimand and the question is how to combine the single-instrument estimands into an overall estimate. The different IV estimators discussed above imply different answers to this question, and the goal of this section is to characterize the behavior of the IV estimands β_W as functions of the single-instrument estimands b.

3.1. TSLS

The two stage least squares objective $Q_{TSLS}(\beta)$ is quadratic in β, so under Assumption 2 we can solve analytically for

$$\beta_{TSLS}(b) = \frac{E[Z_t X_t]' E\left[Z_t Z_t'\right]^{-1} E[Z_t Y_t]}{E[Z_t X_t]' E\left[Z_t Z_t'\right]^{-1} E[Z_t X_t]} = \sum_{i=1}^{k} w_i b_i$$

where

$$w_i = \frac{E[Z_t X_t]' E\left[Z_t Z_t'\right]^{-1} e_i e_i' E[Z_t X_t]}{E[Z_t X_t]' E\left[Z_t Z_t'\right]^{-1} E[Z_t X_t]}$$

for e_i the ith standard basis vector. Note that $\sum_i w_i = 1$ so the TSLS estimand is a linear combination of the single-instrument estimands with weights summing to one. The weights w_i are not necessarily positive, however, so we cannot in general interpret $\beta_{TSLS}(b)$ as a weighted average of the single-instrument estimands.[6]

3.2. TSGMM

The TSGMM objective is again quadratic in β, so we can solve for

$$\beta_{TSGMM}(b) = \sum_{i=1}^{k} w_i(b) b_i$$

where

$$w_i(b) = \frac{E[Z_t X_t]' \Sigma^{-1}(\beta_{TSLS}(b)) e_i e_i' E[Z_t X_t]}{E[Z_t X_t]' \Sigma^{-1}(\beta_{TSLS}(b)) E[Z_t X_t]}.$$

[6] In models with heterogenous treatment effects, Angrist and Imbens (1995) and Kolesar (2013) give results which characterize the TSLS estimand as a weighted average of particular causal effects under monotonicity conditions. Their results do not imply, however, that the weights w_i above are positive.

Thus for a given weight matrix the TSGMM estimand is a linear combination of the single-instrument estimands with weights which sum to one, where the weights themselves now depend on b through the first-stage estimand. The TSGMM estimand will generally be nonlinear in the single-instrument estimands, except in the homoskedastic case where it coincides with TSLS. Assumption 1 implies that $\Sigma(\beta)$ is everywhere full rank, and thus that $\beta_{TSGMM}(b)$ is a continuous (and, in fact, differentiable) function of b.

3.3. LIML

Matters become more interesting when we consider LIML. The well-known characterization of LIML as a k-class estimator (see e.g. Hausman, 1983) implies the following,

Lemma 2. *Let*

$$\Lambda(b) = \begin{pmatrix} \lambda_1(b) & \lambda_2(b) \\ \lambda_2(b) & \lambda_3 \end{pmatrix} =$$
$$\begin{pmatrix} E[Z_tY_t]' E\left[Z_tZ_t'\right]^{-1} E[Z_tY_t] & E[Z_tY_t]' E\left[Z_tZ_t'\right]^{-1} E[Z_tX_t] \\ E[Z_tY_t]' E\left[Z_tZ_t'\right]^{-1} E[Z_tX_t] & E[Z_tX_t]' E\left[Z_tZ_t'\right]^{-1} E[Z_tX_t] \end{pmatrix} = \begin{pmatrix} b'\Xi b & b'\Xi\iota \\ \iota'\Xi b & \iota'\Xi\iota \end{pmatrix}$$

for $\Xi = D(E[Z_tX_t]) E\left[Z_tZ_t'\right]^{-1} D(E[Z_tX_t])$. *For* $\varphi(b)$ *the largest root of the quadratic equation*

$$\left(\lambda_1(b) + \phi\sigma_Y^2\right)\left(\lambda_3 + \phi\sigma_X^2\right) - \left(\lambda_2(b) + \phi\sigma_{XY}\right)^2 = 0,$$

the LIML estimand $\beta_{LIML}(b)$ *is given by*

$$\beta_{LIML}(b) = \frac{\lambda_2(b) + \varphi(b)\sigma_{XY}}{\lambda_3 + \varphi(b)\sigma_X^2} \tag{8}$$

whenever the denominator is nonzero.

The only terms in $\beta_{LIML}(b)$ which depend on b are λ_2 and φ, both of which change continuously in b. Thus, over most of the parameter space β_{LIML} will be continuous in b. As might be expected, however, β_{LIML} behaves strangely when the denominator in (8) crosses zero.

To formally discuss this issue, define the irregular set

$$\mathcal{I} = \left\{b : \lambda_3 + \varphi(b)\sigma_X^2 = 0\right\} \tag{9}$$

as the set of values b such that the denominator in (8) is zero.

Proposition 1. *For the irregular set* $\mathcal{I}$ *defined in (9):*

1. $b \in \mathcal{I}$ *if and only if*
$$\frac{\lambda_1(b)}{\sigma_Y^2} \geq \frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}.$$
2. *For* $\mathcal{B}$ *the set of single-instrument estimands* b *such that the IV over-identifying restrictions hold (defined in (7)),* $\mathcal{B} \cap \mathcal{I} = \emptyset$.
3. *For* $b \in \mathcal{I}$,
$$\beta_{LIML}(b) = \begin{cases} \{-\infty, +\infty\} & \text{if } \frac{\lambda_1(b)}{\sigma_Y^2} > \frac{\lambda_2(b)}{\sigma_{XY}}, \\ \mathbb{R} \cup \{-\infty, +\infty\} & \text{if } \frac{\lambda_1(b)}{\sigma_Y^2} = \frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}. \end{cases}$$

Note that $\lambda_2(b)$ and $\lambda_1(b)$ are linear and quadratic in b, respectively. Consequently $\mathcal{I}$ consists of the set of points on the $(k-1)$-dimensional hyperplane $\left\{b : \lambda_2(b) = \sigma_{XY}\frac{\lambda_3}{\sigma_X^2}\right\}$ lying weakly outside the ellipsoid $\left\{b : \lambda_1(b) = \sigma_Y^2\frac{\lambda_3}{\sigma_X^2}\right\}$ centered at the origin. The exclusion of the interior of this ellipsoid ensures that the irregular set does not intersect the parameter space $\mathcal{B}$ for the correctly specified IV model, and so reconciles the peculiar behavior of the LIML estimand on $\mathcal{I}$ with its well-understood behavior under correct specification. Note that on the set

$$\mathcal{S} = \left\{b : \frac{\lambda_1(b)}{\sigma_Y^2} = \frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}\right\},$$

$\beta_{LIML}(b)$ is the full extended real line. Hence even though I consider a fixed non-zero value of $E[Z_tX_t]$, and thus the instrument relevance condition holds, the minimizer of the LIML population objective Q_{LIML} is not uniquely defined for $b \in \mathcal{S}$.

Interpreting $\mathcal{I}$ and $\mathcal{S}$. To better understand the irregular set $\mathcal{I}$, note that $\frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$ if and only if

$$\frac{b'\Xi\iota}{\iota'\Xi\iota} = \frac{E\left[Z_tX_t\right]' E\left[Z_tZ_t'\right]^{-1} E\left[Z_tY_t\right]}{E\left[Z_tX_t\right]' E\left[Z_tZ_t'\right]^{-1} E\left[Z_tX_t\right]} = \frac{\sigma_{XY}}{\sigma_X^2}. \tag{10}$$

From the results of Section 3.1, however, the left hand side is equal to the TSLS estimand. Using the definitions of σ_{XY} and σ_X^2, however, the OLS estimand from regressing Y_t on X_t can be written as

$$\beta_{OLS}(b) = \frac{E\left[X_tY_t\right]}{E\left[X_t^2\right]} = \frac{b'\Xi\iota + \sigma_{XY}}{\iota'\Xi\iota + \sigma_X^2}.$$

Thus, we see that (10) is equivalent to

$$\beta_{TSLS}(b) = \frac{b'\Xi\iota}{\iota'\Xi\iota} = \frac{b'\Xi\iota + \sigma_{XY}}{\iota'\Xi\iota + \sigma_X^2} = \beta_{OLS}(b).$$

Hence, $\frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$ if and only if the TSLS and OLS estimands coincide.

Turning next to the condition that $\frac{\lambda_1(b)}{\sigma_Y^2} \geq \frac{\lambda_2(b)}{\sigma_{XY}}$, note that we can re-write this inequality as

$$\frac{b'\Xi b}{\sigma_Y^2} \geq \frac{\iota'\Xi\iota}{\sigma_X^2},$$

which, using the definitions of σ_Y^2 and σ_X^2, is equivalent to

$$\frac{E\left[Z_tY_t\right]' E\left[Z_tZ_t'\right]^{-1} E\left[Z_tY_t\right]}{E\left[Y_t^2\right]} \geq \frac{E\left[Z_tX_t\right]' E\left[Z_tZ_t'\right]^{-1} E\left[Z_tX_t\right]}{E\left[X_t^2\right]}.$$

If the vector of controls W_t includes a constant, however, (X_t, Y_t, Z_t) all have mean zero, and the left hand side is the R^2 from the reduced-form regression of Y_t on Z_t. Likewise, the right hand side is the R^2 from the first-stage regression of X_t on Z_t. Thus, if the vector of controls includes a constant, $\frac{\lambda_1(b)}{\sigma_Y^2} \geq \frac{\lambda_2(b)}{\sigma_{XY}}$ if and only if the R^2 for the reduced-form exceeds that for the first-stage.

Summing up, if the vector of controls W_t includes a constant we see that $b \in \mathcal{I}$ if and only if (i) $\beta_{TSLS}(b) = \beta_{OLS}(b)$ and (ii) the reduced-form R^2 exceeds the first-stage R^2. Likewise, $b \in \mathcal{S}$ if and only if $\beta_{TSLS}(b) = \beta_{OLS}(b)$ and the reduced-form and first-stage have the same R^2.

Behavior outside of $\mathcal{I}$. It is also interesting to understand the behavior of β_{LIML} when $b \notin \mathcal{I}$. To do so, it is helpful to consider the structure of the first-order conditions $\frac{\partial}{\partial\beta}Q_{LIML}(\beta; b) = 0$. We know that if $\beta_{LIML}(b) = \beta \in \mathbb{R}$ we must have $\frac{\partial}{\partial\beta}Q_{LIML}(\beta; b) = 0$, so if we define

$$\mathcal{F}_{LIML}(\beta) = \left\{b : \frac{\partial}{\partial\beta}Q_{LIML}(\beta; b) = 0\right\}$$

to be the set of values b such that the LIML first order conditions are satisfied for a given β, then the set of b such that $\beta_{LIML}(b) = \beta$ must be a subset of $\mathcal{F}_{LIML}(\beta)$,

$$\{b : \beta_{LIML}(b) = \beta\} \subseteq \mathcal{F}_{LIML}(\beta). \tag{11}$$

The next lemma characterizes $\mathcal{F}_{LIML}(\beta)$.

Proposition 2. *$\mathcal{F}_{LIML}(\beta)$ is an ellipsoid,*

$$\mathcal{F}_{LIML}(\beta) = \left\{b : (b - A(\beta))' \,\Xi\, (b - A(\beta)) = C(\beta)\right\}$$

where

$$A(\beta) = \frac{\beta^2\sigma_X^2 - \sigma_Y^2}{-2\sigma_{XY} + 2\beta\sigma_X^2}\iota$$

and

$$C(\beta) = \left(\sigma_Y^2 - 2\beta\sigma_{XY}^2 + \beta^2\sigma_X^2\right)^2 \iota'\Xi\iota.$$

Thus, the set of values b satisfying the LIML first order condition is a ellipsoid with center $A(\beta)$. Consequently, by (11) the level sets of $\beta_{LIML}(b)$ are subsets of ellipsoids. One can confirm that $\mathcal{S} \subset \mathcal{F}_{LIML}(\beta)$ for all β, as must be the case given (11) together with the result in part (3) of Proposition 1. Moreover, one can show that any $b \in \mathcal{I} \setminus \mathcal{S}$ must lie outside of $\mathcal{F}_{LIML}(\beta)$ in the sense that

$$(b - A(\beta))' \, \Xi \, (b - A(\beta)) > C(\beta) .$$

Indeed, for any sequence of single-instrument IV estimands approaching a point in $\mathcal{I} \setminus \mathcal{S}$, the LIML estimand $\beta_{LIML}(b)$ diverges.

Corollary 1. *For any sequence $b_n \to b \in \mathcal{I} \setminus \mathcal{S}$ such that $b_n \notin \mathcal{I}$ for all n:*

1. $\lim_{n\to\infty} |\beta_{LIML}(b_n)| \to +\infty$
2. *If $\lambda_2(b_n) > \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3$ for all n then $\beta_{LIML}(b_n) \to +\infty$, while if $\lambda_2(b_n) < \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3$ for all n then $\beta_{LIML}(b_n) \to -\infty$.*

To interpret this result, note that $\lambda_2(b) > \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3$ implies $\beta_{2SLS}(b) > \beta_{OLS}(b)$, while the reverse holds for $\lambda_2(b) < \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3$. Thus, when the TSLS estimand approaches the OLS estimand from above the LIML estimand diverges to $+\infty$, while when the TSLS estimand approaches from below the LIML estimand diverges to $-\infty$.

3.4. CUGMM

There is no known closed-form expression for the continuous updating GMM estimator $\hat{\beta}_{CUGMM}$ in non-homoskedastic IV models, and the behavior of $\beta_{CUGMM}(b)$ is correspondingly harder to characterize. Nonetheless, if one considers the set on which the CUGMM first order conditions are satisfied for a given β

$$\mathcal{F}_{CUGMM}(\beta) = \left\{ b : \frac{\partial}{\partial \beta} Q(\beta; b) = 0 \right\}$$

one can show that these sets are again ellipsoids. Thus, since for $\beta \in \mathbb{R}$,

$$\{b : \beta_{CUGMM}(b) = \beta\} \subseteq \mathcal{F}_{CUGMM}(\beta),$$

the level sets of β_{CUGMM} are again subsets of ellipsoids.

Proposition 3. *$\mathcal{F}_{CUGMM}(\beta)$ is an ellipsoid*

$$\mathcal{F}_{CUGMM}(\beta) = \left\{ b : \left(b - A^*(\beta)\right)' B^*(\beta) \left(b - A^*(\beta)\right) = C^*(\beta) \right\}$$

where

$$A^*(\beta) = \left(I_k\beta - \Omega^{-1}(\beta)\left(-\widetilde{\Gamma}_{12} - \widetilde{\Gamma}_{12}' + 2\beta\widetilde{\Gamma}_{22}\right)^{-1} \right)\iota$$

$$B^*(\beta) = \Omega(\beta)\left(-\widetilde{\Gamma}_{12} - \widetilde{\Gamma}_{12}' + 2\beta\widetilde{\Gamma}_{22}\right)\Omega(\beta)$$

and

$$C^*(\beta) = \iota'\left(-\widetilde{\Gamma}_{12} - \widetilde{\Gamma}_{12}' + 2\beta\widetilde{\Gamma}_{22}\right)^{-1}\iota$$

for $\widetilde{\Gamma}_{ij} = D(E[ZX_t])^{-1}\Gamma_{ij}D(E[ZX_t])^{-1}$.

As Proposition 3 highlights, the CUGMM estimand has a structure similar to the LIML estimand, in that the contours of the CUGMM estimand are again subsets of ellipsoids. Unlike in the case of LIML, however, the matrix $B^*(\beta)$ which defines the "shape" of these ellipsoids now depends on β. Moreover, there is not in general a point where all of the sets $\mathcal{F}_{CUGMM}(\beta)$ intersect, and thus there does not in general exist a value b such that $\beta_{CUGMM}(b)$ is equal to the extended real line. Nonetheless, in the next Section 1 find that the CUGMM estimand $\beta_{CUGMM}(b)$ exhibits behavior similar to the LIML estimand $\beta_{LIML}(b)$ in an example calibrated to data.

4. IV estimands in an example

To illustrate the analytic results above, Figs. 1–4 plot the contours of the IV estimands β_W as functions of single-instrument estimands b in a calibration based on Yogo (2004). Yogo studies the effect of weak instruments on estimation of the elasticity of intertemporal substitution using a linear Euler equation model and data from a number of countries. Here I calibrate all elements of ψ other than $E[Z_tY_t]$ to values estimated from the quarterly US data series used by Yogo, which covers the period from the third quarter of 1947 to the last quarter of 1998. Yogo considers a number of specifications, and here I take the outcome variable Y_t to be real consumption growth and the endogenous regressor X_t to be the real interest rate. Yogo

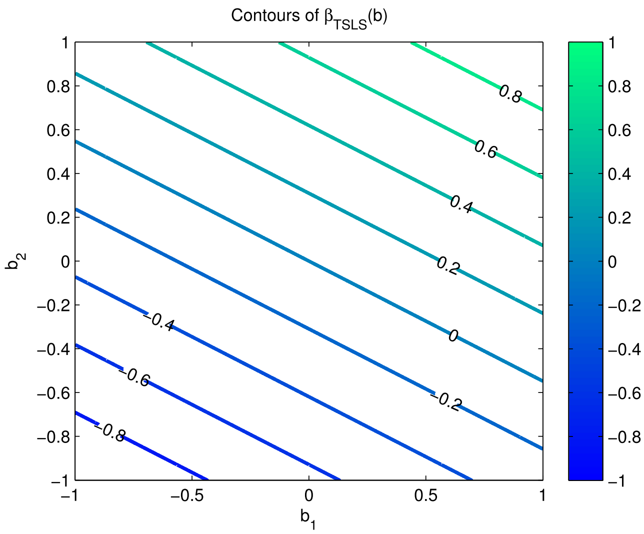

Fig. 1. Contours of two-stage least squares estimand $\beta_{TSLS}(b)$ as a function of single-instrument estimands b in calibration to US quarterly data from Yogo (2004).

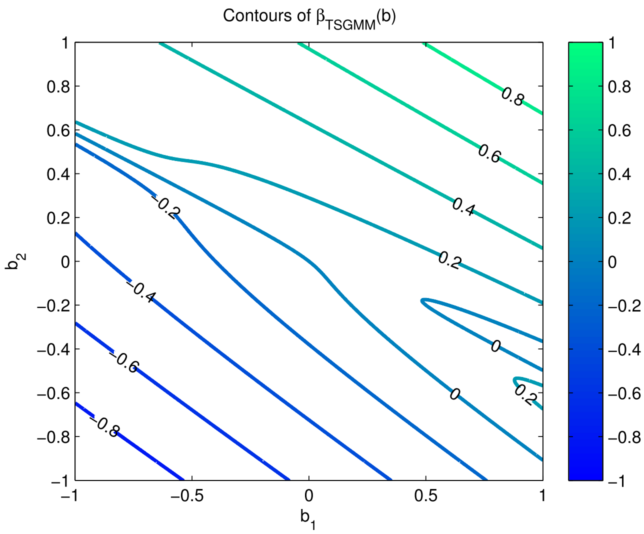

Fig. 2. Contours of two-step GMM estimand $\beta_{TSGMM}(b)$ as a function of single-instrument estimands b in calibration to US quarterly data from Yogo (2004).

finds that identification-robust Anderson–Rubin confidence sets for the coefficient on the real interest rate are empty in this dataset, suggesting model misspecification.[7] See Yogo (2004) for further discussion of the data.

Yogo's analysis uses four instruments, namely the two-period lags of consumption growth, the dividend–price ratio, the nominal interest rate, and inflation. In order to plot the contours of $\beta_W(b)$ here I restrict attention to two instruments, specifically lagged consumption growth and the dividend–price ratio, which I select because they yield easy-to-read plots. Note that while Yogo's analysis is motivated by a concern with weak instruments, here I fix $E[Z_t X_t]$ at its (non-zero) estimate from the data so weak instruments do not drive the results.

Figs. 1–4 bear out the analytical results of the previous section. In particular, we see that the TSLS estimand $\beta_{TSLS}(b)$ is linear in b, while the TSGMM estimand is continuous in b but nonlinear. The contours of the LIML estimand are subsets of ellipsoids, and all contours intersect at two points. The CUGMM estimand is in many ways similar to the LIML estimand but its behavior is somewhat more irregular, particularly for small β.

[7] The Anderson–Rubin confidence set considered by Yogo assumes the data are homoskedastic, and Yogo finds non-empty confidence sets when instead considering the S statistic of Stock and Wright (2000) which relaxes this homoskedasticity assumption. S statistic confidence sets are, however, found to be empty in a linear GMM specification which treats both real interest rates and real equity returns as endogenous regressors.

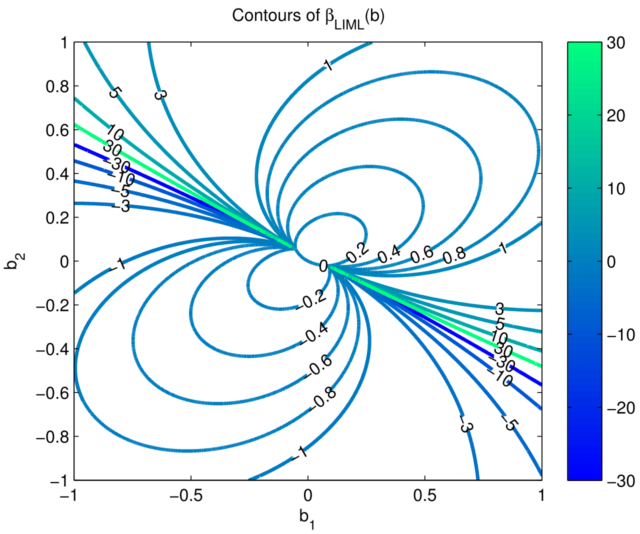

Fig. 3. Contours of limited information maximum likelihood estimand β_{LIML} (b) as a function of single-instrument estimands b in calibration to US quarterly data from Yogo (2004).

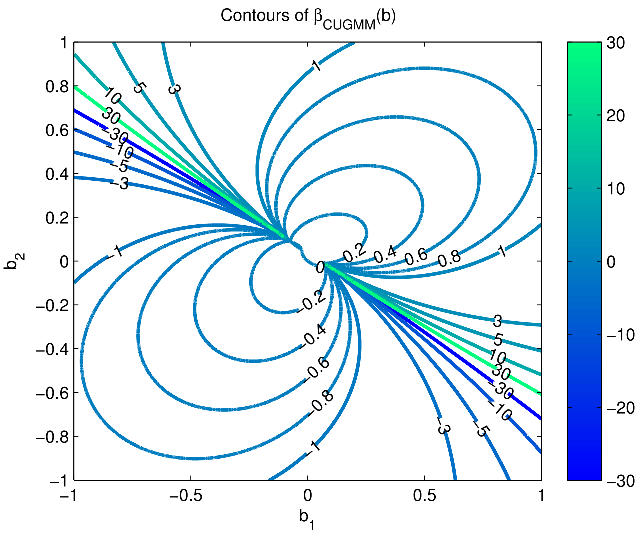

Fig. 4. Contours of continuously updating GMM estimand β_{CUGMM} (b) as a function of single-instrument estimands b in calibration to US quarterly data from Yogo (2004).

5. Conclusion

When the over-identifying restrictions of the classical IV model fail, common IV estimators converge to distinct probability limits. Characterizing these limits as a function of the single-instrument IV estimands, I find that the LIML estimand is discontinuous and, further, is sometimes equal to the full parameter space. If the set of controls includes a constant, these issues arise when the OLS and TSLS estimands are equal and the reduced-form R^2 is weakly larger than the first-stage R^2. While complete analytical results for CUGMM are more elusive, the contours of the CUGMM estimator resemble those of LIML, and the two estimands display similar behavior in a calibration to data from Yogo (2004).

These results do not necessarily imply that we ought to favor one estimator over another: when single-instrument IV estimands differ, the choice among different IV estimators amounts to a choice of how best to summarize these disparate estimands. Moreover, one might argue that the extreme behavior observed for the LIML estimand is in part an artifact how we have parameterized the model. If one instead considers the circular parametrization of the IV model as in Chamberlain (2007), for example, then the LIML estimand is continuous on $\mathcal{I} \setminus \mathcal{S}$. Even with this reparameterization, however, the LIML estimand is not uniquely defined on $\mathcal{S}$. Overall, the highly nonlinear and discontinuous behavior of the LIML estimand suggests that caution is warranted when interpreting LIML estimates in misspecified models.

Acknowledgments

I am grateful to Jushan Bai, Anna Mikusheva, Whitney Newey, Christoph Rothe, Miikka Rokkanen, Jim Stock, participants in the Fall 2015 Conference in Honor of Jerry Hausman, and two anonymous referees for helpful comments. Support from the Silverman (1978) Family Career Development Chair at MIT and from the National Science Foundation under grant number 1654234 is gratefully acknowledged.

Appendix

Proof of Lemma 1. Follows immediately from the Continuous Mapping Theorem, see e.g. Theorem 2.3 of van der Vaart (2000). □

Proof of Lemma 2. One may note that the same argument which allows us to derive the expression for $\widehat{\beta}_{LIML}$ from the objective function $\widehat{Q}_{LIML}(\cdot)$ likewise allows us to derive the expression for $\beta_{LIML}(\psi)$ from the objective function $Q(\cdot;\psi)$. For completeness, however, I provide a direct proof.

One can express the LIML estimator as

$$\widehat{\beta}_{LIML} = \frac{X'Y - \widehat{k}_{LIML}X'M_ZY}{X'X - \widehat{k}_{LIML}X'M_ZX} = \frac{X'P_ZY + \left(1-\widehat{k}_{LIML}\right)X'M_ZY}{X'P_ZX + \left(1-\widehat{k}_{LIML}\right)X'M_ZX}$$

for $\widehat{k}_{LIML}$ the smallest root of

$$\det\left(\begin{pmatrix} Y'Y & Y'X \\ Y'X & X'X \end{pmatrix} - k\begin{pmatrix} Y'M_ZY & Y'M_ZX \\ Y'M_ZX & X'M_ZX \end{pmatrix}\right) = 0$$

or, equivalently, $\widehat{\varphi} = \left(1-\widehat{k}_{LIML}\right)$ the largest root of

$$\det\left(\begin{pmatrix} Y'P_ZY & Y'P_ZX \\ Y'P_ZX & X'P_ZX \end{pmatrix} + \phi\begin{pmatrix} Y'M_ZY & Y'M_ZX \\ Y'M_ZX & X'M_ZX \end{pmatrix}\right) =$$
$$\left(Y'P_ZY + \phi T\widehat{\sigma}_Y^2\right)\left(X'P_ZX + \phi T\widehat{\sigma}_X^2\right) - \left(Y'P_ZX + \phi T\widehat{\sigma}_{XY}\right)^2 = 0.$$

Assumption 1 implies that

$$\frac{1}{T}\begin{pmatrix} Y'P_ZY & Y'P_ZX \\ Y'P_ZX & X'P_ZX \end{pmatrix} \to_p \Lambda(b)$$

for $\Lambda(b)$ as defined in the text, while

$$\frac{1}{T}\begin{pmatrix} Y'M_ZY & Y'M_ZX \\ Y'M_ZX & X'M_ZX \end{pmatrix} \to_p \begin{pmatrix} \sigma_Y^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_X^2 \end{pmatrix}.$$

Thus, for all ϕ

$$\frac{1}{T^2}\left(Y'P_ZY + \phi T\widehat{\sigma}_Y^2\right)\left(X'P_ZX + \phi T\widehat{\sigma}_X^2\right) - \frac{1}{T^2}\left(Y'P_ZX + \phi T\widehat{\sigma}_{XY}\right)^2 \to_p$$
$$\left(\lambda_1 + \phi\sigma_Y^2\right)\left(\lambda_3 + \phi\sigma_X^2\right) - \left(\lambda_2 + \phi\sigma_{XY}\right)^2. \tag{12}$$

Since the largest root of a quadratic equation (when it exists) is a continuous function of the coefficients, and the structure of the problem implies that such a root always exists in the present setting, the Continuous Mapping Theorem implies that $\widehat{\varphi} \to_p \varphi$ for φ the largest root of (12). Thus, again applying the Continuous Mapping Theorem,

$$\widehat{\beta}_{LIML} \to_p \beta_{LIML}(b) = \frac{\lambda_2(b) + \varphi(b)\sigma_{XY}}{\lambda_3 + \varphi(b)\sigma_X^2}$$

provided the denominator $\lambda_3 + \varphi(b)\sigma_X^2$ is nonzero. □

Proof of Proposition 1. I first prove part (1). Suppose that $\lambda_3 + \varphi(b)\sigma_X^2 = 0$. Since $\varphi(b)$ is the largest root of (12),

$$\left(\lambda_1(b) + \varphi(b)\sigma_Y^2\right)\left(\lambda_3 + \varphi(b)\sigma_X^2\right) - \left(\lambda_2(b) + \varphi(b)\sigma_{XY}\right)^2 = 0.$$

It must therefore be the case that $\lambda_2(b) + \varphi(b)\sigma_{XY} = 0$ as well, implying that $\frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$. Note, further, that (12) is quadratic in ϕ and tends to infinity as $\phi \to \pm\infty$ (since positive-definiteness of the limit in (2) implies that $\sigma_{XY}^2 < \sigma_X^2\sigma_Y^2$), and thus that a necessary and sufficient condition for a root ϕ^* of (12) to be the largest root (or one of the largest roots if the roots are equal) is that the derivative of (12) at ϕ^* is non-negative. Thus, since the derivative of (12) with respect to ϕ is

$$\sigma_Y^2\left(\lambda_3 + \phi\sigma_X^2\right) + \sigma_X^2\left(\lambda_1(b) + \phi\sigma_Y^2\right) - 2\sigma_{XY}\left(\lambda_2(b) + \phi\sigma_{XY}\right),$$

a value ϕ^* with

$$\lambda_3 + \phi^* \sigma_X^2 = \lambda_2 (b) + \phi^* \sigma_{XY} = 0$$

is the largest root (or one of the largest roots, if the roots are equal) if and only if

$$\lambda_1 (b) + \phi^* \sigma_Y^2 = \lambda_1 (b) - \frac{\lambda_3}{\sigma_X^2} \sigma_Y^2 \geq 0.$$

Hence the necessary and sufficient condition for $\lambda_3 + \varphi (b) \sigma_X^2 = 0$ is that $\frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$ and $\frac{\lambda_1(b)}{\sigma_Y^2} \geq \frac{\lambda_3}{\sigma_X^2}$.

I next prove part (2). For $b \in \mathcal{B}$, $b_i = \tilde{\beta}$ for all i which implies that $\lambda_1 (b) = \tilde{\beta}^2 \lambda_3$ and $\lambda_2 (b) = \tilde{\beta} \lambda_3$ for all i, so $\frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$ for $b \in \mathcal{B}$ if and only if $\tilde{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} = \rho \frac{\sigma_Y}{\sigma_X}$ (for $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$). On the other hand $\frac{\lambda_1(b)}{\sigma_Y^2} \geq \frac{\lambda_2(b)}{\sigma_{XY}}$ for $b \in \mathcal{B}$ requires that $\tilde{\beta} \geq \frac{\sigma_Y}{\rho \sigma_X}$ if $\rho > 0$, and that $\tilde{\beta} \leq \frac{\sigma_Y}{\rho \sigma_X}$ if $\rho < 0$. Since I have assumed $\rho \neq 0$, these requirements can be satisfied only if $\rho = 1$ or $\rho = -1$ which would imply that the right hand side in (2) has reduced rank and so is ruled out by Assumption 1.

Finally, I prove part (3). Note that

$$Q_{LIML} (\beta; b) = \frac{\lambda_1 (b) - 2\beta \lambda_2 (b) + \beta^2 \lambda_3}{\sigma_Y^2 - 2\beta \sigma_{XY} + \beta^2 \sigma_X^2}.$$

Taking the first order condition with respect to β yields that at any local minimum $\tilde{\beta}$ of Q_{LIML},

$$\frac{-2\lambda_2 (b) + 2\lambda_3 \tilde{\beta}}{\sigma_Y^2 - 2\tilde{\beta} \sigma_{XY} + \tilde{\beta}^2 \sigma_X^2} - \left(-2\sigma_{XY} + 2\sigma_X^2 \tilde{\beta}\right) \frac{\lambda_1 (b) - 2\tilde{\beta} \lambda_2 (b) + \tilde{\beta}^2 \lambda_3}{\left(\sigma_Y^2 - 2\tilde{\beta} \sigma_{XY} + \tilde{\beta}^2 \sigma_X^2\right)^2} = 0$$

which, with some algebra, can be shown to hold if and only if

$$\sigma_Y^2 \left(-2\lambda_2 (b) + 2\lambda_3 \tilde{\beta}\right) - \lambda_1 (b) \left(-2\sigma_{XY} + 2\sigma_X^2 \tilde{\beta}\right) + 2\tilde{\beta}^2 \lambda_2 (b) \sigma_X^2 - 2\tilde{\beta}^2 \lambda_3 \sigma_{XY} = 0. \tag{13}$$

For $b \in \mathcal{I}$, $\frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$ so (13) becomes

$$\sigma_Y^2 \frac{\lambda_3}{\sigma_X^2} \left(-2\sigma_{XY} + 2\sigma_X^2 \tilde{\beta}\right) - \lambda_1 (b) \left(-2\sigma_{XY} + 2\sigma_X^2 \tilde{\beta}\right)$$
$$= \left(\sigma_Y^2 \frac{\lambda_3}{\sigma_X^2} - \lambda_1 (b)\right) \left(-2\sigma_{XY} + 2\sigma_X^2 \tilde{\beta}\right) = 0.$$

If $\sigma_Y^2 \frac{\lambda_3}{\sigma_X^2} - \lambda_1 (b) = 0$ (so $\frac{\lambda_1(b)}{\sigma_Y^2} = \frac{\lambda_3}{\sigma_X^2}$), then this condition holds for all β and $Q_{LIML} (\beta; b)$ does not depend on β, implying that $\beta_{LIML} (b) = \mathbb{R} \cup \{-\infty, \infty\}$. If, on the other hand, $\sigma_Y^2 \frac{\lambda_3}{\sigma_X^2} - \lambda_1 (b) \neq 0$ then the unique solution to (13) is $\tilde{\beta} = \frac{\sigma_{XY}}{\sigma_X^2}$. Plugging this back into the LIML objective gives

$$Q_{LIML} (\beta; b) = \frac{\lambda_1 (b) - 2\frac{\sigma_{XY}}{\sigma_X^2} \lambda_2 (b) + \frac{\sigma_{XY}^2}{\sigma_X^4} \lambda_3}{\sigma_Y^2 - 2\frac{\sigma_{XY}}{\sigma_X^2} \sigma_{XY} + \frac{\sigma_{XY}^2}{\sigma_X^4} \sigma_X^2} =$$
$$\frac{\lambda_1 (b) - \frac{\sigma_{XY}}{\sigma_X^2} \lambda_2 (b)}{\sigma_Y^2 - \frac{\sigma_{XY}}{\sigma_X^2} \sigma_{XY}} > \frac{\lambda_2 (b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$$

where the second equality uses that $\frac{\lambda_2(b)}{\sigma_{XY}} = \frac{\lambda_3}{\sigma_X^2}$ for $b \in \mathcal{I}$, and the last inequality follows from the fact that $\frac{\lambda_1(b)}{\sigma_Y^2} > \frac{\lambda_2(b)}{\sigma_{XY}}$ whenever $\sigma_Y^2 \frac{\lambda_3}{\sigma_X^2} - \lambda_1 (b) \neq 0$ and $b \in \mathcal{I}$. Since $\lim_{\beta \to \pm\infty} Q_{LIML} (\beta; b) = \frac{\lambda_3}{\sigma_X^2}$, this implies that the unique interior solution to the first order conditions is not a global minimum. Since $Q_{LIML} (\beta; b)$ is continuous and everywhere differentiable in β, this is only possible if $\lim_{\beta \to \pm\infty} Q_{LIML} (\beta; b) = \inf_\beta Q_{LIML} (\beta; b)$, from which it follows that $\beta_{LIML} (b) = \{-\infty, +\infty\}$. □

Proof of Proposition 2. This follows immediately from Proposition 3, using the fact that we can recover LIML as a special case of CUGMM by setting

$$(\Gamma_{11}, \Gamma_{12}, \Gamma_{22}) = \left(\sigma_Y^2 E\left[Z_t Z_t'\right], \sigma_{XY} E\left[Z_t Z_t'\right], \sigma_X^2 E\left[Z_t Z_t'\right]\right).$$

In this case

$$-\widetilde{\Gamma}_{12} - \widetilde{\Gamma}_{12}' + 2\beta \widetilde{\Gamma}_{22} = \left(-2\sigma_{XY} + 2\beta \sigma_X^2\right) \Xi$$

and

$$\Omega(\beta) = \frac{1}{\sigma_Y^2 - 2\beta\sigma_{XY} + \beta^2\sigma_X^2}\Xi.$$

Plugging these expressions into the result of Proposition 3 and dividing through by $\frac{-2\sigma_{XY}+2\beta\sigma_X^2}{\sigma_Y^2-2\beta\sigma_{XY}+\beta^2\sigma_X^2}$ completes the proof. □

Proof of Corollary 1. Fix a point $b \in \mathcal{I} \setminus \mathcal{S}$. Suppose we have a sequence of points $b_n \to b$, where $b_n \notin \mathcal{I}$ for all n. I first show that $|\beta_{LIML}(b_n)| \to \infty$.

Define

$$R(\beta; b) = (b - A(\beta))' \,\Xi\, (b - A(\beta)) = \lambda_1(b) - 2a(\beta)\lambda_2(b) + a(\beta)^2\lambda_3 \tag{14}$$

for $a(\beta) = \frac{\beta^2\sigma_X^2-\sigma_Y^2}{-2\sigma_{XY}+2\beta\sigma_X^2}$, and recall that by Proposition 2, $\tilde{b} \in \mathcal{F}(\beta)$ if and only if $R\left(\beta; \tilde{b}\right) = C(\beta)$. Let b^* be an arbitrary point in $\mathcal{S}$, where the proof of part (3) of Proposition 1 establishes that $b^* \in \mathcal{F}(\beta)$ for all β. Note that since $b \in \mathcal{I} \setminus \mathcal{S}$

$$R(\beta; b) - R\left(\beta; b^*\right) = \lambda_1(b) - \lambda_1\left(b^*\right) > 0. \tag{15}$$

where we have used that $b, b^* \in \mathcal{I}$ implies $\lambda_2(b) = \lambda_2(b^*)$, and that $b^* \in \mathcal{S}, b \notin \mathcal{S}$ implies $\lambda_1(b) > \lambda_1(b^*)$.

Thus, the point b lies outside $\mathcal{F}(\beta)$ for all β. Indeed, since (under the norm $\|x\|_\Xi = \sqrt{x'\Xi x}$) $\mathcal{F}(\beta)$ is a circle centered at $A(\beta)$, the distance from b to $\mathcal{F}(\beta)$ is

$$d(b, \mathcal{F}(\beta)) = \inf_{b \in \mathcal{F}(\beta)} \|b - \tilde{b}\|_\Xi = \sqrt{R(\beta; b)} - \sqrt{R(\beta; b^*)} = \sqrt{R(\beta; b)} - \sqrt{C(\beta)}$$

where the final equality follows by the definition of $\mathcal{F}(\beta)$.

Note that $d(b, \mathcal{F}(\beta))$ is a continuous function of β. Thus, for any $L > 0$

$$\inf_{\beta \in [-L, L]} d(b, \mathcal{F}(\beta)) = \epsilon(L) > 0 \tag{16}$$

since otherwise there must exist some $\tilde{\beta} \in [-L, L]$ with $d\left(b; \mathcal{F}\left(\tilde{\beta}\right)\right) = 0$, which would contradict (15). Note, further, that $\epsilon(L)$ is decreasing in L by definition. Since $b_n \to b$, we know that $\|b - b_n\|_\Xi \to 0$. Thus, since $b_n \in \mathcal{F}(\beta_{LIML}(b_n))$ by Lemma 2 along with (11), we know that $d(b, \mathcal{F}(\beta_{LIML}(b_n))) \to 0$ as $n \to \infty$. Given (16), however, this implies that $|\beta_{LIML}(b_n)| \to \infty$.

Finally, suppose the sequence of points b_n satisfies $\lambda_2(b_n) > \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3$ for all n. I claim that in this case $\beta_{LIML}(b_n) \to +\infty$. The proof proceeds by contradiction: in particular, suppose $\beta_{LIML}(b_n) \not\to +\infty$. Then by the argument in the previous paragraph there exists a subsequence b_{n_r} such that $b_{n_r} \to b$ as $r \to \infty$, $\lambda_2\left(b_{n_r}\right) > \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3$, and $\beta_{LIML}\left(b_{n_r}\right) \to -\infty$. To simplify notation I assume $b_{n_r} \equiv b_n$.

Using (14), for $b^* \in \mathcal{S}$ we have that for $\beta_n = \beta_{LIML}(b_n)$

$$R(\beta_n; b_n) - R\left(\beta_n, b^*\right) = \lambda_1(b_n) - \lambda_1\left(b^*\right) - 2a(\beta_n)\left(\lambda_2(b_n) - \lambda_2\left(b^*\right)\right).$$

Since $b_n \to b$ and $\lambda_1(b) > \lambda_1(b^*)$, there exists an N_1 such that for $n > N_1$, $\lambda_1(b_n) > \lambda_1(b^*)$. Likewise, since $a(\beta) < 0$ for $\beta < -\frac{\sigma_Y}{\sigma_X}$, there exists N_2 such that for $n > N_2$, $a(\beta_n) < 0$. Thus, since we assumed that $\lambda_2(b_n) > \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3 = \lambda(b^*)$,

$$-2a(\beta_n)\left(\lambda_2(b_n) - \lambda_2\left(b^*\right)\right) > 0$$

which implies that $R(\beta_n; b_n) - R(\beta_n, b^*) > 0$ for $n \geq \max\{N_1, N_2\}$. This implies that $b_n \notin \mathcal{F}(\beta_n)$, so we have reached a contradiction. Thus, $\beta_{LIML}(b_n) \to +\infty$. An argument along the same lines establishes that if $\lambda_2(b_n) < \frac{\sigma_{XY}}{\sigma_X^2}\lambda_3$ then $\beta_{LIML}(b_n) \to -\infty$, and so completes the proof. □

Proof of Proposition 3. Note that

$$Q_{CUGMM}(\beta; b) = (b - \iota\beta)' \,\Omega(\beta)\, (b - \iota\beta)$$

where

$$\Omega(\beta) = \left(\widetilde{\Gamma}_{11} - \beta\left(\widetilde{\Gamma}_{12} + \widetilde{\Gamma}_{12}'\right) + \beta^2\widetilde{\Gamma}_{22}\right)^{-1}$$

and $\widetilde{\Gamma}_{ij} = D(E[ZX_t])^{-1}\,\Gamma_{ij} D(E[ZX_t])^{-1}$. Thus,

$$\frac{\partial}{\partial\beta}Q_{CUGMM}(\beta; b) = -2\iota'\Omega(\beta)(b - \iota\beta) - (b - \iota\beta)'\,\Omega(\beta)\left(-\widetilde{\Gamma}_{12} - \widetilde{\Gamma}_{12}' + 2\beta\widetilde{\Gamma}_{22}\right)\Omega(\beta)(b - \iota\beta).$$

Defining

$$\Psi(\beta) = \Omega(\beta)\left(-\widetilde{\Gamma}_{12} - \widetilde{\Gamma}_{12}' + 2\beta\widetilde{\Gamma}_{22}\right)\Omega(\beta)$$

and completing the square in b yields

$$\frac{\partial}{\partial \beta} Q_{CUGMM}(\beta; b) =$$

$$-\left(b - \iota\beta + \Psi(\beta)^{-1}\Omega(\beta)\iota\right)' \Psi(\beta)\left(b - \iota\beta + \Psi(\beta)^{-1}\Omega(\beta)\iota\right) + \iota'\Omega(\beta)\Psi(\beta)^{-1}\Omega(\beta)\iota =$$

$$-\left(b + \left(I_k\beta - \Psi(\beta)^{-1}\Omega(\beta)\right)\iota\right)' \Psi(\beta)\left(b + \left(I_k\beta - \Psi(\beta)^{-1}\Omega(\beta)\right)\iota\right) + \iota'\left(-\widetilde{\Gamma}_{12} - \widetilde{\Gamma}_{21} + 2\beta\widetilde{\Gamma}_{22}\right)^{-1}\iota$$

which immediately implies the result. □

References

Angrist, J., Imbens, G., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. J. Amer. Statist. Assoc. 90, 431–442.

Chamberlain, G., 2007. Decision theory applied to an instrumental variable model. Econometrica 75, 609–652.

Hall, A.R., Inoue, A., 2003. The large sample behaviour of the generalized method of moments estimator in misspecified models. J. Econometrics 114 (2), 361–394.

Hausman, J., 1983. Handbook of Econometrics, Vol. 1. In: Griliches, Z., Intriligator, M. (Eds.), North-Holland Publishing Company, pp. 391–448, chapter Specification and Estimation of Simultaneous Equation Models.

Imbens, G., Angrist, J., 1994. Identificaton and estimation of local average treatment effects. Econometrica 62, 467–475.

Kolesar, M., 2013. Estimation in an instrumental variables model with treatment effect heterogenity. Unpublished Manuscript.

Lee, S., 2017. A consistent variance estimator for 2sls when instruments identify different lates. J. Bus. Econom. Statist. Forthcoming.

Mogstad, M., Santos, A., Torgovitsky, A., 2018. Using instrumental variables for inference about policy-relevant treatment effects. Econometrica 86, 1589–1619.

Stock, J., Wright, J., 2000. Gmm with weak identification. Econometrica 68, 1055–1096.

van der Vaart, A., 2000. Asymptotic Statistics. Cambridge University Press.

Yogo, M., 2004. Estimating the elasticity of intertemporal substitution when instruments are weak. Rev. Econ. Statist. 86, 797–810.