Underidentification?*

Manuel Arellano
CEMFI
arellano@cemfi.es

Lars Peter Hansen
University of Chicago
lhansen@uchicago.edu

Enrique Sentana
CEMFI
sentana@cemfi.es

July 31, 2009

Abstract

We develop methods for testing the hypothesis that an econometric model is underidentified and inferring the nature of the failed identification. By adopting a generalized-method-of moments perspective, we feature directly the structural relations and we allow for nonlinearity in the econometric specification. We establish the link between a test for overidentification and our proposed test for underidentification. If, after attempting to replicate the structural relation, we find substantial evidence against the overidentifying restrictions of an augmented model, this is evidence against underidentification of the original model.

1 Introduction

It is common in econometric practice to encounter one of two different phenomena. Either the data are sufficiently powerful to reject the model, or the sample evidence is sufficiently weak so as to suspect that identification is tenuous. The early simultaneous equations literature recognized that underidentification is testable, but to date such tests are uncommon in econometric practice despite the fact that there are many situations of economic interest in which seemingly point identified models may be only set identified.

We adopt a generalized-method-of-moments (GMM) perspective and provide a way to test for underidentification (an I test) using statistics that are commonly employed as tests for overidentification (J tests). More specifically, we consider an augmented structural model

*This is an substantially revised and extended version of Arellano et al. (1999). We thank Javier Alvarez, Raquel Carrasco, Jesús Carro and especially Francisco Peñaranda and Javier Mencía for able research assistance at various stages of this project. Also, we thank Andres Santos and Azeem Shaikh for helpful conversations about our general approach to this problem. Arellano and Sentana gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 and Hansen gratefully acknowledges support from the National Science Foundation through grant SES0519372.

in which the moment conditions are satisfied by a curve instead of a point. In this context, our proposal is to test for underidentification by testing for overidentification in the augmented model. This leads us to adapt or extend standard overidentifying testing methods available in the literature. If it is possible to estimate a curve without statistically rejecting the overidentifying restrictions of the augmented model, then we may conclude that the original econometric relation could be *not identified*, or equivalently, that it is underidentified. In contrast, rejections provide evidence that the original model is indeed point identified.

We consider in progression three different estimation environments: linear in parameters models (section 3), models with nonlinear restrictions on the parameters (section 4), and finally, more fundamental nonlinearities (section 5). Throughout we develop specific examples in detail to illustrate the nature and the applicability of I testing. In the next section we provide a more detailed overview of the paper.

2 Overview

As in Hansen (1982), suppose that $\{x_t\}$ is an observable stationary and ergodic stochastic process[1] and let $\mathbb{P}$ be a parameter space that we take to be a subset of $\mathbb{R}^k$. Introduce a function $f(x,\cdot) : \mathbb{P} \to \mathbb{R}^p$ for each x. The function f is jointly Borel measurable and it is continuously differentiable in its second argument for each value of x. Finally suppose that $E|f(x_t,\beta)| < \infty$ for each $\beta \in \mathbb{P}$.

In light of this assumption we define $Ef(x_t,\beta) = \bar{f}(\beta)$ for each $\beta \in \mathbb{P}$. GMM estimation uses the equation:

$$\bar{f}(\beta) = 0 \tag{1}$$

to identify a parameter vector β_0. When β_0 is identified, it is the unique solution to (1), otherwise there will be multiple solutions.

We pose an alternative estimation problem as a device to formally explore underidentification. We consider a parameterization of the form: $\beta = \pi(\theta)$, where $\theta \in \Theta$ for some conveniently chosen domain Θ with a corresponding norm and π is a continuous function

[1] As elsewhere in the econometrics literature, analogous results can be obtained using other data generating processes. For cross-sectional and panel extensions of Hansen (1982) formulation see the textbooks by Hayashi (2000) and Arellano (2003), respectively.

with range $\mathbb{P}$. For example, suppose that

$$\pi(\theta) = \begin{bmatrix} \theta \\ \tau(\theta) \end{bmatrix} \tag{2}$$

so that θ is the first component of the parameter vector. We then explore a set of such functions that is restricted appropriately.[2]

As an alternative identification condition, we require $\bar{f}[\pi(\theta)] = 0$ for all $\theta \in \Theta$ if, and only if $\pi = \pi_0$.

If we can successfully identify a nonconstant function π_0 that realizes alternative values in the parameter space, then we cannot uniquely identify a single parameter vector β_0 from the moment conditions (1). Thus the parameter vector β_0 is underidentified. Conveniently, this estimation problem looks like a standard estimation problem except that we seek to estimate a function instead of a finite-dimensional parameter vector. This naturally leads to a test of underidentification of β_0 based on an attempt to identify π_0. The resulting test is the counterpart to the GMM overidentification test of Sargan (1958) and Hansen (1982). Henceforth, we shall refer to this test as an I test.

Our development of a statistical test for underidentification leads naturally to the question of how to estimate π_0 efficiently. One approach would be to use one of the standard GMM objective functions and try to construct an estimator of π_0 as an approximate minimizer of a quadratic form. In this paper we explore a rather different approach. As in Sargan (1959), our approach requires that we put an explicit structure on the lack of identification and this structure will be evident when we construct our parameterization of π used in our alternative (and weaker) identification condition 2.

To motivate the efficiency gain, consider the special case in which Θ consists of two different known values, $\theta^{[1]}$ and $\theta^{[2]}$, say, and $\pi(\theta)$ can be written as in (2).

Example 2.1. *In this case we seek to identify $\pi_0(\theta^{[1]})$ and $\pi_0(\theta^{[2]})$, with $\pi_0(\theta^{[1]}) \neq \pi_0(\theta^{[2]})$. We can map this into a standard GMM problem where we simply stack or duplicate the moment conditions. This leads us to consider GMM estimators of the unknown elements of $[\pi(\theta^{[1]}), \pi(\theta^{[2]})]$ as approximate solutions to:*

$$A_N \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} f[x_t, \pi_N(1)] \\ f[x_t, \pi_N(2)] \end{bmatrix} = 0. \tag{3}$$

[2]See section 5.1 for further details.

for a selection matrix A_N that is $2(k-1)$ by $2p$. Form

$$K(i,j) = \lim_{N\to\infty} \frac{1}{N} E\left[\sum_{t=1}^{N} \left(f[x_t, \pi_0(i)] - \bar{f}[\pi_0(i)]\right) \sum_{s=1}^{N} \left(f[x_s, \pi_0(j)] - \bar{f}[\pi_0(j)]\right)'\right]$$

and

$$D(i) = \left.\frac{\partial \bar{f}}{\partial \beta_2}\right|_{\pi_0(i)}.$$

where we have partitioned β as (β_1, β_2) so that it conforms with the partition of $\pi(\theta)$ in (2). From Hansen (1982), the efficient limiting selection matrix to be used in (3) *will be*

$$A^* = \begin{bmatrix} D(1)' & 0 \\ 0 & D(2)' \end{bmatrix} \begin{bmatrix} K(1,1) & K(1,2) \\ K(2,1) & K(2,2) \end{bmatrix}^{-1}.$$

If by chance $K(1,2) = 0$, the efficient selection matrix would simplify to be

$$\begin{bmatrix} D(1)'K(1,1)^{-1} & 0 \\ 0 & D(2)'K(2,2)^{-1} \end{bmatrix},$$

and there would be no gain from efficient estimation. In our applications $K(1,2)$ is typically different from zero so there are potential gains for joint estimation. Thus we are lead to modify the GMM objective function when we seek to estimate the function π_0 instead of β_0 and use this as the basis of a GMM-based test of underidentification.

We include this two-point example only by way of illustration. For the estimation problems in this paper we seek to identify subspaces in the case of linear models and curves in the case of nonlinear models and design our GMM problem accordingly.

Our work is related to two different strands of the literature that have gained prominence in recent years. One is the weak instruments literature (see *e.g.* Stock et al. (2002)), which maintains the assumption that the rank condition is satisfied, but only just. To relate to this line of research suppose that Θ is an interval and consider an interior point θ^*. Suppose that π is differentiable at θ^*. Then under appropriate regularity conditions:

$$\left[\left.\frac{\partial \bar{f}(\beta)}{\partial \beta}\right|_{\beta^*}\right] \left[\left.\frac{d\pi(\theta)}{d\theta}\right|_{\theta^*}\right] = 0, \tag{4}$$

where $\beta^* = \pi(\theta^*)$. In other words, the matrix

$$\left[\left.\frac{\partial \bar{f}(\beta)}{\partial \beta}\right|_{\beta^*}\right]$$

has reduced rank for any θ^* in the interior of Θ. In contrast the weak instruments literature considers the reduced rank as the limit of a sequence of data generating models indexed by the sample size.[3] In our analysis such a sequence could be interesting as a local specification under the alternative hypothesis of identification. We seek to infer the specific manner in which identification may fail whereas the weak instrument literature focuses on developing reliable standard errors and tests of hypotheses about a unique true value of β.

The other strand is the set estimation literature (see *e.g.* Chernozhukov et al. (2007)), which often assumes up-front that $E[f(x;\beta)] = 0$ for some manifold of values of β that includes β_0, possibly because β_0 is unequivocally underidentified *a priori*, and whose objective is to make inferences about this manifold.[4] In contrast, in this paper we focus on models in which β_0 is not unequivocally underidentified. Given this focus, unstructured underidentification will not be of interest in general. By adding a particular structure to the identification failure, we are led to alter the usual GMM objective in order to estimate efficiently the one-dimensional function π that parameterizes the potential lack of identification.

3 Linear in the Parameters

We first study the identification of an econometric model that is linear in parameters, in which case we can write (1) as:

$$E(\Psi_i)\alpha = 0, \tag{5}$$

where α is a $k+1$-dimensional unknown parameter vector in the null space of the population matrix $E(\Psi_i)$, and Ψ_i is an r by $k+1$ matrix constructed from data.[5] If there is a solution α_0 to this equation, then any scale multiple of α_0 will also be a solution. Thus from a statistical perspective, we consider the problem of identifying a direction. To go from a direction to the parameters of interest requires an additional *scale normalization* of the form $q'\alpha = 1$, where q is a $k+1$ vector that is specified *a priori*. For instance, we could choose q to be a member of the canonical basis, which would restrict one of the components of α to be one

[3] Typically in this literature the rank is not just reduced but is zero in the limit.

[4] Some of this literature also considers moment inequalities as a source of underidentification. Our analysis does not cover this situation.

[5] Therefore, we consider not only models which are linear in both variables and parameters, but also the non-linear in variables but linear in parameters models discussed in chapter 5 of Fisher (1966), which combine different non-linear transformations of the same variables.

as in:

$$\alpha = \begin{bmatrix} 1 \\ -\beta \end{bmatrix}$$

Alternatively, we could choose $q = \alpha$ so that $|\alpha| = 1$, together with a sign restriction on one of the nonzero coefficients as in:

$$\alpha = \begin{bmatrix} +\sqrt{1 - |\beta|^2} \\ \beta \end{bmatrix}$$

where $|\beta| \leq 1$. Neither of these approaches can be employed without loss of generality, however. The particular application dictates how to select the parameters of interest from this direction.[6]

Consider now an alternative specification that reflects a precise form of underidentification. Let

$$\pi(\theta) = \theta\alpha^{[1]} + (1 - \theta)\alpha^{[2]} \tag{6}$$

where we restrict $\alpha^{[1]}$ to have ones in the first two positions and $\alpha^{[2]}$ to a have one in the first position and a zero in the second position.[7] The null hypothesis of interest is:

$$E\left(\Psi_i\right)\pi(\theta) = 0 \ \forall\theta \in \mathbb{R} \tag{7}$$

and some $\alpha^{[1]}$ and $\alpha^{[2]}$. Given the linear structure of (5), it suffices to check this restriction at two values of θ, say $\theta = 0$ and $\theta = 1$. This leads us to efficiently estimate those $\alpha_0^{[1]}$ and $\alpha_0^{[2]}$ for which

$$\left.\begin{array}{l} E(\Psi_i)\alpha^{[1]} = 0 \\ E(\Psi_i)\alpha^{[2]} = 0 \end{array}\right\} \tag{8}$$

The duplicated moment conditions (8) give us a direct link to the rank condition familiar in the econometrics literature. Suppose the order condition ($r \geq k$) is satisfied, but not necessarily the rank condition. Thus the maximal possible rank of the matrix $E(\Psi_i)$ is $\min\{r, k+1\}$. Model (5) is said to be *identified* when $E(\Psi_i)$ has rank k in which case its null space is precisely one dimensional. When $r > k$ and the model is identified, it is said

[6]Sensitivity to the choice of normalization can be avoided in GMM by using the approach of Hillier (1990) and Alonso-Borrego and Arellano (1999) or by using the continuously-updated estimator of Hansen et al. (1996). As a consequence, our more general rank formulation can be explored using such methods.

[7]Strictly speaking these are more than just normalizations. Other normalizations (see section 3.1) are also not only possible but also desirable in some applications.

to be *overidentified* because the rank of the matrix $E(\Psi_i)$ now must not be full. Instead of having maximal rank $k+1$, $E(\Psi_i)$ has reduced rank k. This implication is known to be testable and statistical tests of overidentification are often conducted in practice.

In contrast, model (5) is said to be *underidentified* when the rank of $E(\Psi_i)$ is less than k. In this case the null space of $E(\Psi_i)$ will have more than one dimension. A single normalization will no longer select a unique element from the parameter space. By focusing on (6), our approach puts an explicit structure on the lack of identification, as illustrated by (8). Thus, we initially make the following assumption (see section 3.1.2 for other possibilities):

Hypothesis 3.1. *$E(\Psi_i)$ has rank $k-1$.*

Under this hypothesis the set of solutions to equation (5) is two-dimensional. To test for this lack of identification, we think of (8) as a new *augmented model*. We attempt to determine $(\alpha^{[1]}, \alpha^{[2]})$ simultaneously and ask whether they satisfy the combined overidentifying moment restrictions (8). If they do, then we may conclude that the original econometric relation is *not identified* or equivalently is underidentified. Thus by building an augmented equation system, we may pose the null hypothesis of underidentification as a hypothesis that the augmented equation system is overidentified. Rejections of the overidentifying restrictions for the augmented model provide evidence that the original model is indeed identified. Posed in this way, underidentification can be tested simply by applying appropriately an existing test for overidentification. For instance, a standard J test for overidentification, such as those of Sargan (1958) and Hansen (1982), is potentially applicable to the augmented model. This test will be our I test.

The following example illustrates our formulation.

Example 3.1. *Suppose that $r=1$ and $k=1$. Write*

$$E(\Psi_i) = \begin{bmatrix} a_1 & a_2 \end{bmatrix}$$

For there to be identification in the sense that we consider, at least one of the entries of this vector must be different from zero. If we normalize the first entry of $\alpha' = \begin{bmatrix} 1 & -\beta \end{bmatrix}$ to be one, then we obtain the more restrictive rank condition condition that $a_2 \neq 0$. The "normalization" rules out the case that $E(\Psi_i)$ is of the form $\begin{bmatrix} 0 & a_2 \end{bmatrix}$ and $\alpha' = \begin{bmatrix} \alpha_1 & 0 \end{bmatrix}$. Our notion of identification includes this possibility.

To understand better implementation, in the remainder of this section we consider as examples three specific situations: single equation IV, multiple equations with cross-equation restrictions, and sequential moment conditions.

3.1 Single equation IV

Example 3.2. *Suppose the target of analysis is a single equation from a simultaneous system:*

$$y_i \cdot \alpha = u_i \tag{9}$$

where the scalar disturbance term u_i is orthogonal to an r-dimensional vector z_i of instrumental variables:

$$E\left(z_i u_i\right) = 0. \tag{10}$$

Form:

$$\Psi_i = z_i y_i'.$$

Then orthogonality condition (10) is equivalent to α satisfying the moment relation (5).

For this example we duplicate the moment conditions as in (8), and study the simultaneous overidentification of those $2r$ moment conditions. To proceed with the construction of a test, we have to rule out the possibility that $\alpha^{[1]}$ and $\alpha^{[2]}$ are proportional. One strategy is to restrict $\alpha^{[2]}$ to be orthogonal to $\alpha^{[1]}$. Two orthogonal directions can be parameterized with $2k-1$ parameters, k parameters for one direction and $k-1$ for the orthogonal direction. However, there is not a unique choice of orthogonal directions to represent a two-dimensional space. There is an additional degree of flexibility. A new direction can be formed by taking linear combinations of the original two directions and a corresponding orthogonal second direction. Thus the number of required parameters is reduced to $2k-2$, and the number of overidentifying restrictions for the I test of underidentification is $2r-2k+2$.

In practice, we can impose the normalizing restrictions $|\alpha^{[1]}| = |\alpha^{[2]}| = 1$ by using spherical coordinates, force $\alpha^{[1]} \cdot \alpha^{[2]} = 0$, and set the first entry of $\alpha^{[2]}$ to zero. This works provided that all vectors in the null space of $E(z_i y_i{}')$ do not have zeros in the first entry. Alternatively, we could restrict the top two rows $(\alpha^{[1]}, \alpha^{[2]})$ to equal an identity matrix of order two. This rules out the possibility of a vector in the null space that is identically zero

in its first two entries, but this may be of little concern for some applications.[8] When $k = 1$, both approaches boil down to setting $(\alpha^{[1]}, \alpha^{[2]}) = I_2$ so that the $2r$ moment conditions:

$$E\left(z_i y_i'\right) = 0$$

can be represented without resort to parameter estimation. As a result, the "identified" set will be the whole of $\mathbb{R}^2$.

Example 3.1 could emerge as a special case of example 3.2 with $r = 1$ and $k = 1$. Notice that our underidentification test in this case tests simultaneously the restriction that $a_1 = 0$ and $a_2 = 0$. More generally, when $r \geq 2$ our test considers simultaneously $E(z_i y_{1,i}) = 0$ and $E(z_i y_{2,i}) = 0$. The resulting I test is different from the test for the relevance of instruments in a model with a normalization restriction on one variable to be estimated by say two-stage least squares. Such a test would examine only $E\left(z_i y_{2,i}\right) = 0$.

In contrast, when $k > 1$, some parameters must be inferred as part of implementing the I test. The estimated parameters can then be used for efficiently estimating the identified linear set by exploiting (6). To illustrate this point, suppose a normalized relationship between three endogenous variables with instrument vector z_i:

$$E\left[z_i\left(y_{0i} - \alpha_1 y_{1i} - \alpha_2 y_{2i}\right)\right] = 0.$$

Now z_i need not be uncorrelated to the three endogenous variables for underidentification. Lack of correlation with two linear combinations of them is enough. For example, we may write the null of underidentification as

$$H_0 : E\left[\begin{array}{c} z_i\left(y_{0i} - \beta y_{2i}\right) \\ z_i\left(y_{1i} - \gamma y_{2i}\right) \end{array}\right] = 0.$$

If H_0 holds, for any α_1^*

$$E\left\{z_i\left[y_{0i} - \alpha_1^* y_{1i} - \left(\beta - \gamma\alpha_1^*\right) y_{2i}\right]\right\} = 0,$$

so that the observationally equivalent values (α_1^*, α_2^*) are contained in the line $\alpha_2^* = \beta - \gamma\alpha_1^*$.

A time series example is a forward-looking Phillips curve as in Galí et al. (2001), where the components of y denote current inflation, future inflation, and a measure of aggregate

[8] Once again, it is desirable to construct a test statistic of underidentification using a version of the test of overidentifying restrictions that is invariant to normalization

demand, whereas the components of z consist of lags of the previous variables, and of other variables such as the output gap and wage inflation. There are theoretical and empirical considerations to suggest that a null like H_0 is plausible in this context. For example, lack of higher-order dynamics in a new Keynesian macro model has been shown to be a source of underidentification of a hybrid Phillips curve with lagged inflation by Nason and Smith (2008). Relatedly, Cochrane (2007) also raises similar concerns regarding the identification of Taylor rules by Clarida et al. (2000) and others.

3.1.1 Related Literature

Tests of underidentification in a single structural equation were first considered by Koopmans and Hood (1953) and Sargan (1958). When the model is correctly specified and identified, the rank of $E(z_i y_i')$ is k. Under the additional assumptions that the error term u_i is a conditionally homoskedastic martingale difference, an asymptotic chi-square test statistic of overidentifying restrictions with $r-k$ degrees of freedom is given by $N\lambda_1$, where

$$\lambda_1 = \min_{\alpha} \frac{\alpha' Y' Z \left(Z'Z\right)^{-1} Z' Y \alpha}{\alpha' Y' Y \alpha}, \tag{11}$$

and $Z'Y = \sum_{i=1}^{N} z_i y_i'$, etc. Thus λ_1 is the smallest characteristic root of $Y'Z\left(Z'Z\right)^{-1}Z'Y$ in the metric of $Y'Y$. (See Anderson and Rubin (1949) and Sargan (1958)). This a version of the J test for overidentification, and it does not require that we normalize α.

Koopmans and Hood (1953) and Sargan (1958) indicated that when the rank of $E(z_i y_i')$ is $k-1$ instead, if λ_2 is the second smallest characteristic root, $N(\lambda_1+\lambda_2)$ has an asymptotic chi-square distribution with $2(r-k)+2$ degrees of freedom. These authors suggested that this result could be used as a test of the hypothesis that the equation is underidentified and that any possible equation has an iid error term.

The statistic $N(\lambda_1+\lambda_2)$ has a straightforward interpretation in terms of our approach. Indeed, it can be regarded as a continuously-updated GMM test of overidentifying restrictions of the augmented model (8), subject to the additional restrictions on the error terms mentioned previously. To see this, let $A = \begin{bmatrix}\alpha^{[1]} & \alpha^{[2]}\end{bmatrix}$ and consider the minimizer of

$$\begin{bmatrix}\alpha^{[1]\prime} Y' Z & \alpha^{[2]\prime} Y' Z\end{bmatrix} \left(A' Y' Y A \otimes Z'Z\right)^{-1} \begin{bmatrix} Z'Y\alpha^{[1]} \\ Z'Y\alpha^{[2]}\end{bmatrix}$$

subject to $A'Y'YA = I_2$. The constraint restricts the sample covariance matrix of the disturbance vector to be an identity matrix. It uses the positive definite matrix $Y'Y$ to define orthogonal directions when duplicating equations, which is convenient for this application. In light of this normalization, the minimization problem may be written equivalently as

$$\min_{A'Y'YA=I_2} \alpha^{[1]\prime}Y'Z\left(Z'Z\right)^{-1}Z'Y\alpha^{[1]} + \alpha^{[2]\prime}Y'Z\left(Z'Z\right)^{-1}Z'Y\alpha^{[2]}, \tag{12}$$

and the minimized value coincides with $\lambda_1 + \lambda_2$ (Rao (1973), page 63). A comparison of (12) with (11) makes clear that the I test will be numerically at least as large as the J test, a result that is a special case of Proposition B.2 in Appendix B. This comparison also shows that the estimate of α obtained from (11) coincides with the estimate of $\alpha^{[1]}$ obtained from (12), so that in this special case the optimal point estimate belongs to the optimal linear set estimate.

More recently, Cragg and Donald (1993) considered single equation tests of underidentification based on the reduced form. For the single equation model, the rank of the matrix $E\left(\Psi_i\right)$ is the same as that of

$$P = E\left(\Psi_i\right)'\left[E(z_i z_i')\right]^{-1} = E\left(y_i z_i'\right)\left[E(z_i z_i')\right]^{-1}.$$

This is the matrix of coefficients of the reduced form system of population regressions of the entries of y_i onto z_i. Suppose the second component of y_i is the first component of z_i. Partition P as:

$$P = \begin{bmatrix} \Pi_1 & \Pi_2 \\ I & 0 \end{bmatrix}.$$

The nullity of P and hence $E\left(\Psi_i\right)$ is the same as the nullity of Π_2. Cragg and Donald (1993) construct a minimum chi-square test statistic that enforces then rank restriction in Π_2.[9] Their statistic can also be related to our approach. As we show in Appendix A, under the assumption that u_i is a conditionally homoskedastic martingale difference, the Cragg-Donald statistic minimizes

$$\begin{bmatrix} \alpha^{[1]\prime}Y'Z & \alpha^{[2]\prime}Y'Z \end{bmatrix}\left(A'Y'MYA \otimes Z'Z\right)^{-1}\begin{bmatrix} Z'Y\alpha^{[1]} \\ Z'Y\alpha^{[2]} \end{bmatrix}$$

[9]Cragg and Donald also considered an alternative null of no identifiability in an equation with the coefficient of one of the endogenous variables normalized to unity. This was a rank restriction in the submatrix of Π_2 that excludes the row corresponding to the normalized entry.

subject to $A'Y'MYA = I_2$ where $M = I - Z\left(Z'Z\right)^{-1}Z'$. Moreover, a Cragg-Donald statistic that is robust to heteroskedasticity and/or serial correlation can be reinterpreted as a continuously updated GMM criterion of the augmented structural model using MYA as errors in the weight matrix. Since the difference between YA and MYA at the truth is of small order, using one form of errors or the other is asymptotically irrelevant.

While the Cragg and Donald (1993) approach is straightforward to implement in the single-equation case, it is more difficult to implement in some models with cross-equation restrictions. This difficulty can emerge because we must simultaneously impose the restrictions on the reduced form together with the rank deficiency. In example 3.2, this is easy to do, and it is also feasible in the applications to linear observable factor pricing models of asset returns carried out by Cragg and Donald (1997) and Burnside (2007), but not in more general models as we will illustrate in sections 3.2 and 3.3.

3.1.2 Underidentification of a higher dimension

Although the null hypothesis 3.1 is the natural leading case in testing for underidentification, it is straightforward to extend the previous discussion to situations in which the underidentified set is of a higher dimension. Suppose that the rank of $E(\Psi_i)$ is $k-j$ for some j. Then we can write all the admissible equations as linear combinations of the $(j+1)r$ orthogonality conditions

$$E(\Psi_i)\left(\alpha^{[1]}, \alpha^{[2]}, ..., \alpha^{[j+1]}\right) = 0. \tag{13}$$

If we impose $(j+1)^2$ normalizing restrictions on $(\alpha^{[1]}, \alpha^{[2]}, ..., \alpha^{[j+1]})$ to avoid indeterminacy,[10] the effective number of parameters is $(j+1)(k+1) - (j+1)^2 = (j+1)(k-j)$ and the number of moment conditions is $(j+1)r$ under the assumption that there are no redundancies. Therefore, by testing the $(j+1)(r-k+j)$ overidentifying restrictions in (13) we test the null that α is underidentified of dimension j against the alternative of underidentification of dimension less than j or identification. Henceforth, we shall refer to those tests as I_j tests.

[10] For instance, we may make the top $j+1$ rows of $A^{[j+1]} = (\alpha^{[1]}, \alpha^{[2]}, ..., \alpha^{[j+1]})$ equal to the identity matrix of order $j+1$. More generally, we can impose the $(j+1)^2$ normalizing restrictions $A^{[j+1]\prime}A^{[j+1]} = I_{(j+1)}$ and $a_{i\ell} = 0$ for $\ell > i$, where $a_{i\ell}$ denotes the (i,ℓ)-th element of $A^{[j+1]}$.

3.2 Multiple equations with cross-equation linear restrictions

We next consider examples with multiple equations with common parameters.[11]

Example 3.3. *Consider the following two equation model with cross-equation restrictions:*

$$\begin{aligned} \alpha \cdot \begin{bmatrix} y_{1,i} \\ y_{3,i} \end{bmatrix} &= u_{1,i} \\ \alpha \cdot \begin{bmatrix} y_{2,i} \\ y_{3,i} \end{bmatrix} &= u_{2,i} \end{aligned}$$

where $y_{1,i}$, $y_{2,i}$ are scalars. Let z_i denote an r^*-dimensional vector of instrumental variables appropriate for both equations.

$$\begin{aligned} E\left(z_i u_{1,i}\right) &= 0 \\ E\left(z_i u_{2,i}\right) &= 0. \end{aligned}$$

Form:

$$\Psi_i = \begin{bmatrix} z_i y_{1,i} & z_i y_{3,i}{}' \\ z_i y_{2,i} & z_i y_{3,i}{}' \end{bmatrix}.$$

Thus $r = 2r^*$. We transform this equation system to obtain an equivalent one by forming:

$$\Psi_i^* = \begin{bmatrix} z_i(y_{1,i} - y_{2,i}) & 0 \\ z_i y_{1,i} & z_i y_{3,i}{}' \end{bmatrix} \tag{14}$$

implying that

$$E\left[z_i(y_{1,i} - y_{2,i})\right] = 0 \tag{15}$$

In this example, duplicating (15) would induce a degeneracy because equation (15) does not depend on parameters. Instead these r^* moment conditions should be included just once. The I test is implemented by again parameterizing a two-dimensional subspace with $2k-2$ free parameters. There are $3r^* < 2r$ composite moment conditions to be used in estimating these free parameters. Thus the degrees of freedom of the I test is $3r^* - 2k + 2$.

This I test includes (15) among the moment conditions to be tested even though these conditions do not depend on the unknown parameters. If these moment conditions were excluded, then it would matter if the second row block of Ψ_i^* in (14) is replaced by $\begin{bmatrix} z_i y_{2,i} & z_i y_{3,t}{}' \end{bmatrix}$. By including (15) among the moment conditions to be tested this change is inconsequential.

[11] Interestingly, Kim and Ogaki (2009) suggest to use models with cross equation restrictions to try to break away from the potential identifiability problems that affect single equation IV estimates.

An extended version of this example arises in log-linear models of asset returns such as those studied by Hansen and Singleton (1983) and others. Such models have a scalar $y_{3,i}$ given by consumption growth expressed in logarithms. The variables $y_{1,i}$ and $y_{2,i}$ are the logarithms of gross returns. In addition there are separate constant terms in each equation that capture subjective discounting and lognormal adjustments. By differencing the equations we obtain a counterpart to (15) except that a constant term needs to be included. Duplication continues to induce a degeneracy because this constant term is trivially identified.

Example 3.4. *Consider a normalized four-input translog cost share equation system. After imposing homogeneity of degree 1 in prices and dropping one equation to take care of the adding-up condition in cost shares we have*

$$y_{ji} = \beta_{j1}p_{1i} + \beta_{j2}p_{2i} + \beta_{j3}p_{3i} + v_{ji} \qquad (j = 1, 2, 3) \tag{16}$$

where y_{ji} denotes the cost share of input j, and p_{ji} is the log price of input j relative to the omitted input.[12] *The underlying cost function implies the following three cross-equation symmetry constraints*

$$\beta_{jk} = \beta_{kj} \qquad j \neq k. \tag{17}$$

Moreover, prices are endogenous (possibly due to data aggregation) and an r -dimensional vector of instruments z_i is available:

$$E(z_i v_{ji}) = 0 \qquad (j = 1, 2, 3) \tag{18}$$

In the absence of the symmetry restrictions, the order condition is satisfied if $r \geq 3$. It would appear that the parameters may be just identified with $r = 2$ when the symmetry restrictions are taken into account, for in that case the order condition is satisfied. However, it turns out that such system has reduced rank 5 by construction.

To test for underidentification, we duplicate the original moment conditions, introduce suitable normalizations, and drop redundant moments, obtaining

$$E[z_i(y_{ji} - \gamma_{j2}p_{2i} - \gamma_{j3}p_{3i})] = 0 \qquad (j = 1, 2, 3) \tag{19}$$

$$E[z_i(p_{1i} - \gamma_{02}p_{2i} - \gamma_{03}p_{3i})] = 0 \tag{20}$$

[12] See (Berndt, 1991, page 472). For simplicity we abstract from intercepts and log output terms since they have no effect on our discussion.

Since there are $4r$ orthogonality conditions and 8 parameters, with $r = 2$ the augmented set of moments does not introduce any overidentifying restrictions. For arbitrary r, (19)-(20) imply that (18) is satisfied for any β^*_{j1}, and for $\beta^*_{j2}, \beta^*_{j3}$ $(j = 1, 2, 3)$ such that

$$\beta^*_{j2} = \gamma_{j2} - \beta^*_{j1}\gamma_{02} \qquad \beta^*_{j3} = \gamma_{j3} - \beta^*_{j1}\gamma_{03}. \tag{21}$$

Thus, if we do not impose symmetry, the identified set will be of dimension three $(\beta^*_{11}, \beta^*_{21}, \beta^*_{31})$ and will be characterized by the eight γ parameters in (19)-(20). However, one restriction must be imposed on those parameters for the augmented model to characterize observationally equivalent values of the original β parameters satisfying the symmetry constraints. To see this, note that, subject to the cross-restrictions, (19)-(20) imply that (18) are satisfied as before for any β^*_{11} (and for β^*_{12} and β^*_{13} as in (21)), but only for $\beta^*_{21} = \beta^*_{12}$ so that

$$\beta^*_{21} = \gamma_{12} - \beta^*_{11}\gamma_{02},$$

and for β^*_{22} and β^*_{23} such that

$$\beta^*_{22} = \gamma_{22} - (\gamma_{12} - \beta^*_{11}\gamma_{02})\gamma_{02} \qquad \beta^*_{23} = \gamma_{23} - (\gamma_{12} - \beta^*_{11}\gamma_{02})\gamma_{03}.$$

Equally, they are satisfied only for $\beta^*_{31} = \beta^*_{13}$ so that

$$\beta^*_{31} = \gamma_{13} - \beta^*_{11}\gamma_{03},$$

and for β^*_{32} and β^*_{33} such that

$$\beta^*_{32} = \gamma_{32} - (\gamma_{13} - \beta^*_{11}\gamma_{03})\gamma_{02} \qquad \beta^*_{33} = \gamma_{33} - (\gamma_{13} - \beta^*_{11}\gamma_{03})\gamma_{03}.$$

Moreover, the restriction $\beta^*_{32} = \beta^*_{23}$ implies that the admissible values of the coefficients in the augmented model must satisfy for any β^*_{11}:

$$\gamma_{32} - (\gamma_{13} - \beta^*_{11}\gamma_{03})\gamma_{02} = \gamma_{23} - (\gamma_{12} - \beta^*_{11}\gamma_{02})\gamma_{03}$$

or

$$\gamma_{32} - \gamma_{23} = \gamma_{13}\gamma_{02} - \gamma_{12}\gamma_{03}. \tag{22}$$

Thus, after enforcing symmetry the identified set is of dimension one (β^*_{11}) and depends on seven parameters only. The I test for this problem is a test of overidentifying restrictions

based on the moments (19)-(20) subject to (22). Enforcing (22) reduces the set of observationally equivalent parameters under the null, but this is the right way to proceed since the existence of other β's that satisfy the instrumental-variable conditions but not the symmetry conditions should not be taken as evidence of underidentification of the model.[13]

3.3 Sequential moment conditions

Consider next an example with an explicit time series structure.

Example 3.5. *Suppose that*

$$y_{i,t+2} = \begin{bmatrix} v_{i,t+2} & v_{i,t+1} & ... & v_{i,t-\ell} \end{bmatrix}'$$

for a scalar process $\{v_{i,t} : t = 1, 2, ...\}$. *Thus* $k = \ell + 2$. *Form:*

$$\alpha \cdot y_{i,t+2} = u_{i,t+2}$$

where

$$E\left[z_{i,t} u_{i,t+2}\right] = 0$$

for $t = 1, ...$ *and* $\alpha \neq 0$. *Thus*

$$E\left[z_{i,t} {y_{i,t+2}}'\right] \alpha = 0. \tag{23}$$

The dimension of the vector $z_{i,t}$ *varies with* t. *This dependence is relevant in a panel data setting in which the number of time periods is small relative to the number of individuals.*[14] *Assume that there is no redundancy among the entries of* $z_{i,t}$. *That is,* $E\left(z_{i,t} z_{i,t}'\right)$ *is nonsingular. Moreover, assume that the entries of* $z_{i,t-1}$ *are among the entries of* $z_{i,t}$.

For this model to be underidentified, we must be able to find an $\alpha^* \neq \alpha$, both distinct from zero, such that α^* also satisfies equation system (23). Since α and α^* are distinct and linear combinations of α and α^* must satisfy (23), it follows that

$$E\left[z_{i,t} {y^*_{i,t+1}}'\right] \gamma = 0 \tag{24}$$

for $t = 1, 2, ...$ where

$$y^*_{i,t+1} = \begin{bmatrix} v_{i,t+1} & v_{i,t} & ... & v_{i,t-\ell} \end{bmatrix}'$$

[13] Note that when $r = 2$, the model's parameters are not identified, but it is still possible to test the restriction (22) as a specification test of the model.

[14] In a pure time series setting, there is only one i, say $i = 1$ but T is large.

and γ is not degenerate and has k entries.

Conversely, suppose that moment conditions (24) are in fact satisfied. Notice that

$$E\left[z_{i,t}{y^*_{i,t+2}}'\right]\gamma = 0$$

because

$$E\left[z_{i,t+1}{y^*_{i,t+2}}'\right]\gamma = 0,$$

where this latter equation is just (24) shifted one time period forward. As a consequence, both

$$\begin{aligned}\alpha &= \begin{bmatrix}\gamma' & 0\end{bmatrix}' \\ \alpha^* &= \begin{bmatrix}0 & \gamma'\end{bmatrix}'.\end{aligned}$$

necessarily satisfy (23). Thus the Itest for underidentification naturally leads us to test an alternative set of moment conditions with one less free parameter given by (24). Identification of the parameter vector α from (23) up to scale requires that we reject moment equations (24) up to scale.

In a panel data setting, the I test is built from moment conditions (24) for $t = 1, 2, ..., T$ and large N. This construction of the I does not simply duplicate moments conditions, as this would lead to a degeneracy or repetition of moment conditions. Instead, the time series structure naturally leads to an alternative equation system to be studied. Also we could construct a collection of reduced form equations by projecting $y_{i,t+2}$ onto $z_{i,t}$ for each i and explore the restrictions imposed on coefficients. The reduced-form coefficients would necessarily be time dependent, and they would include some implicit redundancies. For this example, it is particularly convenient to work directly with the original structural equation system.

A concrete example of this estimation comes from Arellano and Bond (1991). They consider the estimation of a scalar autoregression with a fixed effect. In this example there is an underlying process $\{v_{i,t} : t = 0, 1, ..\}$. Form the scalar $\Delta v_{i,t} = v_{i,t} - v_{i,t-1}$ and construct $z_{i,t}$ to include $v_{i,0}, v_{i,1}, ..., v_{i,t}$. By taking first differences the fixed effect is eliminated from the estimation equation. When there is a unit root, this differencing reduces the order of the autoregression, but in general the order is not reduced. The I test checks whether in fact the order can be reduced.

We illustrate this using an AR(2) model for panel data with an individual specific intercept η_i:

$$\alpha_1(v_{i,t+2} - \eta_i) = -\alpha_2(v_{i,t+1} - \eta_i) - \alpha_3(v_{i,t} - \eta_i) + u_{i,t+2} \ (t = 3, ..., T), \tag{25}$$

and

$$E\left(u_{i,t}|v_{i,1}, ..., v_{i,t-1}; \eta_i\right) = 0. \tag{26}$$

Taking the first differences of equation (25) eliminates the fixed effect. Following Arellano and Bond (1991), consider GMM estimation of α_1 and α_2 based on a random sample $\{v_{i,1}, ..., v_{i,T} : i = 1, ..., N\}$ and the unconditional moment restriction:

$$E[z_{i,t}(\alpha_1 \Delta v_{i,t+2} + \alpha_2 \Delta v_{i,t+1} + \alpha_3 \Delta v_{i,t})] = 0 \ (t = 1, ..., T - 2). \tag{27}$$

Thus, we have a system of $T - 3$ equations with a set of admissible "instruments" that increases with T, but a common parameter vector α. With $T = 3$ there is a single equation in first differences with two instruments so that α is at best just identified up to scale. We may pin down the scale by letting the residual variance be zero or we could normalize the first coefficient to be unity, in which case the remaining coefficients are the negatives of the familiar autoregressive coefficients.

Returning to our original specification (25), suppose that $\alpha_1 + \alpha_2 + \alpha_3 = 0$. Then

$$\alpha_1(v_{i,t+2} - \eta_i) = -\alpha_2(v_{i,t+1} - \eta_i) - \alpha_3(v_{i,t} - \eta_i) + u_{i,t+2} \ (t = 3, ..., T),$$

Under this parameter restriction the fixed effect is inconsequential and can be dropped. Imposing this zero restriction allows us to rewrite the equation as:

$$\alpha_1 \Delta v_{i,t+2} = -(\alpha_2 + \alpha_1)\Delta v_{i,t+1} + v_{i,t+2}.$$

This first-order AR specification in first-differences is implicitly the specification that is used in building the I test. If this specification fails to satisfy the orthogonality restrictions, then the parameters of the original model cannot be identified using the approach of Arellano and Bond (1991). *Accepting* the hypothesis that underlies the I test is tantamount to assuming accepting an identified AR(2) specification with a unit root.

Up until now we have considered only models that are linear in the variables. We extend this discussion to include models with nonlinearities. In this discussion, it is important to

distinguish two cases. In the first case there is a separation between variables and parameters, and hence the nonlinearity is confined to the parameters. In the second case, the nonlinearities between variables and parameters interact in a more essential way.

4 Nonlinearity in the Parameters

In order to discuss lack of identification in non-linear models, it is important to carefully distinguish the different situations that may arise. We say that $\beta^* \neq \beta_0$ is observationally equivalent to β_0 if and only if $E[f(x_i;\beta^*)] = 0$. The true value β_0 is locally identifiable if there is no observationally equivalent value in a neighborhood of β_0, or more formally, if $E[f(x_i;\beta^j)] \neq 0$ for any sequence β^j such that $\lim_{j\to\infty}\beta^j = \beta_0$ (Fisher (1966)). Similarly, β_0 is globally identifiable if $E[f(x_i;\beta)] \neq 0$ for all $\beta \neq \beta_0$, that is, if there is no observationally equivalent structure anywhere in the admissible parameter space. The order condition $\dim(f) \geq \dim(\beta)$ provides a first check of identification, but this is only necessary. A complement is provided by the rank condition: If $D(\beta) = E\left[\partial f(x,\beta)/\partial\beta'\right]$ is continuous at β_0, and $rank[D(\beta_0)] = \dim(\beta)$, then β_0 is locally identified (Fisher (1966); Rothenberg (1971)).

In contrast to the order condition, this condition is only sufficient. But if $rank[D(\beta)]$ is also constant in a neighborhood of β_0, then the above rank condition becomes necessary too. In a linear model such that $f(x_i;\beta) = \Psi(x_i)\alpha$, where $\Psi(x_i)$ is $r \times \ell + 1$ and β results from α after normalization, the condition $rank\{E[\Psi(x_i)]\} = k$ is necessary and sufficient for both local and global identification. However, as argued in Sargan (1983b,a), there are non-linear models in which the rank condition fails, and yet β_0 is locally identified. In that case, β_0 is said to be first-order underidentified.

Another possibility that can only arise in non-linear models is a situation in which there are either a finite or a countably infinite number of isolated values of β which are observationally equivalent to β_0. In our analysis of nonlinear models we focus on situations where there is a continuum of observationally equivalent structures. We proceed by first showing how the analysis for linear models can be extended to decomposable nonlinear models of the form:

$$E(\Psi_i)\phi(\beta_0) = 0. \tag{28}$$

We extend our previous analysis by replacing the parameter vector α by a nonlinear,

continuously differentiable function $\phi : \mathbb{P} \to \mathbb{R}^{k+1}$ where $\mathbb{P}$ is the closure of an open set in $\mathbb{R}^{\ell}$. We study the nonlinear equation:

Assumption 4.1.

$$E\left(\Psi_i\right)\phi(\beta) = 0.$$

for some β in $\beta \in \mathbb{P}$.

The identification question is only of interest when ϕ is a one-to-one (i.e. injective) function. If there are two distinct parameter values β and β^* for which $\phi(\beta) = \phi(\beta^*)$ then we know *a priori* that we cannot tell β from β^* on the basis of Assumption 4.1. We make the stronger restriction

Assumption 4.2. *For any two values of the parameter vector $\beta \neq \beta^*$ in $\mathbb{P}$, $\phi(\beta) \neq \mathsf{c}\phi(\beta^*)$ for some real number c.*

We know that we can only identify $\phi(\beta)$ up to a proportionality factor. In Assumption 4.2 we ask the nonlinear parameterization to eliminate scale multiples from consideration.

Suppose now that two values $\hat{\beta}$ and $\check{\beta}$ satisfy Assumption 4.1 and are distinct. Thus both $\phi(\hat{\beta})$ and $\phi(\check{\beta})$ are in the null space of the matrix $E(\Psi_i)$. By Assumption 4.2, the vectors $\phi(\hat{\beta})$ and $\phi(\check{\beta})$ are not proportional, that is they are not in the same subspace. Any two linear combinations of $\phi(\hat{\beta})$ and $\phi(\check{\beta})$ must also be in the null space of $E(\Psi_i)$. We now find it fruitful to think of the function ϕ as imposing restrictions on a parameter vector α through the mapping $\phi(\beta) = \alpha$. By thinking of α as the parameter to be estimated, we can use aspects of the approach described previously. Since ϕ is one-to-one, we can uncover a unique β for each α. This leads us to construct the parameter space:

$$\mathbb{Q} \doteq \{\alpha : \alpha = \phi(\beta) \text{ for some } \beta \in \mathbb{P}\}.$$

To study underidentification using our previous approach, we expand the parameter space as follows:

$$\mathbb{Q}^* \doteq \{\alpha : \alpha = \mathsf{c}_1\alpha_1 + \mathsf{c}_2\alpha_2, \alpha_1 \in \mathbb{Q}, \alpha_2 \in \mathbb{Q}, \mathsf{c}_1, \mathsf{c}_2 \in \mathbb{R}\}.$$

Notice that if

$$E(\Psi_i)\alpha = 0$$

for two values of α in $\mathbb{Q}$, then there is a two-dimensional subspace of solutions to this equation in $\mathbb{Q}^*$. This problem is not just a special case of our earlier analysis because $\mathbb{Q}^*$ is not necessarily a linear space.

4.1 An illustrative example

To illustrate how nonlinearity in parameters can alter the analysis, we use an example that is closely related to the non-linear IV model with serially correlated errors considered by Sargan (1959). Nevertheless, it differs in an important way because in our case the valid instrumental variables are predetermined but not necessarily strictly exogenous.[15]

Example 4.1. *Consider a time series example:*

$$x_i \cdot \beta_1 = u_i + \gamma_1 \cdot w_i \tag{29}$$

$$u_i = \beta_2 u_{i-1} + \gamma_2 \cdot w_i. \tag{30}$$

where $\{w_i\}$ *is a multivariate martingale difference sequence. Suppose also that* z_{i-1} *is a linear function of* $w_{i-1}, w_{i-2}, ...$ *The process* $\{u_i\}$ *is unobservable to the econometrician, but*

$$x_i \cdot \beta_1 - \beta_2(x_{i-1} \cdot \beta_1) = (\gamma_1 + \gamma_2) \cdot w_i - \beta_2 \gamma_1 \cdot w_{i-1}.$$

Let

$$\Psi_i = \begin{bmatrix} z_{i-2}x_i' & -z_{i-2}x_{i-1}' \end{bmatrix},$$

and consider identification of β *based on:*

$$E(\Psi_i)\phi(\beta) = 0$$

where

$$\phi(\beta) = \begin{bmatrix} \beta_1 \\ \beta_2\beta_1 \end{bmatrix}. \tag{31}$$

To achieve identification requires that we impose an additional normalization, say $|\beta_1| = 1$. We may wish to restrict $|\beta_2| < 1$. Since have not restricted $\gamma_2 \cdot w_i$ to be uncorrelated with

[15] In his Presidential address to the Econometric Society Sargan (1983a) studied a static model with the same mathematical structure, while Sargan (1983b) analyzed a dynamic multivariate version.

u_{t-1}, the unobserved (to the econometrician) process $\{u_i\}$ can be stationary and still satisfy equation (30). Thus when $|\beta_2| > 1$,

$$u_i = -\sum_{j=1}^{\infty} (\beta_2)^{-j} w_{i+j}$$

is a stationary process that satisfies (30). Notice, however, in this case $u_i + \gamma_1 \cdot w_i$ is orthogonal to z_{i-1} so there is an additional moment restriction at our disposal. As is well known the case of $|\beta_2| = 1$ requires special treatment.

Consider two parameter choices (β_1, β_2) and (β_1^*, β_2^*). Without loss of generality write

$$\beta_1^* = \mathsf{c}\beta_1 + \mathsf{d}\eta_1 \tag{32}$$

where $\mathsf{c} = \beta_1 \cdot \beta_1^*$, $|\eta_1| = 1$ and $\eta_1 \perp \beta_1$, and impose that $\mathsf{c}^2 + \mathsf{d}^2 = 1$ to guarantee that $|\beta_1^*| = 1$ too.

In line with the linear case assume that $rank[E(\Psi_i)] = k - 1$ so that its nullity is 2. This means that if there are other observationally equivalent structures, they must satisfy

$$E(\Psi_i) \begin{bmatrix} \mathsf{c}\beta_1 + \mathsf{d}\eta_1 \\ \mathsf{c}\beta_2^*\beta_1 + \mathsf{d}\beta_2^*\eta_1 \end{bmatrix} = 0 \tag{33}$$

Given the partly linear, partly non-linear structure of the model only the following two locally underidentified situations may arise.

4.1.1 Only β_1 identified

There is one *arguably special* way in which identification can break down. Suppose that

$$E(z_{i-2}x'_{i-1})\beta_1 = 0,$$

and hence

$$E(z_{i-1}x'_i)\beta_1 = 0 \tag{34}$$

for some β_1. This phenomenon can occur for one of two reasons. First perhaps the choice z_{i-2} is unfortunate. Alternatively, $x_i \cdot \beta_1$ may depend only on current and possibly future values of the martingale difference sequence $\{w_i\}$. As we have seen, this may happen when $|\beta_2| > 1$ or in the degenerate case when u_i is identically zero ($\gamma_1 = 0$).[16]

[16] In the case in which $|\beta_2| > 1$ we may identify β_2 from other moment conditions.

Note that this equivalent to choosing $\mathsf{d} = 0$ and $\beta_2 \neq \beta_2^*$ in (33). For this same β_1, it is also required that

$$E(z_{i-2}x'_{i-1})\beta_1 = 0$$

Typically, there will be common entries in z_{i-1} and z_{i-2}. Let z^*_{i-1} be a random vector formed after eliminating these redundancies in order that $E\left[z^*_{i-1}z^{*\prime}_{i-1}\right]$ is nonsingular. Then the I test for β_2 is based on:

$$E(z^*_{i-1}x'_i)\beta_1 = 0.$$

In other words, if the composite disturbance term $u_i + \gamma_1 \cdot w_i$ is orthogonal to z^*_{i-1}, then β_2 is not identified via the moment conditions. This I test is implemented by estimating the econometric relationship without quasi-differencing, and then testing the resulting overidentifying restrictions. Of course, if the null hypothesis underlying the I test is accepted, there are other moment conditions that could be used to identify β_2 given β_1.

Notice in this case there is a continuum of values of the composite parameter vector β that satisfy the moment conditions under the null hypothesis of the I test, but only a single value of β_1, which our procedure will estimate efficiently.

This test is closely related but not identical to the underidentification test proposed by Sargan (1959) for the non-linear in parameters model that he studied. The augmented set of moment conditions that he considered were (34) and

$$E(\Psi_i)\begin{bmatrix}\beta_1^* \\ \beta_2^*\beta_1^*\end{bmatrix} = 0,$$

where he implicitly chose β_2^* so that the sample covariance matrix of $x'_i\beta_1$ and $(x'_i - \beta_2^* x'_{i-1})\beta_1^*$ were 0. Apart from our modern emphasis on symmetric normalization and robustness to serial correlation and heteroskedasticity, the main difference with his approach is that we impose the restriction $\beta_1 = \beta_1^*$, which, in parallel with a gain in estimation efficiency, leads to a reduction in the number of degrees of freedom and the resulting gain in power, and also eliminates the need to choose two arbitrary values for β_2.

As we mentioned previously, we could allow for the value of β_2 to have an absolute value greater than one. In this case identification of β_2 will fail unless we replace z_{i-2} by z_{i-1}.

4.1.2 Only β_2 is identified

Suppose now there is a vector $\beta_1^* \neq \beta_1$ such that

$$\alpha^* = \begin{bmatrix} \beta_1^* \\ \beta_2 \beta_1^* \end{bmatrix}$$

satisfies the moment conditions:

$$E\left(\Psi_i\right) \alpha^* = 0.$$

Since any linear combination of α and α^* must satisfy moment conditions, we can choose $\mathsf{c} = 0$ in (32) so that

$$\begin{bmatrix} \eta_1 \\ \beta_2 \eta_1 \end{bmatrix}$$

should also satisfy the moment conditions (33). This gives rise to a second I test. We parameterize two orthonormal directions η_1 and β_1 along with a *single* parameter β_2. When β_1 has only two components, we are free to set β_1 and η_1 equal to the two coordinate vectors and freely estimate only the parameter β_2. In that case the moment conditions of the I test can be expressed as

$$E[z_{i-2}(y_i - \beta_2 y_{i-1})] = 0, \ i = 1, 2$$

More generally, under the null hypothesis associated with this I test there is a two-dimensional plane of (non-normalized) values of the original parameter vector β_1 that satisfy the moment conditions, but only value of β_2. After normalisation, the manifold of observationally equivalent structures will be given by (32) with $\mathsf{c}^2 + \mathsf{d}^2 = 1$. In this sense, note that if $E[z_{i-2}(y_i - \beta_2 y_{i-1})] = 0$ for some i, then all the β_1 coefficients will be identified except the one corresponding to y_i.

Importantly, this test is different from a linear test of $rank[E(\Psi_i)] = k - 1$ derived along the lines of section 3.1, since such a test would not impose that the observationally equivalent structures must satisfy (31).

Once again, as a by-product of our procedure we will obtain efficient GMM estimators of β_2, and the parameters β_1 and η_1 that characterize the identifed set through (32).

5 Fundamental nonlinearity

We now explore the underidentification problem when there is a more fundamental nonlinearity of the parameters in the moment conditions. Recall that in the linear model discussed

in section 3, underidentification implies that we can estimate a line, which we chose to implicitly parameterize by means of two parameter vectors. Similarly, in the non-linear in parameters model discussed in section 4, we also implicitly parameterize a curve as a function of a finite number of parameters. The natural extension for a fully nonlinear model is to estimate a one-dimensional curve. As in the linear and non-linear in parameters models, joint estimation of the curve implies improvements in statistical efficiency. The tools developed by Hansen (1982, 1985) and Carrasco and Florens (2000) can be extended to this application.

As in section 2, we pose the inferential problem as one in which a function, π of a scalar θ is estimated. We restrict θ to be in a compact interval Θ. For each value of θ, a hypothetical parameter vector, say $\pi(\theta)$, satisfies the population moment conditions. Such a function could feature θ as the first entry of the parameter vector so the first coordinate of π is the identity function, as in (2), but this is only one possibility. Associated with the function π is a curve

$$\mathcal{C} = \{\pi(\theta) : \theta \in \Theta\}$$

in the parameter space.[17]

5.1 Estimation Environment

Conveniently, this estimation problem looks like a standard problem except that we seek to estimate a function instead of a finite dimensional parameter vector. Suppose that $\{x_t\}$ is a stationary and ergodic stochastic process (but see footnote 1).

Assumption 5.1. *Let $\mathbb{P}$ be a compact subset of $\mathbb{R}^k$.*

Introduce a function $f(x, \cdot) : \mathbb{P} \to \mathbb{R}^p$ for each x. The function f is jointly Borel measurable and at the very least continuous in its second argument for each value of x. Thus $f(x_t, \cdot)$ is a p-dimensional random function on $\mathbb{P}$ or a random element.

Assumption 5.2. *$E|f(x_t, \beta)| < \infty$ for each $\beta \in \mathbb{P}$.*

In light of this assumption we define $Ef(x_t, \beta) = \bar{f}(\beta)$ for each $\beta \in \mathbb{P}$.
As in Hansen (1982), we also assume:

[17] Underidentification of a higher dimension arises when θ is a vector instead of a scalar, as in section 3.1.2.

Assumption 5.3. *$f(x_t, \cdot)$ is first-moment continuous for each $\beta \in \mathbb{P}$.*

Under this assumption $\bar{f}$ is continuous in β. This continuity condition along with a point-wise (in β) Law of Large Numbers implied by ergodicity gives a Uniform Law of Large Numbers (see Hansen (1982)).

We are interested in extending the usual GMM estimation framework by considering parameterizations of the form $\pi(\theta)$, where π is a continuous function with range $\mathbb{P}$ and $\theta \in \Theta$.

Assumption 5.4. *Π is a compact set of admissible functions defined using the supnorm.*

From the Arzelà–Ascoli Theorem it suffices that there be uniform bound on the functions in Π and that the functions be equicontinuous. The uniform bound comes from the compactness of $\mathbb{P}$ (Assumption 5.1).

Consider next first-moment continuity. Notice that

$$|f[x, \pi(\theta)] - f[x, \tilde{\pi}(\theta)]| \leq \sup_{\beta \in \mathbb{P}, \tilde{\beta} \in \mathbb{P}, |\beta - \tilde{\beta}| \leq \epsilon} |f(x, \beta) - f(x, \tilde{\beta})|$$

provided that $\|\pi - \tilde{\pi}\| \leq \epsilon$. This simple inequality implies that the first-moment continuity restriction given in Assumption 5.3 extends to the parameter space Π.

Given that we now seek to identify a function π_0 instead of a vector β_0, under the null hypothesis of underidentification of β our new "identification condition" requires that:

Assumption 5.5. *$\bar{f}[\pi(\theta)] = 0$ for all $\theta \in \Theta$ if, and only if $\pi = \pi_0$.*

This assumption rules out constant functions in the sup-norm closure of the set Π. More generally, it rules out the possibility that there exists $\tilde{\pi}$ such that

$$\{\tilde{\pi}(\theta) | \theta \in \Theta\} \subset \{\pi_0(\theta) | \theta \in \Theta\} \tag{35}$$

for some $\tilde{\pi} \neq \pi_0$, in which case there would exist two functions in this closure for which the image of one function is a proper subset of the other. Note that (35) is ruled out *a priori* if we use parameterization (2) given by:

$$\pi(\theta) = \begin{bmatrix} \theta \\ \tau(\theta) \end{bmatrix} \tag{36}$$

because the first coordinate of π is allowed to vary.

A consistent estimator can be obtained for π_0 in the usual fashion except that our parameter is now a function. In other words, consistency is a straightforward extension. It turns out, however, that statistical efficiency is altered in a more fundamental way.

As we previously observed in expression (4) of section 2, the matrix

$$E\left[\frac{\partial f[x_t, \pi_0(\theta)]}{\partial \beta}\right]$$

is of rank $k-1$ for all values of θ in the interior of Θ when $\bar{f}$ is continuously differentiable in β and π_0 is differentiable in θ.[18] This rank failure confirms the underidentification of β over a range of values of θ and hence along a one-dimensional curve in the parameter space. Part of our econometric challenge is to make inferences about the function π_0 used to parameterize this curve. This leads to explore efficient estimation of π_0 using some limit theory approximations:

Assumption 5.6. $\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\left[f(x_t,\cdot)-\bar{f}\right]$ *converges to a Gaussian random element with covariance function:*

$$K(\alpha,\beta)=\lim_{N\to\infty}\frac{1}{N}E\left(\sum_{t=1}^{N}[f(x_t,\alpha)-\bar{f}(\alpha)]\sum_{s=1}^{N}[f(x_s,\beta)-\bar{f}(\beta)]'\right).$$

This requires a functional version of a central limit theorem, but it is well understood how to justify this restriction.

5.2 Ignoring Efficiency Gains for Joint Estimation

In what follows we use parameterization (36). For each value of θ we estimate the $k-1$-dimensional parameter vector $\tau(\theta)$. In this section we explore the efficient estimation of $\tau(\theta)$ for each choice of θ as a separate estimation problem. As a consequence, we may apply directly the analysis in Hansen (1982) and the earlier analysis in Sargan (1958, 1959), which involves reducing the moment conditions by introducing a $k-1$ by p matrix A for every choice of θ.

Specifically, construct

$$D(\theta)=E\left[\frac{\partial f[x_t,\pi_0(\theta)]}{\partial\beta}\right]\begin{bmatrix}0_{k-1}\\ I_{k-1}\end{bmatrix}$$

[18] Wright (2003) focuses on testing this restriction at a single value in the parameter space under the assumption that the true parameter is locally identified but not first-order identified.

where 0_{k-1} is a row vector of zeros and I_{k-1} is an identity matrix of dimension $k-1$. If we imitate standard GMM optimization, we conclude that a "point-wise efficient" selection matrix is:

$$A(\theta) = D(\theta)' \left(K[\pi_0(\theta), \pi_0(\theta)]\right)^{-1}. \tag{37}$$

By premultiplying the selection matrix by a nonsingular matrix, possibly distinct for each value of θ, we preserve statistical efficiency. This "point-wise efficient" selection can be implemented by choosing a weighting matrix

$$\left(K[\pi_0(\theta), \pi_0(\theta)]\right)^{-1}$$

for each θ.

Notice since the covariance matrix $K[\pi_0(\theta), \pi_0(\theta)]$ depends on θ, point-wise efficiency cannot be achieved by any fixed weighting matrix, as typically used in a quadratic form minimization. Thus in this environment, optimally selecting the weighting matrix, and hence the selection matrix by choice of θ improves (first-order) asymptotic efficiency.[19]

While this one-value-at-time approach improves efficiency, it ignores correlation across the values of θ. As a consequence, further improvement is possible as was illustrated in Example 2.1.

5.3 Efficiency

To characterize further efficiency gains, we need to construct a covariance operator and its inverse. Our development will be informal in places, and we defer to subsequent research a more rigorous analysis.

Following Carrasco and Florens (2000) we construct a covariance operator $\mathcal{K}$ as:

$$\mathcal{K}\phi(\theta) \doteq \int K[\pi_0(\theta), \pi_0(\vartheta)]\phi(\vartheta)d\vartheta$$

where ϕ maps Θ into $\mathbb{R}^p$. The coordinate functions of ϕ are restricted to be in an L^2 space defined using a conveniently chosen measure on $[0,1]$. We denote this space L^2_p. For notational simplicity we use Lebesgue measure, but in some applications other measures may

[19] Set inference methods in GMM settings typically use a fixed weighting matrix and ignore these efficiency improvements.

turn out to be more convenient. We take the operator $\mathcal{K}$ to be bounded. Notice that

$$\int \phi^* \cdot [\mathcal{K}(\phi)] = \int \phi \cdot [\mathcal{K}(\phi^*)].$$

Thus $\mathcal{K}$ is a self-adjoint operator.

When the kernel K is continuous on $\Theta \times \Theta$, the operator $\mathcal{K}$ is necessarily bounded and it has a discrete spectrum.[20] In this case we may represent the covariance operator as:

$$\mathcal{K}\phi = \sum_{j=1}^{\infty} \lambda_j \phi_j \int \phi \cdot \phi_j$$

where the $\{\lambda_j : j = 1, ...\}$ are decreasing and

$$\sum_{j=0}^{\infty} \lambda_j^2 < \infty,$$

and the $\{\phi_j : j = 1, 2, ...\}$ are orthonormal and complete. We consider two other operators that are constructed from $\mathcal{K}$. The square root operator is:

$$\mathcal{K}^{1/2}\phi = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \phi_j \int \phi \cdot \phi_j,$$

and its inverse:

$$\mathcal{K}^{-1/2}\phi = \sum_{j=1}^{\infty} \frac{1}{\sqrt{\lambda_j}} \phi_j \int \phi \cdot \phi_j.$$

In fact the operator $\mathcal{K}^{1/2}$ generally fails to have an inverse on all of L_p^2 because either there may be only a finite number of nonzero eigenvalues or there may be an infinite number of eigenvalues in which case the λ_j converge to zero. We focus on the latter case.

For a given ϕ^*, we construct a solution ϕ to the equation:

$$\mathcal{K}^{1/2}\phi = \phi^*$$

by forming:

$$\phi = \sum_{j=1}^{\infty} \frac{< \phi^*, \phi_j >}{\sqrt{\lambda_j}} \phi_j$$

which is well-defined when

$$\sum_{j=1}^{\infty} \frac{< \phi^*, \phi_j >^2}{(\lambda_j)} < \infty.$$

[20] In fact $\mathcal{K}$ is a Hilbert-Schmidt operator.

With this in mind, let

$$H_p^2 = \left\{ \phi : \sum_{j=1}^{\infty} \frac{< \phi_j, \phi >^2}{\lambda_j} < \infty \right\}.$$

In our analysis of efficiency, we assume that $\bar{f}[\pi_0(\cdot)]$ is in H_p^2.

The counterpart to an efficient selection matrix of the form given in Sargan (1958, 1959), and Hansen (1982, 1985) is:

$$D(\theta)' \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \eta_N} \phi_j(\theta) \int_0^1 \phi_j(\vartheta) \cdot \frac{1}{N} \sum_{t=1}^{N} f[x_t, \hat{\pi}_N(\vartheta)] d\vartheta = 0, \tag{38}$$

where η_N is a regularization parameter that decays to zero with the sample size. We include the *regularization parameter* because of the error that is present in approximating the function $\bar{f}(\beta)$ by $\frac{1}{N} \sum_{t=1}^{N} f(x_t, \beta)$. The estimated function $\hat{\pi}_N(\theta)$, defined as the solution to (38) for all θ, is infeasible because λ_j and ϕ_j are population quantities. This selection can be implemented by solving

$$\min_{\pi \in \Pi} \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \eta_N} \left(\int_0^1 \phi_j(\theta) \cdot \frac{1}{N} \sum_{t=1}^{N} f[x_t, \pi(\theta)] d\theta \right)^2. \tag{39}$$

The objective function (39) is the operator counterpart to the quadratic minimization problem often used in GMM estimation. More specifically, it is a version of the continuum of moment condition objective function of Carrasco and Florens (2000) extended to the estimation of the (infinite-dimensional) parameter vector π. Carrasco and Florens (2000) provide justification and discuss implementation for their related GMM setting.

In terms of efficiency it is most convenient to work with the counterpart to an information matrix. Thus we construct an information operator $\mathcal{J}$. To construct the operator domain, let ψ map Θ into $\mathbb{R}^{k-1}$, and let

$$\tilde{H}_{k-1}^2 = \{ \psi : D\psi \in H_p^2 \}.$$

Then

$$\mathcal{J}\psi(\theta) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} D'(\theta) \phi_j(\theta) \left[\int_{\Theta} \phi_j \cdot (D\psi) \right].$$

Notice that

$$\begin{aligned}\int_{\Theta} \psi \cdot (\mathcal{J}\psi) &= \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \left[\int_{\Theta} \phi_j \cdot (D\psi)\right]^2 \\ &\geq \frac{1}{\lambda_1} \sum_{j=1}^{\infty} \left[\int_{\Theta} \phi_j \cdot (D\psi)\right]^2 \\ &= \frac{1}{\lambda_1} \int_{\Theta} (D\psi) \cdot (D\psi)\end{aligned}$$

since eigenfunctions are a complete orthonormal sequence of functions in the space. While the information operator $\mathcal{J}$ is defined on the restricted domain $\tilde{H}^2_{k-1}$, its inverse, which we denote $\mathcal{C}$, can be extended to a larger domain of functions $\tilde{L}^2_{k-1}$ consisting of ψ such that $D\psi \in L^2_p$. Our conjectured and (hopefully sharp) efficiency bound for integral averages of the parameter estimator

$$\sqrt{N} \int_0^1 \psi(\theta) \cdot [\hat{\tau}_N(\theta) - \tau_0(\theta)] d\theta$$

is given by

$$\int_{\Pi} \psi \cdot (\mathcal{C}\psi)).$$

We suspect that this bound can be extended to a broader class of linear functionals of the parameter π including functions that evaluate π at individual points, however, this will be left for subsequent research.

It is interesting to relate the inferential problems in the previous sections with the one in this section. The main difference is that the linearity of (5) and (28) implies that the resulting operator $\mathcal{K}$ would only have a finite number of positive eigenvalues. Once we take this fact into account, though, the curves that we will estimate with the procedure that we have developed in this section will coincide with the curves that we implicitly estimated using the procedures developed in sections 3 and 4.

To see why, consider for instance the linear in parameters model (5), and suppose that instead of (6) we seek to estimate a non-linear parametric curve with the following structure

$$\pi(\theta) = \theta \cdot \alpha^{[1]} + (1 - \theta) \cdot \alpha^{[2]} + \begin{bmatrix} \theta \\ \upsilon(\theta) \end{bmatrix} \tag{40}$$

Further, assume that $\pi(\theta)$ can be uniquely identified from the continuum of moment conditions (7). We know that for each possible υ the linear span of the image will be finite-dimensional. As we show in appendix B, the method proposed in this section will select

$\upsilon(\theta) = 0\ \forall\theta$ in order to keep the dimension of the linear span as small as possible, in this case two.

5.4 Testing

Suppose that $\pi_0(\theta)$ is a known function of θ, say $\pi_0(\theta) = \theta$. Under full identification there is a unique but unknown parameter vector, given by say $\beta_0 = \pi_0(\theta_0)$, but we wish to test for underidentification by pre-specifying π_0 but not θ_0. By assumption, estimation of π_0 is unnecessary. This is a special case of our analysis, but it is also a special case of the analysis of Carrasco and Florens (2000). While estimation has been pushed aside, the "overidentification" test of Carrasco and Florens (2000) is directly applicable to this problem as a test of underidentification.

More generally, an overidentification test could be constructed analogously to that of Carrasco and Florens (2000) by scaling appropriately the minimized sample counterpart to (39). The resulting test could produce a normal distribution as the limit of a sequence of appropriately scaled (approximate) chi-square distributions with an arbitrarily large number of degrees of freedom. An alternative approximation that incorporates the role of regularization leads instead to an approximate quadratic form in normal variables. Our experience suggests that such an Imhof (1961)-style approximation becomes an attractive alternative to the limiting normal distribution (see Appendix C for a further discussion of this point).

To illustrate the previous discussion, we use an asset pricing example from Hansen and Singleton (1982), which is closely related to Example 3.3.[21] For simplicity, we consider the case of a single asset.[22]

Example 5.1. *Consider:*

$$E\left\{z_t\left[y_{1t}\exp(-\varphi y_{2t}) - exp(-\rho)\right]\right\} = 0, \tag{41}$$

where y_{1t} denotes the gross return on some risky financial asset over period t, y_{2t} denotes the continuously compounded rate of growth in consumption of a representative agent with

[21] This example has also been explored in Hansen et al. (1996), Stock and Wright (2000) and Kleibergen (2005) in their analysis of continuously-updated GMM and weak identification.

[22] Multiple asset versions of this example typically lead to substantial empirical evidence against the asset pricing model. Similarly, empirical analyses with Treasury bill data also reject the model. In contrast, aggregate data on equities provides only weak evidence about the parameters of interest. See for instance, section 3 and Figure 1 of Hansen and Singleton (1996) for further discussion of these issues.

time-separable expected utility preferences of the isoelastic variety over the same period, z_t is a vector of $p \geq 2$ instrumental variables known in period $t-1$, φ is the reciprocal of the elasticity of intertemporal substitution and ρ is the rate of time preference.

If the joint distribution of y_{1t} and y_{2t} conditional on z_t does not depend on z_t,[23] then (41) will be compatible with any pair of values of φ and ρ that satisfy the following moment condition

$$E\left[y_{1t}\exp(-\varphi y_{2t}) - \exp(-\rho)\right] = 0. \tag{42}$$

Equation (42) can be understood as defining a curve in φ, ρ space that is the locus of all the structures that are observationally equivalent to the true one. In particular we can trace out values for the subjective rate of time preference for a range of values of the elasticity of intertemporal substitution. Under specific distributional assumptions, such as joint normality and homoskedasticity of $logy_{1t}$ and y_{2t} conditional on z_t, it is possible to obtain a parametric expression for such a locus. The approach that we have developed in this section makes such parametric assumptions unnecessary. All we need to do to test for the identifiability of model (41) is to apply our procedures to:

$$E\left\{z_t\left[y_{1t}\exp(-\theta y_{2t}) - \exp(-\tau(\theta))\right]\right\} = 0.$$

6 Conclusions

In instrumental variables or generalized-method-of-moments estimation of an econometric model it is useful to have a statistical test designed to ascertain whether the model is underidentified. Indeed Koopmans and Hood (1953) (see page 184) wrote:

> "It is ... natural to abandon without further computation the set of restrictions strongly rejected by the (likelihood ratio) test. Similarly, it is natural to apply a test of identifiability before proceeding with the computation of the sampling variance of estimates ... and to forego any use of the estimates, if the indication of nonidentifiability is strong."

[23] Note that the independence assumption of the joint distribution of y_{1t} and y_{2t} given z_t is much stronger than the equivalent condition in the linear case discussed in example 3.3. On the other hand, conditional independence is only a sufficient condition for the failure of identification in this nonlinear model.

While it was recognized in the early econometric literature on simultaneous equations systems that underidentification is testable, to date such tests are uncommon in econometric practice. Nevertheless, many econometric models of interest often imply a large number of moment restrictions relative to the number of unknown parameters and are therefore seemingly overidentified. However, this situation is often coupled with informal evidence that identification may be at fault. In those cases, an identification test in conjunction with some specificity about the potential nature of the identification failure will help to assess to what extent the sample is informative about the parameters of interest.

In this paper we proposed a method for constructing tests of underidentification based on the structural form of the equation system. We regard underidentification as a set of overidentifying restrictions imposed on an augmented structural model. Therefore, our proposal is to test for underidentification by testing for overidentification in the augmented model using either standard overidentifying testing methods available in the literature, or some generalizations developed in this paper. A by-product of our analysis is an estimate of a direction or a curve that shows the parameter-tradeoffs that have comparable empirical implications.

Our idea for how to build a test or underidentification is straightforward: estimate a curve instead of a point and test the resulting overidentification. If it is possible to construct such a curve without statistical rejection, then the original model is likely to be underidentified. But if the attempt fails statistically, then the null hypothesis is rejected and we may conclude the model is identified.

We show that our approach can be used not only for single equation linear models, but also for systems with cross-equation restrictions, possibly with different valid instruments for different equations. We also extend our methods to models which are non-linear in the parameters, as well as to fundamentally non-linear models in which there is a one-dimensional manifold of observationally equivalent structures.

In summary, the approach we develop in the paper for linear and nonlinear models has the following characteristics in common:

1) We use the structural specification and exploit the fact that if β_0 is not identified, then there will be a manifold of $\beta's$ that will satisfy the original moment conditions.

2) We implicitly parameterize this underidentified manifold, and write all the implied moment conditions as an extended system with either a finite or a continuum of moment conditions.

3) Then we simply compute the overidentification test of the extended system.

4) As a by-product, we obtain an "efficient estimator" of the underidentified manifold.

We do not provide an omnibus underidentification test, but a general approach to test for underidentification in situations in which the characteristics of the identified set of interest are either theoretically or empirically motivated.

Although we posed the target of the estimation to be a function π_0, it would perhaps more natural to pose the target to be a curve in the parameter space $\mathbb{P}$ and to develop estimation methods that are invariant to how the curve is parameterized. We hope to address this point and to provide some additional formality in subsequent research. In their study of observable factor models, Nagel and Singleton (2009) show that taking account of the conditioning information in an efficient way substantially alters the assessment of competing linear asset pricing models. Thus another important topic for future research is to incorporate conditional moment restrictions and to explore more generally the extent to which underidentification remains an important concern in practice.

Appendices

A The Cragg and Donald test of underidentification

Cragg and Donald (1993) considered single equation tests of underidentification based on the reduced form. Let us partition y_i into a $(p+1)$- and a r_1-dimensional vectors of endogenous and predetermined variables, respectively, $y_i = (y_{1i}', z_{1i}')'$, so that $k = p+r_1$ and $z_i = (z_{1i}', z_{2i}')'$, where z_{2i} is the vector of r_2 instruments excluded from the equation. Moreover, let Π and $\widehat{\Pi} = Y_1'Z(Z'Z)^{-1}$ be the $(p+1) \times r$ matrices of population and sample reduced form linear-projection coefficients, respectively. With this notation and the partition $\Pi = (\Pi_1, \Pi_2)$ conforming to that of z_i, α is identified up to scale if and only if the rank of Π_2 is p, but it is underidentified if the rank is $p-1$ or less.

To test for underidentification Cragg and Donald considered the minimizer of the minimum distance criterion

$$T[vec(\widehat{\Pi} - \Pi)]'V^{-1}vec(\widehat{\Pi} - \Pi) \tag{43}$$

subject to the restriction that the rank of Π_2 is $p-1$. Under the null of lack of identification and standard regularity conditions, this provides a minimum chi-square statistic with $2(r-k)+2$ degrees of freedom, as long as V is a consistent estimate of the asymptotic variance of $vec(\widehat{\Pi})$.

If the rank of Π_2 is $p-1$, there are two linearly independent vectors, denoted by Γ, such that $\Pi_2'\Gamma = 0$. For some ordering of the rows of Π_2, we can normalize Γ as $\Gamma' = (I_2, \Gamma_2')$. Partitioning Π_2 accordingly as $\Pi_2' = (\Pi_{21}', \Pi_{22}')$, we then have that $\Pi_{21}' = -\Pi_{22}'\Gamma_2$. To enforce the rank restriction, Cragg and Donald considered Π as a function of Π_1, Π_{22} and Γ_2.

To relate (43) to our framework, write the augmented model

$$\begin{aligned} y_i \cdot \alpha &= u_i \\ y_i \cdot \alpha^* &= v_i \end{aligned}$$

as a complete system by adding to it $p-1$ reduced form equations, and denote it by

$$By_{1i} + Cz_i = u_i^\dagger$$

where $B = (B_1', B_2')'$, $C = (C_1', C_2')'$, $B_2 = \left(0_{p-1,2} \vdots I_{p-1}\right)$, and $(B_1, C_1) = A'$. To visualize the mapping between the structural parameters and the Cragg-Donald parameterization of

the rank restriction, let us introduce the partitions $C_1 = (C_{11}, 0)$, $C_2 = (C_{21}, C_{22})$ and $B_1 = (B_{11}, B_{12})$. We then have that $\Pi_{22} = -C_{22}$ and $\Pi_{21} = B_{11}^{-1} B_{12} C_{22}$, so that $\Gamma_2 = -B_{11}^{-1} B_{12}$. Π_1 is unrestricted with $-B_{11}^{-1}(C_{11} - B_{12}C_{21})$ as the first component and $-C_{21}$ as the second.

Then noting that

$$\widehat{\Pi} - \Pi(A, C_2) = [Y_1' - \Pi(A, C_2) Z'] Z(Z'Z)^{-1} = \left(Y_1' + B^{-1} C Z'\right) Z(Z'Z)^{-1} = B^{-1} U^{\dagger\prime} Z(Z'Z)^{-1},$$

so that

$$vec(\widehat{\Pi} - \Pi) = (B \otimes Z'Z)^{-1} \sum_{i=1}^{N} (u_i^{\dagger} \otimes z_i),$$

(43) can be expressed as

$$\sum_{i=1}^{N} (u_i^{\dagger} \otimes z_i)' [(B \otimes Z'Z) V (B' \otimes Z'Z)]^{-1} \sum_{i=1}^{N} (u_i^{\dagger} \otimes z_i), \tag{44}$$

which is in the form of a continuously updated GMM criterion that depends on (α, α^*) and the coefficients C_2 in the additional $p-1$ reduced form equations. Since B does not depend on the latter, those parameters can be easily concentrated out of the criterion. A convenient feature of this criterion is that it is invariant to normalization through the updating of B while V is kept fixed.

Specifically, using a standard result on the irrelevance of unrestricted moments Arellano (2003) (see pages 196–197), criterion (44) concentrated with respect to C_2 can be shown to equal:

$$(\alpha' Y'Z, \alpha^{*\prime} Y'Z) [(B_1 \otimes Z'Z) V (B_1' \otimes Z'Z)]^{-1} \begin{pmatrix} Z'Y\alpha \\ Z'Y\alpha^* \end{pmatrix}.$$

An optimal weight matrix under classical errors is $V = Y_1' M Y_1 \otimes (Z'Z)^{-1}$, where $M = I - Z(Z'Z)^{-1} Z'$, in which case the concentrated criterion boils down to

$$(\alpha' Y'Z, \alpha^{*\prime} Y'Z) (A' Y'MYA \otimes Z'Z)^{-1} \begin{pmatrix} Z'Y\alpha \\ Z'Y\alpha^* \end{pmatrix}. \tag{45}$$

Its minimizer subject to $A' Y'MYA = I$ coincides with the sum of the two smallest characteristic roots of $Y'Z(Z'Z)^{-1} Z'Y$ in the metric of $Y'MY$, which is one of the (non-robust) test statistics discussed by Cragg and Donald.

Next, an optimal weight matrix under heteroskedastic errors is

$$V = (I \otimes Z'Z)^{-1} \sum_i (\widehat{\varepsilon}_i \widehat{\varepsilon}_i' \otimes z_i z_i') (I \otimes Z'Z)^{-1}$$

where $\widehat{\varepsilon}_i$ is a reduced-form residual (the i-th column of $Y_1'M$). In this case the concentrated criterion becomes

$$(\alpha' Y' Z, \alpha^{*\prime} Y' Z) \left(\sum_i A' \widetilde{y}_i \widetilde{y}_i' A \otimes z_i z_i' \right)^{-1} \begin{pmatrix} Z' Y \alpha \\ Z' Y \alpha^* \end{pmatrix} \tag{46}$$

where $\widetilde{y}_i$ denotes the i-th column of $Y'M$, so that the values of components of $\widetilde{y}_i$ that correspond to predetermined explanatory variables are identically zero.

To conclude, both robust and non-robust Cragg-Donald criteria can be regarded as continuously-updated GMM criteria of the augmented structural model using $\widetilde{y}_i' A$ as errors. Since the difference between $A' y_i$ and $A' \widetilde{y}_i$ at the truth is of small order, using one or the other is asymptotically irrelevant. Similar remarks can be made for optimal weight matrices under autocorrelated errors.

B Estimating Finite-Dimensional Specifications of π

We begin by considering a general GMM estimation result, which will prove useful for our purposes. Suppose the moment conditions used in GMM estimation can be partitioned as

$$f(x_t, \beta) = \begin{bmatrix} f^{[1]}\left(x_t, \beta^{[1]}\right) \\ f^{[2]}\left(x_t, \beta^{[1]}, \beta^{[2]}\right) \end{bmatrix}.$$

Let

$$g_N(\beta) = \frac{1}{N} \sum_{t=1}^{N} f(x_t, \beta) = \begin{bmatrix} \frac{1}{N} \sum_{t=1}^{N} f^{[1]}\left(x_t, \beta^{[1]}\right) \\ \frac{1}{N} \sum_{t=1}^{N} f^{[2]}\left(x_t, \beta^{[1]}, \beta^{[2]}\right) \end{bmatrix}.$$

Let $V_N(\beta)$ be the asymptotic covariance estimator used in a continuously-weighted GMM estimation, whose partition we denote by:

$$V_N(\beta) = \begin{bmatrix} V_N^{[11]}\left(\beta^{[1]}\right) & V_N^{[12]}(\beta) \\ V_N^{[21]}(\beta) & V_N^{[22]}(\beta) \end{bmatrix}.$$

We compare GMM objectives for estimating $\beta_0^{[1]}$ alone using the first set of moment conditions versus estimating the entire vector β_0 using the full set of moment conditions.

Lemma B.1.

$$\min_{\beta \in \mathbb{P}} g_N(\beta)' [V_N(\beta)]^{-1} g_N(\beta) \geq \min_{\beta \in \mathbb{P}} g_N^{[1]}\left(\beta^{[1]}\right)' \left[V_N^{[11]}\left(\beta^{[1]}\right) \right]^{-1} g_N^{[1]}\left(\beta^{[1]}\right)$$

.

Proof. Form

$$\tilde{f}(x_t, \beta, \gamma) = \begin{bmatrix} f^{[1]}(x_t, \beta^{[1]}) \\ f^{[2]}\left(x_t, \beta^{[1]}, \beta^{[2]}\right) - \gamma \end{bmatrix}$$

and construct similarly $\tilde{g}_N(\beta, \gamma)$. The proof follows in three steps.

1. $$\min_{\beta \in \mathbb{P}} g_N(\beta)' [V_N(\beta)]^{-1} g_N(\beta) \geq \min_{\beta \in \mathbb{P}, \gamma} \tilde{g}_N(\beta, \gamma)' [V_N(\beta)]^{-1} \tilde{g}_N(\beta, \gamma)$$

   The right-hand side minimization problem will not have a unique solution but this does not matter.

2. $$\min_{\gamma} \tilde{g}_N(\beta, \gamma)' [V_N(\beta)]^{-1} \tilde{g}_N(\beta, \gamma) = g_N^{[1]}(\beta^{[1]})' \left[V_N^{[11]}\left(\beta^{[1]}\right)\right]^{-1} g_N^{[1]}(\beta^{[1]}) \tag{47}$$

   This follows by using the first-order conditions for γ to show that

   $$g_N^{[2]}(\beta) - \gamma = \left(\begin{bmatrix} 0 & I \end{bmatrix} [V_N(\beta)]^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & I \end{bmatrix} [V_N(\beta)]^{-1} \begin{bmatrix} I \\ 0 \end{bmatrix} g_N^{[1]}\left(\beta^{[1]}\right).$$

   Substitute this outcome into the objective function on the left-hand side of (47) and apply the partitioned inverse formula to establish equality with the right-hand side of (47).

3. Finally,

   $$\begin{aligned} \min_{\beta \in \mathbb{P}, \gamma} \tilde{g}_N(\beta, \gamma)' [V_N(\beta)]^{-1} \tilde{g}_N(\beta, \gamma) &= \min_{\beta \in \mathbb{P}} \min_{\gamma} \tilde{g}_N(\beta, \gamma)' [V_N(\beta)]^{-1} \tilde{g}_N(\beta, \gamma) \\ &= \min_{\beta \in \mathbb{P}} g_N^{[1]}(\beta^{[1]})' \left[V_N^{[11]}\left(\beta^{[1]}\right)\right]^{-1} g_N^{[1]}(\beta^{[1]}). \end{aligned}$$

The conclusion follows from these three steps. □

We apply this result to an estimation problem where f_2 corresponds to the moment conditions added when we replicate the original moment conditions, and $\beta^{[2]}$ is introduced to parameterize the additional econometric relation when then model is underidentified. The previous lemma is not directly applicable to this problem because when we replicate moment conditions we add restrictions on the initial parameter vector $\beta^{[1]}$. However, restricting $\beta^{[1]}$ shrinks the parameter space $\mathbb{P}$ in the minimization problem given in the left-side of Lemma B.1 and hence can only increase the minimized objective function. Thus a corollary of this lemma is

Corollary B.2. *Consider the r moment conditions*

$$E[\tilde{f}(x_i, \tilde{\beta})] = 0$$

used to estimate the $k \times 1$ parameter vector $\tilde{\beta}_0$, and denote by I_j the value of the continuously-updated GMM version of the test of the null hypothesis that β is underidentified of dimension j introduced in section 3.1.2. Then, $I_j \geq I_{j-1}$ for any $j \geq 1$.

As a result, if we use continuously-updated GMM and allow for explorations across alternative degrees of underidentification, then the objective will lead us to the smallest allowable degree of underidentification. In particular, if we allow for the estimation of nonlinear curves such as (40) in a model that is fundamentally linear, then the continuously-updated GMM objective will lead us to represent the underidentification by means of a line or at least the segment of a line.

C Imhof-based approximation to the distribution of GMM tests

Let

$$g_N(\beta) = \frac{1}{N} \sum_{t=1}^{N} f(x_t, \beta),$$

and define

$$M = \lim_{T \to \infty} Var \left[\sqrt{N} g_N(\beta_o) \right].$$

Since the purpose of this appendix is to explain the application of Imhof (1961) results in our context, in what follows we will abstract from estimation issues by assuming that β_o is known.

As shown by Hansen (1982), under certain regularity conditions the quadratic form

$$N g_N(\beta_o) M^{-1} g_N(\beta_o)$$

will converge in distribution to a χ^2 random variable with p degrees of freedom as $N \to \infty$.

If the matrix M is ill-conditioned, the quality of the previous approximation can be rather poor. To address this problem, we could use the Tikhonov version of the generalized inverse,

and replace the above criterion function by

$$\begin{aligned} & Ng_N'(\beta_o)M^{1/2}(\eta_N I_p + M^2)^{-1}M^{1/2}g_N(\beta_0) \\ = \; & \sqrt{N}g_N(\beta_o)'W\Delta^{1/2}W'(\eta_N I_p + W\Delta^2 W')^{-1}W\Delta^{1/2}W'\sqrt{N}g_N(\beta_o) \\ = \; & \left[\sqrt{N}g_N(\beta_o)'W\Delta^{-1/2}\right]\left[(\eta_N I_p + \Delta^2)^{-1}\Delta^2\right]\left[\Delta^{-1/2}W'\sqrt{N}g_N(\beta_o)\right] \\ = \; & \sum_{j=1}^{p}\frac{\delta_j^2}{\delta_j^2+\eta_N}\left[\sqrt{N}\varepsilon_{j,N}\right]^2, \end{aligned}$$

where $W\Delta W'$ provides the spectral decomposition of M, $\varepsilon_{j,N}$ is the j^{th} entry of the random vector $\varepsilon_N = \Delta^{-1/2}W'g_N(\beta_0)$ and η_N is a regularization parameter. Since $\sqrt{N}\varepsilon_N \to N(0, I_p)$, we will recover the chi-square limiting distribution under the null if we let η_N go to 0 at a suitable rate. But given that for a fixed η_N the above statistic will converge to a diagonal quadratic form in standard normal random variables as $N \to \infty$, we can use Koerts and Abrahamse (1969) implementation of Imhof (1961) procedure for evaluating the probability that a quadratic form of normals is less than a given value (see also Farebrother (1990)). Although the smallest eigenvalue of M, $\delta_{\min}$ say, will generally be strictly positive, from a numerical point of view it makes sense to truncate the previous expression so that we only use those terms for which

$$\frac{\left(\frac{\delta_j^2}{\delta_j^2+\eta_N}\right)}{\left(\frac{\delta_{\max}^2}{\delta_{\max}^2+\eta_N}\right)}$$

exceeds some small threshold. Finally, since under standard regularity conditions the asymptotic distribution of the above tests is unaffected if we replace M with a consistent estimator, in practice we can treat the sample counterparts of δ_j as if they coincided with their population values.

The same analysis can be applied to GMM contexts with a continuum of moment conditions. For simplicity, we again discuss the case in which $\pi_0(\theta)$ is known, in which case our approach and the Carrasco and Florens (2000) approach coincide.

Define v and C as a vector and square matrix, respectively, of dimension N, with elements

$$\begin{aligned} c_{st} &= \frac{1}{N}\left\langle f\left[x_s, \pi_0(\theta)\right], f\left[x_t, \pi_0(\theta)\right]\right\rangle = \int_\Theta f\left[x_s, \pi_0(\theta)\right]' f\left[x_t, \pi_0(\theta)\right] d\theta \\ v_s &= \left\langle g_N\left(\pi_0(\theta)\right), f\left(x_s, \pi_0(\theta)\right)\right\rangle = \frac{1}{N}\sum_{t=1}^{N}\int_\Theta f\left[x_t, \pi_0(\theta)\right]' f\left[x_s, \pi_0(\theta)\right] d\theta = C_{\cdot s}'\iota_N, \end{aligned}$$

where $C_{.s}$ is the s^{th} column of C and ι_N is a vector of N 1's. Consider the spectral decomposition $C = U\Lambda U'$. Then, it is possible to show that the continuum of moment conditions test studied by Carrasco and Florens (2000) is numerically identical to the following expression

$$v'\left[\eta_N I_N + C^2\right]^{-1} v = \iota_N' C\left[\eta_N I_N + C^2\right]^{-1} C\iota_N = \iota_N' U \begin{bmatrix} \frac{\lambda_1^2}{\eta_N+\lambda_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\lambda_N^2}{\eta_N+\lambda_N^2} \end{bmatrix} U\iota_N.$$

Carrasco and Florens (2000) show that under certain conditions on the regularization constant η_N:

$$\frac{v'\left[\eta_N I_N + C^2\right]^{-1} v - p_N(\eta_N)}{\sqrt{q_T(\eta_N)}} \to N(0,1)$$

where

$$\begin{aligned} p_N(\eta_N) &= \sum_{j=1}^{N} \frac{\lambda_j^2}{\lambda_j^2 + \eta_N} \\ q_N(\eta_N) &= 2\sum_{j=1}^{T} \frac{\lambda_j^4}{\left(\lambda_j^2 + \eta_N\right)^2} \end{aligned}$$

As Carrasco and Florens (2000) argue in remark 11 of their paper, their test can also be asymptotically regarded as a centered and standardized version of a diagonal quadratic form in N standard normal variables. Thus we can again attempt to improve the finite sample approximation by using Imhof (1961) results treating the eigenvalues of the empirical matrix C as if they were the true eigenvalues of its population counterpart.

Another advantage of this Imhof approximation is that it will not breakdown when the number of strictly positive eigenfunctions is finite regardless of the sample size. Such a situation arises in the linear and non-linear in parameters models discussed in sections 3 and 4, respectively.

References

Alonso-Borrego, C. and M. Arellano. 1999. Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data. *Journal of Business and Economic Statistics* 17:36–49.

Anderson, T.W. and H. Rubin. 1949. Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations. *Annals of Mathematical Statistics* 20:46–63.

Arellano, M. 2003. *Panel Data Econometrics*. Oxford, England: Oxford University Press.

Arellano, M. and S.R. Bond. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies* 58:277–297.

Arellano, M., L.P. Hansen, and E. Sentana. 1999. Underdentification? *mimeo. CEMFI* .

Berndt, E.R. 1991. *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley.

Burnside, A.C. 2007. Empirical Asset Pricing and Statistical Power in the Presence of Weak Risk Factors. NBER Working Paper 13357.

Carrasco, M. and J.P. Florens. 2000. Generalization of GMM to a Continuum of Moment Conditions. *Econometric Theory* 17:794–834.

Chernozhukov, V., H. Hong, and E. Tamer. 2007. Estimation and inference on identified parameter sets. *Econometrica* 75:1243–1284.

Clarida, R.R., J. Galí, and M. Gertler. 2000. Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory. *Quarterly Journal of Economics* 115:147–180.

Cochrane, J.H. 2007. Identification with Taylor Rules: A Critical Review. *mimeo. Chicago Booth School* .

Cragg, J.G. and S.G. Donald. 1993. Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory* 9:222–240.

———. 1997. Inferring the Rank of a Matrix. *Journal of Econometrics* 76:223–50.

Farebrother, R.W. 1990. The distribution of a quadratic form in normal variables (Algorithm AS 256.3). *Applied Statistics* 39:294–309.

Fisher, F.M. 1966. *The Identificaton Problem in Econometrics*. McGraw-Hill Book Company.

Galí, J., M. Gertler, and J.D. López-Salido. 2001. European Inflation Dynamics. *European Economic Review* 45:1237–1270.

Hansen, Lars Peter and Kenneth J Singleton. 1996. Efficient Estimation of Linear Asset-Pricing Models with Moving Average Errors. *Journal of Business & Economic Statistics* 14 (1):53–68.

Hansen, L.P. 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50:1029–1054.

———. 1985. A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics* 30:203–238.

Hansen, L.P. and K.J. Singleton. 1982. Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica* 50.

———. 1983. Stochastic Consumption, Risk Aversion and the Temporal Behavior of Assest Returns. *Journal of Political Economy* 91.

Hansen, L.P., J. Heaton, and A. Yaron. 1996. Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics* 14:262–280.

Hayashi, F. 2000. *Econometrics*. Princeton, NJ: Princeton University Press.

Hillier, G.H. 1990. On the Normalization of Structural Equations: Properties of Direction Estimators. *Econometrica* 58:1181–1194.

Imhof, J.P. 1961. Computing the distribution of quadratic forms in normal variables. *Biometrika* 48:419–426.

Kim, H. and M. Ogaki. 2009. Purchasing Power Parity and the Taylor Rule. Ohio State University Department of Economics Working Paper No. 09-03.

Kleibergen, F. 2005. Testing Parameters in GMM Without Assuming that They Are Identified. *Econometrica* 73:1103–1123.

Koerts, J. and A.P.J. Abrahamse. 1969. *On the theory and application of the general linear model*. Rotterdam: Rotterdam University Press.

Koopmans, T.C. and W.C. Hood. 1953. The Estimation of Simultaneous Linear Economic Relationships. In *Studies in Econometric Method*, edited by W. C. Hood and T. C. Koopmans, Cowles Commission Monograph Number 14, chap. 6. Wiley.

Nagel, S. and K. J. Singleton. 2009. Estimation and Evaluation of Conditional Asset Pricing Models. Stanford University.

Nason, J.M. and G.W. Smith. 2008. Identifying the New Keynesian Phillips Curve. *Journal of Applied Econometrics* 23:525–551.

Rao, C.R. 1973. *Linear Statistical Inference and Its Applications*. Wiley, second ed.

Rothenberg, T.J. 1971. Identification in parametric models. *Econometrica* 39:577–591.

Sargan, J.D. 1958. The Estimation of Economic Relationships Using Instrumental Variables. *Econometrica* 26:393–415.

———. 1959. The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society Series B* 21:91–105.

———. 1983a. Identification and Lack of Identification. *Econometrica* 51:1605–1633.

———. 1983b. Identification in Models with Autoregressive Errors. In *Studies in Econometrics, Time Series, and Multivariate Statistics*, edited by S. Karlin, T. Amemiya, and L. A. Goodman. New York: Academic Press.

Stock, J.H. and J.H. Wright. 2000. GMM with Weak Identification. *Econometrica* 68:1055–1096.

Stock, J.H., J.H. Wright, and M. Yogo. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics* 20:518–529.

Wright, J.H. 2003. Detecting lack of identification in GMM. *Econometric Theory* 19:322–330.