Underidentification?*

Manuel Arellano (CEMFI)
Lars Peter Hansen (University of Chicago)
Enrique Sentana (CEMFI)

July 12, 1999

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. We provide a way to test for underidentification using a method that is commonly employed as a test for overidentification.

We study the identification of an econometric model that is linear in parameters. We adopt a generalized-method-of-moments (GMM) perspective and write the model as:

$$E(\Psi_t)\alpha = 0 \tag{1.1}$$

where α is a $k+1$-dimensional unknown parameter vector in the null space of the population matrix $E(\Psi_t)$ where Ψ_t is an r by $k+1$ matrix constructed from data.

*Very preliminary draft. We thank Javier Alvarez, Raquel Carrasco, and Francisco Peñaranda for able research assistance, and John Campbell for kindly allowing us to use his data.

We suppose the order condition ($r \geq k$) is satisfied, but not necessarily the rank condition. Thus the maximal possible rank of the matrix $E(\Psi_t)$ is $\max\{r, k+1\}$. The model is said to be *identified* when the null space of $E(\Psi_t)$ is precisely one dimensional. In this case the parameter vector of interest is obtained by imposing a normalization that selects one element from the null space. The selection rule can restrict one of the components of α to be one, or it might require that $|\alpha| = 1$ together with a sign restriction on one of the nonzero coefficients. Identification follows when the matrix $E(\Psi_t)$ has rank k. When $r > k$ and the model is identified, it is said to be *over-identified* because the rank of the matrix $E(\Psi_t)$ now must not be full. Instead of having maximal rank $k + 1$, $E(\Psi_t)$ has reduced rank k. This implication is known to be testable and statistical tests of overidentification are often conducted in practice.

The model is said to be *under-identified* when the rank of $E(\Psi_t)$ is less than k. In this case the null space of $E(\Psi_t)$ will have more than one dimension. A single normalization will no longer select a unique element from the parameter space. Instead there exists another solution α^* not proportional to α such that

$$E(\Psi_t)\alpha^* = 0. \tag{1.2}$$

Tests for underidentification are not common in econometric practice, and the aim of this paper to propose and implement such a test.

To test for the lack of identification, we ask whether there exists another normalized vector α^* that satisfies (1.2). We approach this question by thinking of (1.1) and (1.2) as emerging from a new *augmented model*. We attempt to determine (α, α^*) simultaneously and ask whether they satisfy the combined over-identifying moment restrictions. If they do, then we may conclude that the original econometric relation is *not identified* or equivalently is under-identified. Thus by building an augmented equation system, we may pose the null hypothesis of underidentification as a hypothesis that the augmented equation system is over-identified. Rejections of the over-identifying 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 test for overidentification such as that of Sargan (1958) (and extended by Hansen, 1982) is potentially applicable to the augmented model.

As we will see, there are two complications that must be considered in this implementation. First, under the null hypothesis of underidentification, we are compelled to extend the normalization to extract multiple, linearly independent elements from the null space of $E(\Psi_t)$. For instance, if (α, α^*) satisfies (1.1) and (1.2), then so does any pair of linear combinations of α and α^*. Since the parameter estimates of the augmented model are of no particular interest to us, it is of little consequence which rule is used to achieve identification. Any convenient normalizations will suffice, and it is known how to construct GMM estimators that are insensitive to normalization. Second, when we duplicate the moment relations to achieve identification of the augmented model, we may introduce some redundancy into the system. As a consequence, sometimes we will be compelled to use less than the full $2r$ moment conditions from the augmented system when testing for underidentification. We will provide some guidance as to when to expect redundancy in the moment conditions.

The remainder of the paper is organized as follows. Section 2 considers identification testing in the context of a single structural equation, including comparisons to other approaches. We discuss the relationship of our method with the minimum eigenvalue tests suggested by Koopmans and Hood (1953) and Sargan (1958), and the reduced form approach proposed by Cragg and Donald (1993). Section 3 deals with cross-equation restrictions, discussing an example motivated in the estimation of an intertemporal asset pricing model. Section 4 considers identification testing in autoregressive models with individual effects for short panels. This is an example of a system of equations in which the valid instruments differ for different equations, and the model has a nonstandard reduced form. We provide empirical illustrations and Monte Carlo simulations for the asset pricing and the panel data examples. Finally, section 5 contains the conclusions.

2. Single Equation

To illustrate our method and compare to other approaches, in this section we consider a single equation from a simultaneous system. Suppose that:

$$w_t'\alpha = u_t \tag{2.1}$$

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

$$E\left(z_t u_t\right) = 0. \tag{2.2a}$$

Typically, the first coefficient of α is set to one so that $\alpha = (1, \beta')'$, where β is a k-dimensional vector of parameters. Form:

$$\Psi_t = z_t w_t'.$$

Then orthogonality condition (2.2a) is equivalent to α satisfying the moment relation (1.1).

The parameter vector α will be identified up to scale, if and only if the rank of $E(z_t w_t')$ is equal to k, which requires the order condition $r \geq k$. If the rank of $E(z_t w_t')$ is $k-1$ or less, the equation is not identified regardless of the difference between r and k.

Let us assume that $r \geq k$ but the rank of $E(z_t w_t')$ is actually $k-1$. This means that all the parameter values compatible with (2.2a) will lie in a linear subspace of dimension 2. Then we can write all the admissible equations as linear combinations of the $2r$ moment conditions

$$\begin{aligned} E(z_t w_t')\alpha &= 0 \\ E(z_t w_t')\alpha^* &= 0. \end{aligned}$$

Thus to test for underidentification, we effectively introduce a second equation to the system:

$$w_t'\alpha^* = v_t \tag{2.3}$$

combined with the orthogonality condition:

$$E\left(z_t v_t\right) = 0,$$

and we study the simultaneous overidentification of the two econometric equations (2.1) and (2.3). Even after normalizing each equation, there are two superfluous dimensions to this parameterization. It may be possible, for example, to avoid indeterminacy by choosing the two top rows of (α, α^*) equal to the identity matrix of order two, which eliminates two parameters per equation. Alternatively, we can impose the normalizing restrictions $(\alpha, \alpha^*)'(\alpha, \alpha^*) = I_2$ and set the $(1,2)$-th element of (α, α^*) to zero. In any event, the effective number of parameters is $2k-2$ and the number of moment conditions is $2r$. If the $2(r-k)+2$ over-identifying restrictions for the newly constructed two-equation system are rejected, then we have rejected the underidentification of the original econometric relation. We may be confident in that our single equation model is identified.

In practice, it is desirable to construct a test statistic of underidentification using a version of the test of over-identifying restrictions that is invariant to normalization, such as those based on continuously updated GMM (Hansen, Heaton, and Yaron, 1996), empirical likelihood estimation, or other information-theoretic alternatives (Imbens, 1997, Kitamura and Stutzer, 1997, Imbens, Spady, and Johnson, 1998).

Underidentification of a Higher Order Although the null hypothesis that sets the rank of $E(z_t w_t')$ to $k-1$ is the natural leading case in testing for underidentification, it is straightforward to extend the previous discussion to higher orders of underidentification. Suppose that the rank of $E(z_t w_t')$ is $k-j$ for some j. Then we can write all the admissible equations as linear combinations of the $(j+1)r$ orthogonality conditions

$$E(z_t w_t')(\alpha, \alpha_1^*, ..., \alpha_j^*) = 0. \tag{2.4}$$

If we impose $(j+1)^2$ normalizing restrictions on $(\alpha, \alpha_1^*, ..., \alpha_j^*)$ to avoid indeterminacy,[1] the effective number of parameters is $(j+1)(k+1)-(j+1)^2 = (j+1)(k-j)$ and the number of moment conditions is $(j+1)r$. Therefore, by testing the $(j+1)(r-k+j)$ over-identifying restrictions in (2.4) we test the null that α is underidentified of order j against the alternative of underidentification of order

[1]For instance, we may make the top $j+1$ rows of $\mathcal{A} = (\alpha, \alpha_1^*, ..., \alpha_j^*)$ equal to the identity matrix of order $j+1$. More generally, we can impose the $(j+1)^2$ normalizing restrictions $\mathcal{A}'\mathcal{A} = I_{(j+1)}$ and $a_{i\ell} = 0$ for $\ell > i$, where $a_{i\ell}$ denotes the (i, ℓ)-th element of $\mathcal{A}$.

less than j or identification.

Other Approaches Tests of underidentification in a single structural equation were first considered by Koopmans and Hood (1953) and Sargan (1958).

When $r > k$ and the rank of $E(z_t w_t')$ is k, under the additional assumptions that the error term u_t is conditionally homoskedastic and serially uncorrelated, an asymptotic chi-square test statistic of over-identifying restrictions with $r - k$ degrees of freedom is given by $T\lambda_1$, where

$$\lambda_1 = \min_a \frac{a'W'Z\,(Z'Z)^{-1}\,Z'Wa}{a'W'Wa},$$

and $Z'W = T^{-1}\sum_{t=1}^{T} z_t w_t'$, etc. Thus λ_1 is the smallest characteristic root of $W'Z\,(Z'Z)^{-1}\,Z'W$ in the metric of $W'W$ (Anderson and Rubin, 1949, Sargan, 1958).

Koopmans and Hood, and Sargan indicated that when the rank of $E(z_t w_t')$ is $k - 1$ instead, if λ_2 is the second smallest characteristic root, $T(\lambda_1 + \lambda_2)$ 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 a homoskedastic and non-autocorrelated error.

The statistic $T(\lambda_1 + \lambda_2)$ has a straightforward interpretation in terms of our approach. Indeed, it can be regarded as a continuously updated GMM test of over-identifying restrictions of the augmented model (2.1) and (2.3), subject to the additional restrictions on the error terms mentioned above. To see this, let $A = (a, a^*)$ and consider the minimizer of

$$(a'W'Z, a^{*\prime}W'Z)\,(A'W'WA \otimes Z'Z)^{-1} \begin{pmatrix} Z'Wa \\ Z'Wa^* \end{pmatrix}$$

subject to $A'W'WA = I_2$. The constraint restricts the sample covariance matrix of the disturbance vector to be an identity matrix. It is used as a convenient normalization for the two equation system.[2] The minimization problem may be written equivalently as

$$\min_{a'W'Wa^*=0, a'W'Wa=1, a^{*\prime}W'Wa^*=1} a'W'Z\left(Z'Z\right)^{-1}Z'Wa + a^{*\prime}W'Z\left(Z'Z\right)^{-1}Z'Wa^*,$$

and the minimized value coincides with $\lambda_1 + \lambda_2$ (Rao, 1973, page 63).

More recently, Cragg and Donald (1993) considered single equation tests of underidentification based on the reduced form. Let us partition w_t into a $(p+1)$- and a r_1-dimensional vectors of endogenous and predetermined variables, respectively, $w_t = (y_t', z_{1t}')'$, so that $k = p + r_1$ and $z_t = (z_{1t}', z_{2t}')'$, where z_{2t} is the vector of r_2 instruments excluded from the equation. Moreover, let Π and $\widehat{\Pi} = Y'Z(Z'Z)^{-1}$ be the $(p+1) \times r$ matrices of population and sample reduced form linear-projection coefficients, respectively. With this notation and the partition $\Pi = (\Pi_1, \Pi_2)$ corresponding to that of z_t, if the rank of Π_2 is p, α is identified up to scale, 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

$$T[vec(\widehat{\Pi} - \Pi)]'V^{-1}vec(\widehat{\Pi} - \Pi) \tag{2.5}$$

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

[2] This normalization does not fully identify a set of parameters of (2.1) and (2.3), but it is enough to provide an explicit expression for the test statistic.

estimate of the asymptotic variance of $vec(\widehat{\Pi})$.[3]

To relate (2.5) to our framework, write the augmented model (2.1) and (2.3) as a complete system by adding to it $p-1$ reduced form equations, and denote it by

$$By_t + Cz_t = u_t^{\dagger}.$$

Then noting that $vec(\widehat{\Pi} - \Pi) = (B \otimes Z'Z)^{-1} \sum_{t=1}^{T} (u_t^{\dagger} \otimes z_t)$, (2.5) can be expressed as

$$\sum_{t=1}^{T} (u_t^{\dagger} \otimes z_t)' [(B \otimes Z'Z) V (B' \otimes Z'Z)]^{-1} \sum_{t=1}^{T} (u_t^{\dagger} \otimes z_t),$$

which is in the form of a continuously updated GMM criterion that depends on (α, α^*) and the coefficients in the additional $p-1$ reduced form equations. Since B does not depend on the latter, they 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.

3. Cross-Equation Restrictions

In the standard simultaneous equations system, we may test for identification equation by equation using the approach described in Section 2. Moreover, if we were to look at multiple equations simultaneously, our implicit null hypothesis would be that none of the equations are identified. Rejecting this hypothesis we could only conclude that at least one of the equations is identified. We could not conclude that all equations are identified from this one system test. Thus in the

[3] 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.

absence of cross-equation restrictions it seems only interesting to proceed with one equation at a time.[4]

When cross equation restrictions are present matters are different. It now makes sense to look at more than one equation at a time when testing for identification, since parameters are no longer uniquely tied to equations. In so doing, we may encounter a problem of redundancy in our moment conditions, as we now illustrate.

Example 3.1. *Consider the following two equation model:*

$$\begin{aligned} y_{1t} &= \alpha_1 + x_t\beta + u_{1t} \\ y_{2t} &= \alpha_2 + x_t\beta + u_{2t} \end{aligned}$$

where y_{1t}, y_{2t} *and* x_t *are endogenous variables. Let* z_t *denote a vector of instrumental variables appropriate for both equations:*

$$\begin{aligned} E\left(z_t u_{1t}\right) &= 0 \\ E\left(z_t u_{2t}\right) &= 0. \end{aligned}$$

To test for underidentification in the first equation alone we would introduce a second equation and three additional normalizations:

$$\begin{aligned} y_{1t} &= \gamma_1 + v_{1t} \\ x_t &= \gamma_0 + v_{0t}. \end{aligned}$$

[4] Our formulation of the moment conditions (1.1) imposed a single normalization. If the test just described were of interest for multiple equations without cross-equation restrictions, then the effective number of parameters for the augmented model would be altered to account for the multiple normalizations in the original model.

But for the two equation system, we do not want to augment each equation because in both cases we would arrive at the same econometric relation for x_t. Thus to test for underidentification, we are led to study a three equation nonredundant system. Therefore, what is tested is:

$$\begin{aligned} E[z_t(y_{1t} - \gamma_1)] &= 0 \\ E[z_t(y_{2t} - \gamma_2)] &= 0 \\ E[z_t(x_t - \gamma_0)] &= 0. \end{aligned}$$

This example illustrates a common phenomenon. Suppose we look at m equations with g endogenous variables. If the instrumental variables are the same for each of the m equations, then augmenting the m equations to $2m$ equations in the g variables will generate redundant moment conditions whenever $2m$ exceeds g. The maximal number of additional nonredundant equations is $\min\{2m, g\}$. As we will see in the next section, matters are a bit more complicated when instrumental variables appropriate for each equation differ.

An Asset Pricing Model The previous example can be motivated in the GMM estimation of a standard consumption-based capital asset-pricing model. Suppose a representative agent who maximizes expected isoelastic utility over present and future consumption. The Euler equations for the agent's consumption and portfolio allocation decision are given by

$$E_{t-1}[\exp(y_{jt} + \ln\rho - \beta x_t)] = 1 \ (j = 1, ..., m)$$

where x_t denotes the change in log consumption between $t-1$ and t, y_{jt} is the (continuously compounded) return on the j-th financial asset in period t, β is the coefficient of relative risk aversion, and ρ is the discount factor (Hansen and Singleton, 1983). There are m assets, and the conditional expectation is taken with respect to the agent's information set in period $t-1$, which includes past returns and consumption. Moreover if $(x_t, y_{1t}, ..., y_{mt})$ are conditionally jointly normally distributed with a constant covariance matrix, then

$$E_{t-1}(y_{jt} - \alpha_j - \beta x_t) = 0 \ (j = 1, ..., m), \tag{3.1}$$

where the asset-specific intercepts α_j depend on the discount factor, and the conditional variances and covariances of asset returns and consumption growth.

In the example there are two assets, and estimation of the parameters α_j and β is based on the unconditional moment restrictions:

$$E[z_t(y_{jt} - \alpha_j - \beta x_t)] = 0 \tag{3.2}$$

where z_t is a vector of instrumental variables whose values are known in $t-1$. The coefficient of relative risk aversion is identified as the common slope of linear combinations of asset returns and consumption growth that are unpredictable on the basis of the vector of instruments. However, if $cov(z_t, x_t) = cov(z_t, y_{jt}) = 0$ (the null of our test in this example) there will be a multiplicity of linear combinations with the same property, and as a result the true value of β will not be empirically identifiable from (3.2).

Empirical Application to US Data We illustrate the situation discussed above using US annual data on returns and consumption growth for the period 1889-1994. The asset returns are (1) the real commercial paper rate and (2) the real stock return. Consumption growth is the annual growth rate of real non-durables and services consumption. Apart from the constant, the instruments are one lag of the real commercial paper rate, the real consumption growth rate, and the log dividend-price ratio. The data is the same as in section 8.2 of Campbell, Lo and MacKinlay (1997), where further details can be found.

In Table 1 we report two-step and continuously updated GMM estimates of the parameters in the original model as well as those in the augmented model. We also report test statistics of overidentifying restrictions for both the augmented and the original models (denoted the I and the J tests, respectively). The estimates and test statistics are robust to heteroskedasticity but not to serial correlation. Therefore, the I statistic is really testing the null hypothesis that the original system is underidentified and that any possible equation has non-autocorrelated errors. While lack of serial correlation is an implication of (3.1), one could argue that inferences that are robust to serial correlation as well as to heteroskedasticity would be more appropriate, since the exercise is aimed at testing underidentification on the basis of (3.2) alone.

According to our results, there seems to be information in the instruments employed since the I tests reject the null of underidentification. However, the results are not very encouraging for the original specification, since the J tests only marginally accept the overidentifying restrictions (at the one percent level, but not at five percent), and the estimated relative risk aversion parameter has the wrong sign.

Monte Carlo Experiment in the Asset Pricing Setting We generated 10,000 time series of size $T = 100$ from the following model:

$$\begin{aligned} y_{1t} &= \alpha_1 + \beta[\mu + \delta(y_{1(t-1)} + x_{t-1} + w_{t-1})] + \varepsilon_{1t} \\ y_{2t} &= \alpha_2 + \beta[\mu + \delta(y_{1(t-1)} + x_{t-1} + w_{t-1})] + \varepsilon_{2t} \end{aligned}$$

$$
\begin{aligned}
x_t &= \mu + \delta(y_{1(t-1)} + x_{t-1} + w_{t-1}) + \varepsilon_{3t} \\
w_t &= \pi w_{t-1} + \varepsilon_{4t}.
\end{aligned}
$$

This specification ensured that the moment restrictions (3.2) were satisfied with $z_t = (1, y_{1(t-1)}, x_{t-1}, w_{t-1})'$ as in the empirical application. It also had the property that none of the variables' forecasts would improve by using $y_{2(t-1)}$ (the lagged real stock return). We considered one experiment under the null hypothesis of underidentification, setting $\delta = 0$, and another under the alternative of identification with $\delta = 0.05$. In both cases, we set $\alpha_1 = \alpha_2 = 0$, $\mu = 0.05$, $\beta = 1$, and $\pi = 0.9$ (to reflect observed persistence in the log dividend-price ratio). Disturbances were generated as $N(0, I)$, and the initial observations were obtained from the stationary distribution of the process.

Table 2 shows the 10, 5, and 1 percent rejection frequencies for the (heteroskedasticity robust) two-step and continuously updated GMM versions of the I test statistic. Their behaviour is broadly the same, although the continuously updated test is slightly more conservative than the two-step. Size distortion in the experiment conducted under the null is not negligible, as both tests show a tendency to under-reject relative to nominal sizes. The rejection frequencies under the alternative ($\delta = 0.05$) are at least four times those obtained under the null, and give an idea of the power the test can be expected to have for small values of δ in this environment.

4. Sequential Moments: Panel Data AR Models

We now turn to consider systems of equations in which the valid instruments differ for different equations. A leading example is given by autoregressive models with individual effects for short panels. In those cases our approach provides a straightforward way of testing for underidentification, which is specially useful since the models have a nonstandard reduced form. We first discuss the AR(2) case, and subsequently generalize the result to an autoregressive process of an arbitrary order.

Consider a second-order autoregressive model for panel data with an individual specific intercept η_i:

$$y_{it} - \eta_i = \alpha_1(y_{i(t-1)} - \eta_i) + \alpha_2(y_{i(t-2)} - \eta_i) + v_{it} \ (t = 3, ..., T), \tag{4.1}$$

such that $T \geq 4$ but small, $\{y_{i1}, ..., y_{iT}, \eta_i\}$ is an *i.i.d.* random vector and

$$E(v_{it} \mid y_{i1}, ..., y_{i(t-1)}) = 0. \tag{4.2}$$

We consider GMM estimation of α_1 and α_2 based on a random sample of size N $\{y_{i1}, ..., y_{iT}\}_{i=1}^N$ and the unconditional moment restrictions (as in Arellano and Bond, 1991):

$$E[y_i^{t-2}(\Delta y_{it} - \alpha_1 \Delta y_{i(t-1)} - \alpha_2 \Delta y_{i(t-2)})] = 0 \ (t = 4, ..., T). \tag{4.3}$$

where $y_i^s = (y_{i1}, ..., y_{is})'$. Thus, we have a system of $T-3$ equations in first-differences with an expanding set of admissible instruments but common parameters.

With $T = 4$ there is a single equation in first differences with two instruments so that α_1 and α_2 are just identified at most. Testing for underidentification in this case is therefore an example of the situation discussed in section 2. It amounts to testing for overidentification the following four moments involving two unknown coefficients γ_1 and γ_2:[5]

$$E\left[\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \otimes \begin{pmatrix} \Delta y_{i4} - \gamma_1 \Delta y_{i3} \\ \Delta y_{i3} - \gamma_2 \Delta y_{i2} \end{pmatrix}\right] = 0. \tag{4.4}$$

If (4.4) holds, (4.3) will hold not only for the true values α_1 and α_2, but also for any other α_1^* and α_2^* along the line $\alpha_2^* = \gamma_1\gamma_2 - \alpha_1^*\gamma_2$. Note that if the autoregressive process contains a unit root so that $\alpha_1 + \alpha_2 = 1$, the moment conditions (4.4) hold with $\gamma_1 = \gamma_2 = -\alpha_2$.

With $T = 5$ a second equation and three additional instruments become available. Single equation testing for the second equation would be based on:

$$E\left[\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{pmatrix} \otimes \begin{pmatrix} \Delta y_{i5} - \gamma_1 \Delta y_{i4} \\ \Delta y_{i4} - \gamma_2 \Delta y_{i3} \end{pmatrix}\right] = 0.$$

However, the moments $E[(y_{i1}, y_{i2})(\Delta y_{i4} - \gamma_2 \Delta y_{i3})] = 0$ are clearly redundant given those in (4.4) implying that $\gamma_1 = \gamma_2$. Moreover, although associated with the second equation, the restriction $E[y_{i3}(\Delta y_{i4} - \gamma_2 \Delta y_{i3})] = 0$ can be actually tested with $T = 4$.

For larger values of T we obtain a similar pattern of redundancies. Namely, all the moments associated with the second equation in the augmented system, except the last one, are redundant given those for the earlier periods. Therefore, for $T \geq 5$ a test of underidentification will be based on the $(T-1)T/2-1$ moments

$$E\left[y_i^{t-1}\left(\Delta y_{it} - \gamma_1 \Delta y_{i(t-1)}\right)\right] = 0, \ (t = 3, ..., T). \tag{4.5}$$

[5] An equivalent normalization is

$$E\left[\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \otimes \begin{pmatrix} \Delta y_{i4} - \gamma_1^* \Delta y_{i2} \\ \Delta y_{i3} - \gamma_2 \Delta y_{i2} \end{pmatrix}\right] = 0$$

with $\gamma_1^* = \gamma_1\gamma_2$.

Since there is only one unknown coefficient, an I test statistic will have an asymptotic χ^2 distribution with $(T-1)T/2-2$ degrees of freedom provided (4.5) holds.

Generalizing the previous argument, an I test for an autoregressive process of order p with individual effects will be a test for overidentification based on

$$E\left[y_i^{t-1}\left(\Delta y_{it}-\gamma_1\Delta y_{i(t-1)}...-\gamma_{(p-1)}\Delta y_{i(t-p+1)}\right)\right]=0\ (t=p+1,...,T). \quad (4.6)$$

In particular, for a first-order process the relevant orthogonality conditions are

$$E\left[y_i^{t-1}\Delta y_{it}\right]=0,\ \ (t=2,...,T). \quad (4.7)$$

Empirical Illustration As an illustration of the previous results, we present in Table 3 parameter estimates, and I and J test statistics for an AR(2) model of employment using the Arellano-Bond dataset. These data consists of an unbalanced panel of 140 quoted firms from the U.K. for which seven, eight, or nine continuous annual observations are available for the period 1976-1984.

The AR(2) results were reported by Alonso-Borrego and Arellano (1999), who interpreted the large disparities between two-step and continuously updated GMM as indicating that the estimates were much less reliable than what their asymptotic standard errors would suggest. Note that the J test statistics give no indication of misspecification. All the statistics shown in the table are robust to heteroskedasticity.

The I test statistics are borderline, since the null hypothesis that the relationship is a priori unidentified can be marginally rejected at the five percent level but not at one percent. In any event, the I statistic in this case provides a useful qualitative indication that the estimates are not very well identified.

Monte Carlo Simulation We simulated data to acquire some information about the size and power properties of the I test in an environment that is related to the application.

In order to investigate possible size distortion, we simulated 10,000 balanced panels of size $N = 150$ and $T = 7$ from model (4.1) with $v_{it} \sim iid\ N(0, 1)$ and a unit root. Specifically, if we denote by μ_1 and μ_2 the largest and smallest roots of the AR(2) polynomial, so that $\alpha_1 = \mu_1 + \mu_2$ and $\alpha_2 = -\mu_1\mu_2$, we chose $\mu_1 = 1$ and $\mu_2 = 0.4$. The latter was chosen to mimic the estimate obtained for γ_1 with the empirical data. As for initial conditions, we first generated $w_{it} = \mu_2 w_{i(t-1)} + v_{it}$, with initial values drawn independently from its unconditional distribution, and then generated $y_{it} = \mu_1 y_{i(t-1)} + w_{it}$, setting the initial values to zero and discarding the first ten observations. To investigate local power we produced another round of simulations in which the unit root was replaced with $\mu_1 = 0.98$ and individual effects were set to zero. The other features of the experiment remained the same as in the first one.

Table 4 shows some rejection frequencies for the (heteroskedasticity robust) two-step and continuously updated versions of the test statistic. The two forms of the test exhibit a similar performance. Size distortion is small, taking into account that sample size is not large, although there is some tendency to over-reject at the 10 percent significance level. We might expect larger size distortion for larger values of μ_2. Indeed, for $\mu_2 = 1$ the AR(2) model would exhibit a larger degree of underidentification since not only α_1 and α_2 but also γ_1 would be underidentified. If this were the relevant null, an I test could be easily constructed for it, but the I test statistics that assume the uniqueness of γ_1 would not have an asymptotic chi-square distribution.

Rejection frequencies under the chosen alternative are about twice the size of those obtained under the null, so power is not very high in our experiment, but it would obviously increase for smaller μ_1 and larger N.

5. Conclusions

In instrumental variables 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, 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."

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 over-identified. However, this situation is often coupled with informal evidence that identification may be at fault. In those cases, an identification test may provide a useful diagnostic of the extent to which estimates are well identified.

In this paper we have proposed a method for constructing tests of underidentification based on the structural form of the equation system. We regard underidentification as a set of over-identifying restrictions imposed on an augmented structural model. Therefore, our proposal is to test for underidentification by testing for overidentification in the augmented model using standard testing methods that are available in the literature.

We show that our approach can be used not only for single equation models, but also for systems with cross-equation restrictions, possibly with different valid instruments for different equations. As examples we consider intertemporal asset pricing models, and autoregressive models with individual effects for short panels. We also provide empirical calculations and Monte Carlo simulations in order to illustrate the use and finite sample properties of identification tests in those environments.

A relevant issue which is outside the scope of this paper is whether and how these procedures could be extended to testing for underidentification in nonlinear GMM problems.

References

Alonso-Borrego, C. and M. Arellano (1999): "Symmetrically Normalized Instrumental-Variable Estimation Using Panel Data", *Journal of Business & 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. 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.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay (1997): *The Econometrics of Financial Markets*, Princeton University Press.

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

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

Hansen, L.P. and K. Singleton (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns", *Journal of Political Economy*, 91, 249-265.

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

Imbens, G. (1997): "One-step Estimators for Over-identified Generalized Method of Moments Models", *Review of Economic Studies*, 64, 359-383.

Imbens, G., R. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models", *Econometrica*, 66, 333-357.

Kitamura, Y. and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica*, 65, 861-874.

Koopmans, T.C. and W.C. Hood (1953): "The Estimation of Simultaneous Linear Economic Relationships". In: Hood, W.C. and T.C. Koopmans (eds.): *Studies in Econometric Method*, Cowles Commission Monograph No. 14, Chapter 6, Wiley.

Rao, C.R. (1973): *Linear Statistical Inference and Its Applications*, Second Edition, Wiley.

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

Table 1
Consumption-Based Capital Asset-Pricing Model
GMM Estimates from US Annual Data

	Two-step	Continuous-updating
β	-1.533	-3.108
	(1.12)	(1.70)
α_1	.054	.085
	(.020)	(.031)
α_2	.101	.127
	(.025)	(.035)
γ_0	.022	.024
	(.003)	(.003)
γ_1	.022	.024
	(.004)	(.004)
γ_2	.078	.084
	(.015)	(.014)
I test (df)	23.94 (9)	23.93 (9)
p-value (%)	0.4	0.4
J test (df)	13.2 (5)	11.6 (5)
p-value (%)	2.1	4.0

NOTE: The sample period is 1889-1994.
Asymptotic standard errors robust to heteroskedasticity shown in parentheses.

Table 2
Size and Power of the I Tests in the Asset Pricing Example
Rejection Frequencies (%) (df=9)

Nominal	Under the null ($\delta = 0$)		Under the alternative ($\delta = .05$)	
level	Two-step	Continuous-updating	Two-step	Continuous-updating
10	9.0	8.6	33.1	31.8
5	3.7	3.4	19.9	19.7
1	0.4	0.3	5.1	4.3
Mean	9.1	9.1	12.9	12.7
Variance	14.9	14.4	23.7	22.5

NOTE: $10,000$ replications, $T = 100$, $v_{it} \sim iid\ N(0, I)$.

Table 3
AR(2) Employment Models with Individual Effects
GMM Estimates in First Differences from a Panel of U.K. Firms

	Two-step	Continuous-updating
α_1	.320	.092
	(.053)	(.047)
α_2	.022	.218
	(.023)	(.019)
γ_1	.314	.416
	(.022)	(.022)
I test (df)	51.1 (34)	48.8 (34)
p-value (%)	3.0	4.8
J test (df)	32.8 (25)	31.7 (25)
p-value (%)	13.7	16.6

NOTE: Unbalanced panel of 140 companies with 7, 8, or 9 annual observations. The sample period is 1976-1984. Time dummies are included in all equations. Asymptotic standard errors robust to heteroskedasticity shown in parentheses.

Table 4
Size and Power of the I Tests in the Panel Example
Rejection Frequencies (%) (df=19)

Nominal	Under the null ($\mu_1 = 1$)		Under the alternative ($\mu_1 = .98$)	
level	Two-step	Continuous-updating	Two-step	Continuous-updating
10	10.9	10.9	19.4	18.8
5	5.5	5.3	10.2	9.8
1	1.0	0.9	2.3	2.2
Mean	19.8	19.7	21.9	21.8
Variance	36.0	35.5	41.0	40.2

NOTE: $10,000$ replications, $N = 150$, $T = 7$, $v_{it} \sim iid\ N(0,1)$, $\eta_i \equiv 0$.
Smaller root is set to $\mu_2 = 0.4$.