XXVIII Simposio de Análisis Económico
Sevilla, December 11, 2003

Presidential Address: Underidentification?

Manuel Arellano

A lecture from joint work with L. P. Hansen & E. Sentana

- This lecture is concerned with testing for underidentification in instrumental variable models.
- The early econometric literature recognized that underidentification is testable, but to date such tests are uncommon in econometric practice.
- Many econometric models imply a large number of moment restrictions relative to the number of unknown parameters and are therefore *seemingly* overidentified.
- However, this is often coupled with informal evidence that identification may be at fault.
- In those cases, a test of underidentification (denoted an I test here) may provide a useful diagnostic of the extent to which estimates are well identified.

Outline

1. Introduction
   What can and what cannot be tested?
   Meaning and implications of test outcomes
2. A GMM testing approach
3. Identification testing in a single equation
4. Higher order underidentification
5. Taylor rules and Phillips curves
6. Cross-equation restrictions
   Asset pricing
   Demand systems with symmetry
   Sequential moments for panel data
7. Conclusions

1. Introduction

- Take as a first example a relationship between two endogenous variables with one instrument z_t
$$y_t = \beta x_t + u_t.$$
- Identification of β relies on two assumptions:
$$\begin{aligned} Cov\left(z_t, u_t\right) &= 0 \quad \textit{Orthogonality} \\ Cov\left(z_t, x_t\right) &\neq 0 \quad \textit{Relevance} \end{aligned}$$
- Orthogonality is not testable, but relevance is.
- If there is more than one instrument, their compatibility for the same relationship can be tested (Sargan test):
$$Cov\left(z_t, y_t - \beta x_t\right) = 0.$$
- In testing for underidentification the null hypothesis is
$$H_0 : \begin{cases} Cov\left(z_t, y_t\right) = 0 \\ Cov\left(z_t, x_t\right) = 0 \end{cases}$$
- Failure to reject H_0 is not evidence against the model. It is evidence that identification of β from z_t may be at fault.
- Under H_0 the model has no empirical content, and estimates of β are just the reflection of sample noise.
- Ideally, a confidence interval for β would reflect closeness to H_0, but it is based on large sample approximations which may be unreliable when instruments are weak.

- Suppose a relationship between three endogenous variables with instrument vector z_t

$$y_t = \alpha w_t + \beta x_t + u_t \quad Cov\left(z_t, u_t\right) = 0$$

or

$$Cov\left(z_t, y_t\right) - \alpha Cov\left(z_t, w_t\right) - \beta Cov\left(z_t, x_t\right) = 0. \quad (1)$$

- Now z_t need not be uncorrelated to the 3 variables for underidentification. Lack of correlation with two linear combinations of them is enough.

- To see this, write the null of underidentification as

$$H_0 : \begin{cases} Cov\left(z_t, y_t - \gamma_1 x_t\right) = 0 \\ Cov\left(z_t, w_t - \gamma_2 x_t\right) = 0 \end{cases}$$

- If H_0 holds, for any α^*

$$Cov\left(z_t, y_t - \gamma_1 x_t\right) - \alpha^* Cov\left(z_t, w_t - \gamma_2 x_t\right) = 0$$

or

$$Cov\left(z_t, y_t\right) - \alpha^* Cov\left(z_t, w_t\right) - \left(\gamma_1 - \gamma_2 \alpha^*\right) Cov\left(z_t, x_t\right) = 0$$

so that (1) is satisfied not only for the true values but also for any other α^* and β^* such that $\beta^* = \gamma_1 - \gamma_2 \alpha^*$.

- Under H_0 the observationally equivalent values of $\left(\alpha^*, \beta^*\right)$ are contained in the line $\beta^* = \gamma_1 - \gamma_2 \alpha^*$.

- i.e. β is identified given prior knowledge of α.

- When z_t is uncorrelated to the 3 variables, there is "2nd order underidentification". In such case, all values of α^* and β^* are observationally equivalent.

2. A GMM testing approach

- We adopt a GMM perspective and write the model as:
$$E(\Psi_t)\alpha = 0 \tag{2}$$
where Ψ_t is an $r \times (k+1)$ matrix constructed from data.
- α is a $k+1$ unknown parameter vector, which is identified subject to some normalization when rank of $E(\Psi_t)$ is k.
- We suppose the order condition is satisfied: $r \geq k$, but not necessarily the rank condition.
- When $r > k$ and the model is identified, it is *over-identified* and $E(\Psi_t)$ has reduced rank k (the null of J tests).
- We take as our null that there are multiple solutions to (2), which requires that $E(\Psi_t)$ has some prespecified reduced rank, usually $k-1$.
- In this case there exists another solution α^* not proportional to α such that
$$E(\Psi_t)\alpha^* = 0. \tag{3}$$
- We regard (2) and (3) as a *new augmented model*. If (α, α^*) satisfy the combined overidentifying moment restrictions, we conclude that the original model is not identified.
- Therefore, our approach is to test for underidentification by testing for overidentification in the augmented model using standard Sargan/Hansen J tests.

Two complications in this implementation:

(1) Normalization

- The augmented model is itself not identified. Further normalization of (α, α^*) is needed to extract linearly independent elements from the null space of $E(\Psi_t)$.
- Since the parameter estimates of the augmented model are of no particular interest, any convenient normalizations will suffice.
- It is known how to construct GMM estimators that are invariant to normalization.

(2) Redundancy

- The effective number of moment conditions may be less than $2r$. For some examples, some of the additional moments are simply duplicated in the augmented system.
- This is important to sort out in advance, since the degrees of freedom for the I test are given by the effective no. of moment conditions minus the effective no. of parameters.

3. Identification testing in a single equation

- Consider:

$$w_t' \alpha = u_t \tag{4}$$

where u_t is orthogonal to an $r \times 1$ vector z_t of instruments

$$E(z_t u_t) \equiv E(z_t w_t')\alpha = 0.$$

α (of order $k+1$) is identified up to scale iff

$$rank\ E(z_t w_t') = k. \tag{H_o^*}$$

- A standard Sargan test of overidentifying restrictions is a test of H_o^* against the alternative of no relationship.
- In contrast, an I test of underidentification tests the null

$$rank\ E(z_t w_t') = k - 1 \tag{H_o}$$

against the alternative of identification.

- Assume that $r \geq k$ but rank of $E(z_t w_t')$ is $k-1$. Then all parameter values compatible with $E\left(z_t u_t\right) = 0$ lie in a linear subspace of dimension 2, and the admissible equations are 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:

$$w_t' \alpha^* = v_t \tag{5}$$

combined with the orthogonality condition $E\left(z_t v_t\right) = 0$, and study the joint overidentification of the 2 equations.

- 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 I_2, which eliminates two parameters per equation.
- More generally, we can impose the proper normalizing restrictions $(\alpha, \alpha^*)'(\alpha, \alpha^*) = I_2$ and setting 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 $2(r - k) + 2$ overidentifying restrictions for the augmented system are rejected, then we have rejected the underidentification of the original econometric relation.
- In practice, it is desirable to construct a test statistic of underidentification using a version of the test of overidentifying restrictions that is invariant to normalization.
- For example, those based on continuously updated GMM (Hansen, Heaton, & Yaron, 1996) or empirical likelihood estimation (Imbens, 1997, Kitamura & Stutzer, 1997).

4. Underidentification of a Higher Order

- Although the null hypothesis that sets 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

$$E(z_t w_t')(\alpha, \alpha_1^*, ..., \alpha_j^*) = 0. \tag{6}$$

- If we impose $(j+1)^2$ normalizing restrictions on $(\alpha, \alpha_1^*, ..., \alpha_j^*)$ to avoid indeterminacy, 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 $(j+1)(r-k+j)$ overidentifying restrictions in (6) we test the null that α is underidentified of order j against the alternative of underidentification of order 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, as long as u_t is conditionally homoskedastic and serially uncorrelated, an asymptotic χ^2 test statistic of overidentifying restrictions with $r - k$ degrees of freedom is given by $T\lambda_1$, where

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

  is the smallest characteristic root of $W'Z\left(Z'Z\right)^{-1}Z'W$ in the metric of $W'W = \sum_{t=1}^{T} w_t w_t'$, etc. (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.
- It can be regarded as a continuously updated GMM test of overidentifying restrictions of the augmented model, 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 $A'W'WA = I_2$.

- This is given by

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

which coincides with $\lambda_1 + \lambda_2$.

- 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 rank of Π_2 is p, α is identified up to scale, but it is underidentified if rank is $p-1$ or less.

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

$$T[vec(\widehat{\Pi} - \Pi)]'V^{-1}vec(\widehat{\Pi} - \Pi) \tag{7}$$

  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 $vec(\widehat{\Pi})$.

- To relate (7) to our framework, write the augmented model 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, since $vec(\widehat{\Pi} - \Pi) = (B \otimes Z'Z)^{-1} \sum_{t=1}^{T} (u_t^\dagger \otimes z_t)$, (7) 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.

5. Taylor rules and Phillips curves

- We next illustrate the previous ideas by examining the empirical content of recently estimated forward looking versions of Taylor rules and Phillips curves.
- Since these estimates are subject to structural interpretation and have attracted much attention, our illustration has some substantive interest.

Forward looking monetary policy rules

- Clarida, Galí and Getler (*QJE*, 2000) estimate the following quarterly policy reaction function for the US Fed:

$$\begin{aligned} r_t^* - r^* &= \beta \left[E_t(\pi_{t+1}) - \pi^* \right] + \gamma E_t(x_{t+1}) \\ r_t &= \rho_1 r_{t-1} + \rho_2 r_{t-2} + \left[1 - (\rho_1 + \rho_2) \right] r_t^* \end{aligned}$$

- r_t, r_t^*: actual and target nominal Federal Funds rates
- π_t, π^*: actual (GDP deflator) and desired inflation rates
- x_t: output gap (as constructed by the CBO)
- Orthogonality conditions:

$$E \left[z_t \left(r_t - \phi_1 - \phi_2 \pi_{t+1} - \phi_3 x_{t+1} - \phi_4 r_{t-1} - \phi_5 r_{t-2} \right) \right] = 0$$

- Instrumental variables: 4 lags of r_t, π_t, x_t, commodity price inflation, M2 growth, and a “spread” variable
- Subsamples:
  - – 60Q1-79Q2: Pre-Volcker: $\beta < 1$?
  - – 79Q3-96Q4: Volcker and Greenspan: $\beta > 1$?

- $\widehat{\beta}$ is below unity for the pre-Volcker period, and much greater than one for Volcker–Greenspan (Table A).
- From this it is concluded that interest rate policy in the latter period has been much more sensitive to changes in expected inflation than in the former.
- The J test does not reject the overidentifying restrictions for any subsample.
- A possible augmented model is:

$$\begin{aligned} E\left[z_t\left(r_t - \gamma_{11} - \gamma_{12}x_{t+1} - \gamma_{13}r_{t-1} - \gamma_{14}r_{t-2}\right)\right] &= 0 \\ E\left[z_t\left(\pi_{t+1} - \gamma_{21} - \gamma_{22}x_{t+1} - \gamma_{23}r_{t-1} - \gamma_{24}r_{t-2}\right)\right] &= 0 \end{aligned}$$

- Since there are 11 other possible asymmetric normalizations, only normalization-invariant CU versions of the I test are reported.
- The I test fails to reject in both subsamples.
- Using a smaller number of lags as instruments, our conclusions are unaffected.

Table A
Forward Looking Taylor Rules
GMM Estimates from US Quarterly Data

	1960:1-1979:2		1979:3-1996:4	
	2S	CU	2S	CU
β	.834	.813	2.153	4.175
	(.067)	(.074)	(.379)	(.541)
γ	.274	.265	.933	.152
	(.087)	(.080)	(.454)	(.215)
J test (df)	13.08 (20)	12.22 (20)	21.38 (20)	16.15 (20)
p-value (%)	87.4	90.8	37.5	70.7
I test (df)		35.8 (42)		34.05 (42)
p-value (%)		73.8		80.4

NOTES: 1960:1-1979:2 is the pre-Volcker period.

1979:3-1996:4 corresponds to the Volcker-Greenspan era.

2S stands for two-step GMM, CU for continuously-updated GMM.

Asymptotic standard errors robust to heteroskedasticity shown in parentheses.

New Phillips Curves

- Galí, Gertler and López–Salido (*EER*, 2001) estimate forward looking Phillips curves based on marginal cost:
$$\pi_t = \mu + \delta E_t(\pi_{t+1}) + \lambda mc_t$$
and on detrended output:
$$\pi_t = \mu + \delta E_t(\pi_{t+1}) + \kappa \left(y_t - y_t^*\right)$$

- π_t: actual (GDP deflator) inflation rate.

- mc_t: average real marginal cost (measured by log real unit labour costs).

- $y_t - y_t^*$: output gap (measured by detrended GDP).

- GGL argue that the mc version is based on a theory of price setting by monopolistically competitive firms subject to constraints in the frequency of price adjustment.

- The fact that detrended output is not a good approximation to mc in the EU area may explain the empirical failure of the output gap-based Phillips curve.

- Orthogonality conditions:
$$E\left[z_{t-1}\left(\pi_t - \mu - \delta\pi_{t+1} - \lambda mc_t\right)\right] = 0$$

- Error: $\pi_{t+1} - E_t(\pi_{t+1})$ and measurement error in mc_t.

- Instrumental variables: 2 lags of mc_t, output gap and wage inflation, plus lags of π_t (5 Europe, 4 US).

- Sample: 70:Q1-98:Q2.

- Tables B and C: 2S GMM estimates of the mc and output gap versions. The finding that $\lambda > 0$ but $\kappa < 0$ appears to favor the mc version.

- But worrying discrepancies between 2S and CU, specially the estimated negative discount factors using CU.

- The J test never rejects the overidentifying restrictions.

- Augmented model: Order 1

$$\begin{aligned} E\left[z_{t-1}\left(\pi_t - \varphi_{11} - \varphi_{12} mc_t\right)\right] &= 0 \\ E\left[z_{t-1}\left(\pi_{t+1} - \varphi_{21} - \varphi_{22} mc_t\right)\right] &= 0 \end{aligned}$$

- Since there are 11 other possible normalizations, only CU versions of the I_1 test are reported.

- CU tests are robust to autocorrelation. Two bandwidth choices are reported: Newey-West with $\iota = 8$, and an automatic optimal choice (Andrews, 1991) based on optimal instruments.

- The I_1 test does not reject, so there seems to be again insufficient information in the instruments employed.

- We are very grateful to Jordi Galí and David López-Salido for kindly making available their datasets, and to Jesús Carro and Francisco Peñaranda for their excellent research assistance.

Comparisons with F tests

- F tests of significance of slope coefficients in the regression of π_{t+1} on the instruments are testing a subset of the restrictions in our I_2 test.
- The augmented model for order 2 underidentification is:

$$\begin{aligned} E\left[z_{t-1}\left(\pi_t-\mu_1\right)\right] &= 0 \\ E\left[z_{t-1}\left(\pi_{t+1}-\mu_2\right)\right] &= 0 \\ E\left[z_{t-1}\left(mc_t-\mu_3\right)\right] &= 0 \end{aligned}$$

- Since $\mu_1=\mu_2$, many moments in the first set are duplicated in the second, so the I_2 test should be based on:

$$E\left[\begin{array}{c} \left(\begin{array}{c} z_t \\ z_{t-1} \end{array}\right)\left(\pi_{t+1}-\mu_2\right) \\ z_{t-1}\left(mc_t-\mu_3\right) \end{array}\right]=0$$

- Rejection of these restrictions does not guarantee identification of the original parameters.
- The I_2 test fails to reject the null (Tables B and C).
- Table D reports F tests constructed under the null (LM) and the alternative (Wald).
- There is a stark contradiction in that the Wald versions reject but the LM versions do not.

Table B
Forward Looking Phillips Curves (Marginal Cost Version)
GMM Estimates from US and Euro Area Quarterly Data

	US			Euro Area		
	2S (ι=8)	CU (ι=8)	CU [opt. ι]	2S (ι=8)	CU (ι=8)	CU [opt. ι]
δ	.924	-4.679	.904	.914	-1.393	1.055
	(.030)	(2.319)	(.061)	(.041)	(.676)	(.076)
λ	.149	4.093	.178	.088	1.884	-.084
	(.075)	(2.319)	(.241)	(.042)	(.606)	(.077)
J test (df)	5.76 (8)	5.56 (8)	10.47 (8) [0]	8.21 (9)	5.20 (9)	10.20 (9) [3]
p-value (%)	67.4	69.6	23.3	51.3	81.7	33.5
I_1 test (df)		8.93 (18)	9.12 (18) [8]		10.02 (20)	8.30 (20) [11]
p-value (%)		96.1	95.7		96.8	99.0
I_2 test (df)		9.35 (24)	7.09 (24) [12]		10.68 (26)	10.18 (26) [10]
p-value (%)		99.7	99.9		99.7	99.7

NOTES: The sample period is 1970:1-1998:2.

2S stands for two-step GMM, CU for continuously-updated GMM.

HAC standard errors (triangular weights) shown in parentheses. Optimal value of ι shown in brackets.

Table C
Forward Looking Phillips Curves (Output Gap Version)
GMM Estimates from US and Euro Area Quarterly Data

	US			Euro Area		
	2S (ι=8)	CU (ι=8)	CU [opt. ι]	2S (ι=8)	CU (ι=8)	CU [opt. ι]
δ	1.012	1.013	1.019	.990	-.575	.962
	(.026)	(.025)	(.047)	(.018)	(.489)	(.027)
κ	-.021	-.018	-.024	-.003	-.282	.012
	(.007)	(.006)	(.010)	(.007)	(.170)	(.011)
J test (df)	5.06 (8)	5.03 (8)	8.11 (8) [1]	7.93 (9)	6.18 (9)	10.51 (9) [3]
p-value (%)	75.1	75.4	42.3	54.2	72.2	31.1
I_1 test (df)		8.70 (18)	4.85 (18) [22]		8.65 (20)	3.30 (20) [35]
p-value (%)		96.6	99.9		98.7	100
I_2 test (df)		9.9 (24)	5.32 (24) [19]		10.58 (26)	4.48 (26) [24]
p-value (%)		99.5	100		99.7	100

NOTES: The sample period is 1970:1-1998:2.

2S stands for two-step GMM, CU for continuously-updated GMM.

HAC standard errors (triangular weights) shown in parentheses. Optimal value of ι shown in brackets.

Table D

Forward Looking Phillips Curves

Robust GMM versions of F tests from US and Euro Area Quarterly Data

$$\pi_{t+1} = \beta' z_{t-1} + \varepsilon_t \quad H_0 : \beta' = (\mu_2, 0')$$

	US (df=10) (R^2=81.5%)			Euro Area (df=11) (R^2=87.7%)		
	ι=8	ι opt. H_0	ι opt. H_1	ι=8	ι opt. H_0	ι opt. H_1
F_T under H_0	9.06	5.29 [19]	13.78 [4]	10.07	4.79 [24]	19.18 [3]
p-value (%)	52.6	87.1	18.3	52.4	94.1	5.8
F_T under H_1	386.9	920.2 [19]	318.0 [4]	1289.8	3694.4 [24]	955.3 [3]
p-value (%)	0	0	0	0	0	0

NOTES: The sample period is 1970:1-1998:2.

Optimal value of ι shown in brackets.

Monte Carlo simulation of rejection rates

- Wald tests of predictability without strict exogeneity have a known tendency to over-reject in finite samples. We conducted a simulation exercise to further investigate this matter.
- Tables E and F report simulated rejection rates. The conclusion is that Wald over-rejects while LM and I_2 under-reject. However, with an automatic choice of ι (based on optimal instruments) LM and I_2 are more reliable than Wald.

Table E
Rejection rates in Monte Carlo experiments
Under the null of 2nd order underidentification
Nominal rejection rate: 5 percent
Robust GMM F tests

Bandwidth	Wald	LM
$\iota = 0$	10.9	4.6
$\iota = 1$	12.7	3.2
$\iota = 2$	14.7	2.3
$\iota = 3$	16.6	1.6
$\iota = 4$	18.8	1.0
$\iota = 8$	27.0	0.
automatic (H_1)	12.3	3.8
automatic (H_0)	12.3	3.8

NOTE: 10,000 replications, $T = 100$.

Table F

Rejection rates in Monte Carlo experiments

Under the null of 2nd order underidentification

Nominal rejection rate: 5 percent

I_2 tests

Bandwidth	CU	2S
$\iota = 0$	4.9	5.0
$\iota = 1$	1.0	1.0
$\iota = 2$	0.8	0.8
$\iota = 3$	0.5	0.5
$\iota = 4$	0.2	0.3
$\iota = 8$	0.	0.
automatic[1]	2.8	3.2

NOTE: 10,000 replications, $T = 100$.

[1]80.5% $\iota = 0$, 18.5% $\iota = 1$, 1% $\iota = 2$.

6. Cross-equation restrictions

- In the standard simultaneous equations system, we may test for identification equation by equation using the approach just described.
- 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 absence of cross-equation restrictions it seems only interesting to proceed with one equation at a time.
- 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 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 IVs appropriate for both equations:

$$E\left(z_t u_{jt}\right) = 0\ (j = 1, 2).$$

To test for underidentification in the 1st equation we would introduce a 2nd equation and 3 additional normalizations:

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

But for the two equation system, we do not want to augment each equation because in both cases we would arrive at the same 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 $\min\{2m, g\}$.

An Asset Pricing Model The example can be motivated in the GMM estimation of a standard consumption-CAPM.

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

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

x_t is the change in log consumption between $t-1$ and t,
y_{jt} is the return on the j-th financial asset in period t,
β is the coefficient of relative risk aversion,
and ρ is the discount factor.

There are m assets, and $E_{t-1}(.)$ 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

$$E_{t-1}(y_{jt} - \alpha_j - \beta x_t) = 0 \; (j = 1, ..., m), \qquad (8)$$

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 α_j and β is based on the unconditional moment restrictions:

$$E[z_t(y_{jt} - \alpha_j - \beta x_t)] = 0 \tag{9}$$

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 (9).

Empirical Application to US Data As an illustration we use US annual data on returns and consumption growth for 1889-1994 (as in Campbell, Lo and MacKinlay, 1997, 8.2).

The asset returns are:
(1) the real commercial paper rate, and
(2) the real stock return.

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.

In Table 1 we report two-step and continuously updated GMM estimates and test statistics (robust to heteroskedasticity but not to serial correlation) of the original and the augmented models.

There seems to be information in the instruments 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 1 percent, but not at 5 percent), and the estimated relative risk aversion parameter has the wrong sign.

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.

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

$$\begin{aligned} y_{1t} &= \alpha_1 + \beta[\mu + \delta(y_{1(t-1)} + x_{t-1} + w_{t-1})] + \varepsilon_{1t} \\ y_{2t} &= \alpha_2 + \beta[\mu + \delta(y_{1(t-1)} + x_{t-1} + w_{t-1})] + \varepsilon_{2t} \\ x_t &= \mu + \delta(y_{1(t-1)} + x_{t-1} + w_{t-1}) + \varepsilon_{3t} \\ w_t &= \pi w_{t-1} + \varepsilon_{4t}. \end{aligned}$$

This ensured that the original moments were satisfied with $z_t = (1, y_{1(t-1)}, x_{t-1}, w_{t-1})'$ as in the application.

We considered one experiment under the null 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 $\pi = 0.9$. Disturbances were generated as $N(0, I)$, and initial observations were obtained from the stationary distribution.

Table 2 shows some rejection frequencies for the (heteroskedasticity robust) two-step and continuously updated GMM versions of the I test.

Their behaviour is broadly the same, although the continuously updated test is slightly more conservative.

Size distortion in the experiment under the null is not negligible, as both tests show a tendency to under-reject.

The rejection frequencies under the alternative are at least 4 times those under the null, and give an idea of the power the test can be expected to have for small δ.

Table 2

Size and Power of the I Tests in the Asset Pricing Example

Rejection Frequencies (%) (df=9)

Nominal level	Under the null ($\delta = 0$) two-step	Under the null ($\delta = 0$) continuous-updating	Under the alternative ($\delta = .05$) two-step	Under the alternative ($\delta = .05$) 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, $T = 100$, $v_{it} \sim iid\, N(0, I)$.

A Translog Share Equation System Consider a four-input translog cost share equation system. After imposing homogeneity in prices and dropping one equation to take care of the adding-up condition:

$$\begin{aligned} y_{1t} &= \beta_1 p_{1t} + \beta_2 p_{2t} + \beta_3 p_{3t} + v_{1t} \\ y_{2t} &= \beta_4 p_{1t} + \beta_5 p_{2t} + \beta_6 p_{3t} + v_{2t} \\ y_{3t} &= \beta_7 p_{1t} + \beta_8 p_{2t} + \beta_9 p_{3t} + v_{3t} \end{aligned} \tag{10}$$

y_{jt}=cost share of input j, p_{jt}=log price of input j relative to the omitted input. We abstract from intercepts and log output terms since they have no effect here. The cost function implies 3 cross-equation symmetry constraints:

$$\begin{aligned} \beta_4 &= \beta_2 \\ \beta_7 &= \beta_3 \\ \beta_8 &= \beta_6. \end{aligned} \tag{11}$$

Moreover, we assume that prices are endogenous and that r instruments, denoted z_t, are available so that

$$E(z_t v_{1t}) = 0 \tag{12}$$

$$E(z_t v_{2t}) = 0 \tag{13}$$

$$E(z_t v_{3t}) = 0. \tag{14}$$

Without the symmetry restrictions, the order condition is satisfied if $r \geq 3$. With the symmetry restrictions, it appears the parameters may be just identified with $r = 2$, for in that case (12)-(14) is a system of 6 equations with 6 unknowns. But the system has reduced rank 5 by construction, so that the model is underidentified.

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

$$E[z_t(y_{1t} - \gamma_1 p_{2t} - \gamma_2 p_{3t})] = 0 \quad (15)$$

$$E[z_t(p_{1t} - \gamma_3 p_{2t} - \gamma_4 p_{3t})] = 0 \quad (16)$$

$$E[z_t(y_{2t} - \lambda_1 p_{2t} - \lambda_2 p_{3t})] = 0 \quad (17)$$

$$E[z_t(y_{3t} - \delta_1 p_{2t} - \delta_2 p_{3t})] = 0. \quad (18)$$

Note that since there are $4r$ orthogonality conditions and 8 parameters, with $r = 2$ the augmented set of moments does not introduce any overidentifying restrictions. Indeed, in the absence of symmetry restrictions, the moments (15)-(18) are satisfied by the original model (which is not identified). So there is nothing to test in (15)-(18).

In general, (15) and (16) imply that (12) is satisfied for any β_1^*, and for β_2^* and β_3^* such that

$$\begin{aligned} \beta_2^* &= \gamma_1 - \beta_1^* \gamma_3 \quad (19) \\ \beta_3^* &= \gamma_2 - \beta_1^* \gamma_4. \end{aligned}$$

Similarly, (16) and (17) imply that (13) is satisfied for any β_4^*, and for β_5^* and β_6^* such that

$$\begin{aligned} \beta_5^* &= \lambda_1 - \beta_4^* \gamma_3 \\ \beta_6^* &= \lambda_2 - \beta_4^* \gamma_4. \end{aligned}$$

Finally, (16) and (18) imply that (14) is satisfied for any β_7^*, and for β_8^* and β_9^* such that

$$\begin{aligned} \beta_8^* &= \delta_1 - \beta_7^* \gamma_3 \\ \beta_9^* &= \delta_2 - \beta_7^* \gamma_4. \end{aligned}$$

However, one restriction must be imposed on the coefficients in the augmented model for (15)-(18) to characterize observationally equivalent values of the original parameters satisfying the symmetry constraints. To see this, note that, subject to the cross-restrictions, (15)-(18) imply that (12)-(14) are satisfied as before for any β_1^* (and for β_2^* and β_3^* as in (19)), but only for $\beta_4^* = \beta_2^*$ so that

$$\beta_4^* = \gamma_1 - \beta_1^* \gamma_3,$$

and for β_5^* and β_6^* such that

$$\begin{aligned} \beta_5^* &= \lambda_1 - (\gamma_1 - \beta_1^* \gamma_3)\gamma_3 \\ \beta_6^* &= \lambda_2 - (\gamma_1 - \beta_1^* \gamma_3)\gamma_4. \end{aligned}$$

Equally, they are satisfied only for $\beta_7^* = \beta_3^*$ so that

$$\beta_7^* = \gamma_2 - \beta_1^* \gamma_4,$$

and for β_8^* and β_9^* such that

$$\begin{aligned} \beta_8^* &= \delta_1 - (\gamma_2 - \beta_1^* \gamma_4)\gamma_3 \\ \beta_9^* &= \delta_2 - (\gamma_2 - \beta_1^* \gamma_4)\gamma_4. \end{aligned}$$

Moreover, the restriction $\beta_8^* = \beta_6^*$ implies that the admissible values of the coefficients in the augmented model must satisfy for any β_1^*:

$$\delta_1 - (\gamma_2 - \beta_1^* \gamma_4)\gamma_3 = \lambda_2 - (\gamma_1 - \beta_1^* \gamma_3)\gamma_4$$

or

$$\delta_1 - \lambda_2 = \gamma_2 \gamma_3 - \gamma_1 \gamma_4. \qquad (20)$$

Therefore, the I test for this problem is a test of overidentifying restrictions based on the moments (15)-(18) subject to (20).

Enforcing (20) 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.

Note that when $r = 2$, the model's parameters are not identified, but it is still possible to test the restriction (20) as a specification test of the model.

Sequential Moments: Panel Data AR Models We now consider systems of equations in which the valid instruments differ for different equations.

A leading example is given by AR 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.

Consider first an AR(2) model with an individual specific intercept η_i:

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

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

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

We consider estimation of α_1 and α_2 based on a random sample of size N and the unconditional moment restrictions:

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

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 with 2 instruments so that (α_1, α_2) are just identified at most.

Testing for underidentification amounts to testing for overidentification the following 4 moments involving 2 unknown coefficients:

$$E\left[\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \otimes \begin{pmatrix} \Delta y_{i4} - \gamma_1 \Delta y_{i3} \\ \Delta y_{i3} - \gamma_2 \Delta y_{i2} \end{pmatrix}\right] = 0. \qquad (21)$$

If (21) holds, the original moments will hold not only for the true values (α_1, α_2), but also for any other (α_1^*, α_2^*) along the line $\alpha_2^* = \gamma_1\gamma_2 - \alpha_1^*\gamma_2$.

Note that if the AR(2) process contains a unit root so that $\alpha_1 + \alpha_2 = 1$, the moment conditions (21) hold with $\gamma_1 = \gamma_2 = -\alpha_2$.

With $T = 5$ a second equation and 3 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 (21) 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

$$E\left[y_i^{t-1}\left(\Delta y_{it} - \gamma_1 \Delta y_{i(t-1)}\right)\right] = 0, \; (t = 3, ..., T). \quad (22)$$

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 (22) holds.

Generalizing the previous argument, an I test for an AR(p) process with individual effects will be a test for overidentification based on

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

In particular, for an AR(1) process the relevant orthogonality conditions are

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

Empirical Illustration Table 3 shows parameter estimates, and I and J test statistics for an AR(2) model of employment using the Arellano & Bond (1991) dataset.

The data consists of an unbalanced panel of 140 quoted firms from the U.K. for which 7, 8, or 9 continuous annual observations are available for 1976-1984.

The AR(2) results were reported by Alonso-Borrego & 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 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.

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.

Monte Carlo Simulation We simulated 10,000 balanced panels of size $N = 150$ and $T = 7$ from the AR(2) model with $v_{it} \sim iid\, N(0, 1)$ and a unit root.

Specifically, we set the largest and smallest roots of the AR(2) polynomial to $\mu_1 = 1$ and $\mu_2 = 0.4$ respectively (the latter was chosen to mimic the estimated γ_1 from the empirical data).

To investigate local power we conducted another experiment with $\mu_1 = 0.98$, setting individual effects to zero.

Table 4 shows some rejection frequencies for the (heteroskedasticity robust) two-step and continuously updated versions of the test statistic.

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 χ^2 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.

Table 4

Size and Power of the I Tests in the Panel Example

Rejection Frequencies (%) (df=19)

Nominal level	Under the null ($\mu_1 = 1$) two-step	Under the null ($\mu_1 = 1$) continuous-updating	Under the alternative ($\mu_1 = .98$) two-step	Under the alternative ($\mu_1 = .98$) 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 $\mu_2 = 0.4$.

7. Conclusions

- 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.
- 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 Taylor and Phillips rules, intertemporal asset pricing models, and autoregressive models with individual effects for short panels.
- We 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 talk is how these procedures could be extended to testing for underidentification in nonlinear GMM problems.