Finite underidentification*

Enrique Sentana
CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain
<sentana@cemfi.es>

Revised: July 2019

Abstract

I adapt the Generalised Method of Moments to deal with nonlinear models in which a finite number of isolated parameter values satisfy the moment conditions. To do so, I initially study the closely related limiting class of first-order underidentified models, whose expected Jacobian is rank deficient but not necessarily 0. In both cases, my proposed procedures yield parameter estimators and underidentification tests within a standard asymptotically normal framework. I study models with and without separation of data and parameters. Finally, I illustrate the proposed inference procedures with a dynamic panel data model and a non-linear regression model for discrete data.

Keywords: Finite set, Generalised Method of Moments, Identification test.

JEL: C10

*This paper shares its origins with Arellano, Hansen and Sentana (2012). I am extremely grateful to Manuel Arellano and Lars Hansen for multiple discussions, suggestions and advice over the years we worked on extending Arellano, Hansen and Sentana (1999) to nonlinear situations. Gabriele Fiorentini, Arthur Lewbel and João Santos Silva were of great help with the examples mentioned in section 2. I have also benefited from the feedback provided by seminar participants at Alicante, Amsterdam, Austin, Boston University, CEMFI, Chicago, Harvard/MIT, Montreal, Penn and UCL, as well as conference audiences at ESEM 07 (Budapest) and CMES 18 (Shanghai). I thank Juan Carlos Ruíz García, Julio Gálvez and José Gutiérrez for able research assistance. Of course, the usual caveat applies. Financial support from the Spanish Ministries of Education & Science, Science & Innovation, and Economy & Competitiveness through grants SEJ 2005-08880, ECO 2008-00280 and ECO 2014-59262, respectively, is also gratefully acknowledged.

1 Introduction

Identification has been a central issue for the theory and practice of econometrics since at least the early analysis of simultaneous equations at the Cowles Commission (see e.g. Koopmans and Hood (1953)).[1] In the linear in parameters models of the form

$$E[\Psi(x)\alpha] = 0$$

considered by those authors, where x is a vector of observable random variables and $\Psi(x)$ contains $p \times (r+1)$ known functions of data, the observationally equivalent values of the $(r+1) \times 1$ unknown parameter vector α that satisfy the moment conditions above lie on either a one-dimensional linear subspace (the so-called **point identified** case), which nevertheless requires some additional normalisation to pin a unique direction down, or a higher-dimensional linear subspace (the **set identified** case).

However, in non-linear models defined by the unconditional moment conditions:

$$E\left[f\left(x;\beta\right)\right] = \bar{f}(\beta) = 0, \tag{1}$$

where $f(x,\beta)$ contains p influence functions and β is a vector of $k \leq p$ unknown parameters that lie on a subset $\mathbb{P}$ of $\mathbb{R}^k$, other meaningful underidentified situations may arise (see e.g. Fisher (1966) and Rothenberg (1971)):

a. **Uncountable underidentification**: There is a manifold of observationally equivalent values of β that satisfy the moment conditions (1).

b. **Countably infinite underidentification**: There is an infinite but countable number of observationally equivalent values of β that satisfy the moment conditions (1).

c. **Finite underidentification**: There is a finite number of observationally equivalent values of β that satisfy the moment conditions (1).

In addition, there exist other situations which share some underidentification features (see Sargan (1983a,b)):

[1] Throughtout this paper I focus on what Lewbel (2018) calls extremum-based parametric identification, which is linked to the estimation criterion used.

1. **First-order underidentification**: β_0 is the unique solution to (1), at least in an open neighbourhood of β_0, and therefore locally and possibly globally identified,[2] and yet $\text{rank}[\bar{D}(\beta)] < k$ at $\beta = \beta_0$ but not in its neighbourhood, where $\bar{D}(\beta) = E[\partial f(\beta)/\partial \beta']$ is the expected Jacobian of the moment conditions.

2. **Second-order underidentification**: β_0 is the only solution to (1), but $\text{rank}[\bar{D}(\beta)] < k$ **and** the rank of the expected Jacobian of the (vec) Jacobian is also deficient.

These borderline identified cases are closely related to the truly underidentified ones in **a.-c.** Specifically, Sargan (1983a) explained that if $\bar{f}(\beta)$ is analytic and there is rank failure for all higher-order Jacobians, then we go back to the uncountable underidentification in **a.** In turn, I will explain below that **1.** often arises when two observationally equivalent solutions in **c.** become arbitrarily close to each other.

The approach in this paper is closely related to Arellano, Hansen and Sentana (2012), who focused on uncountably underidentified models. They posed the problem as an estimation one where researchers seek to estimate the set over which identification is problematic. Specifically, they considered an augmented structural model in which the moment conditions are satisfied by a curve instead of a point, as in Sargan (1959). They then showed how to estimate the identified curve, providing an efficiency bound for any finite number of points along the curve in their Theorem 5.3. As a by-product, they obtained a test for underidentification by suitably testing for overidentification in the augmented model. If it is possible to estimate a curve without statistically rejecting the overidentifying restrictions of the augmented model, then researchers may conclude that the original econometric relation is uncountably underidentified. In contrast, rejections provide evidence that the original model is indeed point identified.

In this paper, I also impose an explicit structure on the lack of identification, which in turn leads to an alternative estimation problem and its associated underidentification test, but the difference is that I focus on situations in which only a finite number of locally

[2]Formally, β_0 will be locally identifiable if and only if $E[f(x;\beta_j)] \neq 0$ for any sequence $\{\beta_j\}$ such that $\lim_{j\to\infty} \beta_j = \beta_0$, while it will be globally identifiable if there is no observationally equivalent value anywhere in the admissible parameter space $\mathbb{P}$. The order condition $p = \dim(f) \geq \dim(\beta) = k$ provides a first-check of identification, but this is only necessary. A complement is provided by the rank condition: If $\bar{D}(\beta)$ is continuous at β_0, and $\text{rank}[\bar{D}(\beta_0)] = k$, then β_0 is locally identified. In contrast to the order condition, this condition is only sufficient. But if $\text{rank}[\bar{D}(\beta)]$ is also constant in a neighborhood of β_0, then the above rank condition becomes necessary too (see again Fisher (1966) and Rothenberg (1971)).

identified isolated parameter values satisfy (1), as in **c.** For simplicity, I only consider two-point sets, although the results could be extended to any finite number of points.[3]

Before studying such finite underidentified models, though, I study the closely related class of first-order underidentified ones. The reason is threefold. First, in a formal sense that I will characterise below, first-order underidentification can be regarded as the limiting case of finite underidentification when the isolated solutions converge to each other. Second, the methods I propose to deal with the former turn out to be very useful for the purposes of dealing with the latter. Finally, the behaviour of Generalised Method of Moment (GMM) estimators and hypothesis tests in first-order underidentified models has become the focus of increasing attention (see Kleibergen (2005) and Dovonon and Renault (2013)). As mentioned before, in those situations the expected Jacobian of the moment conditions is singular, but not necessarily zero, so that the usual asymptotic theory for standard GMM estimation breaks down. In contrast, the procedures that I propose will restore conventional GMM asymptotics, as the results in Lee and Liao (2018) confirm for the special case of a zero expected Jacobian.

The paper is also somewhat related to two different strands of the literature that have gained prominence in the last two decades. One is the weak instruments literature (see e.g. Stock, Wright and Yogo (2002) and Dufour (2003), or more recently Antoine and Renault (2010)). Papers in this tradition often consider a zero rank Jacobian $\bar{D}(\beta)$ at β_0 as the limit of a sequence of data generating models indexed by the sample size for the purposes of developing reliable standard errors and tests of hypothesis about β_0. By going to the limit and exploiting the additional moment conditions associated to a singular but not necessarily null Jacobian, I restore standard asymptotics. The other strand is the set estimation literature (see e.g. Chernozhukov, Hong and Tamer (2007) or more recently Yildiz (2012)), whose objective is to consistently estimate the set of values of β that satisfy (1). By making the additional assumption that the identified set is finite and modifying the usual GMM objective accordingly, I also estimate the set within a standard asymptotic framework.

Nevertheless, I would like to emphasise that the focus of this paper is more on the

[3]Extensions to countably infinite underidentification in **b.** are conceptually possible (see the rhumb line 3D example in Appendix B), but since to the best of my knowledge there are no interesting economic applications, I will not pursue them.

exploration of the sense in which identification may fail in a given context than on parameter estimation. In that regard, the asymptotically chi-square GMM overidentification restriction statistics that I propose provide natural diagnostics for finite underidentification in one case or first-order underidentification in the other. Furthermore, those tests will remain valid even if some of the maintained assumptions I use to obtain them do not hold. Specifically, the asymptotic distribution of the finite underidentification test that assumes there are only two solutions to (1) holds regardless of the exact number of actual solutions for the same reason that the usual overidentification test of those original moment conditions converges to a limiting chi-square in the presence of more than one locally identified solution. Similarly, the first-order underidentification test that assumes the nullity of the expected Jacobian is 1 will also remain asymptotically valid regardless of the number of zero singular values.

The rest of the paper is organised as follows. In section 2, I review some known situations in which there is either a finite set of observationally equivalent solutions or rank failure of the expected Jacobian in order to highlight the non-trivial features of the more subtle situations I am interested in. Then I study linear in variables but non-linear in parameter models in section 3 and fundamentally non-linear models in section 4. Finally, I conclude in section 5. Some additional details can be found in the appendices, which include 3D representations of all the different identification situations that might occur.

2 Some examples

There are well-known models which systematically give rise to two or more observationally equivalent solutions. The most obvious example is an MA(1) process whose parameters are estimated on the basis of first and second moments of the data. Another trivial example would be a non-linear regression model in which the conditional mean function contains the hyperbolic cosine function $\exp(\beta x)+\exp(-\beta x)$. Lewbel (2012) provides a more interesting example of a simultaneous equations system without exclusion restrictions identified though heteroskedasticity in which there are also two solutions to the moment equations, one positive and one negative. In these non-injective cases, one can suitably restrict the parameter space to achieve point identification. In addition, the two observationally equivalent solutions can be obtained automatically on the basis of one

another.

In other cases, there is generally a unique first-order identified solution, but if the unknown true parameter values satisfy certain restrictions, underidentification issues will arise. An interesting example is the so-called double-indexed model for non-negative data studied by Papadopoulos and Santos Silva (2012). In one of its simplest possible forms, this model parametrises the mean of a non-negative variable y conditional on two weakly exogenous variables x and z as follows:

$$E(y|x,z) = \frac{\exp[(\alpha+\gamma)x + \gamma z]}{1+\exp(\alpha x + \beta z)},$$

where α, β and γ are the parameters of interest. This conditional mean specification is compatible with a zero inflated Poisson model, a hurdle model, and a model with a latent error term for count data among several others (see Papadopoulos and Santos Silva (2012) and the references therein). Let $a(x,z)$ denote a $p \times 1$ vector of functions of x and z used to transform the conditional moment specification above into p unconditional orthogonality conditions in the usual way. If $p \geq 3$, then we will be able to identify α, β and γ provided the true value of β is different from 0. In contrast, if $\beta_0 = 0$ but $\alpha_0 \neq 0$, then Papadopoulos and Santos Silva (2012) point out that there will exist two observationally equivalent solutions: $(\alpha, 0, \gamma)$ and $(-\alpha, 0, \gamma+\alpha)$. Further, it is easy to prove that if $\alpha_0 = \beta_0 = 0$, then γ becomes first-order underidentified even though it is locally identified. In this model, though, those underidentification situations will arise not only asymptotically but also in any finite sample.

Another relatively unknown case is an AR(2) model observed subject to white noise errors, whose parameters are estimated on the basis of first and second moments of the data. When the latent AR(2) process is in fact an AR(1), its second AR root becomes first-order underidentified. Intuitively, the problem is that in a neighbourhood of the true value, the AR(2)+WN model is first-order equivalent to an ARMA(1,1)+WN model, whose parameters are only set identified. In this case, though, a reparametrisation which relies on the $\pm$ square root of the second AR root, as in Rotnitzky et al (2000)), restores standard $\sqrt{T}$ (half) Gaussian asymptotics (see Fiorentini and Sentana (2016)).

In this paper, in contrast, I am particularly interested in more subtle situations in which underidentification depends on parts of the data generating process (DGP) which are not necessarily specified by the moment conditions (1). In those cases, the relationship

between the two observationally equivalent solutions β and β^* or the conditions that lead to a singular expected Jacobian cannot simply be inferred from the true values of certain model parameters.

As in Arellano, Hansen and Sentana (2012), it is convenient to study separately non-linear in parameters but linear in variables models of the form $f(x,\beta) = \Psi(x)\phi(\beta)$, where $\phi(\beta)$ is a non-linear continuously differentiable function, and fundamentally non-linear models, in which no such separation of data and parameters is possible.

I will illustrate my proposed inference procedures with an example for each class:

1. A dynamic panel data model,

2. A non-linear dynamic regression model for discrete data.

For computational reasons, I systematically employ the optimal Continuously Updated GMM estimators (CUE) introduced by Hansen, Heaton and Yaron (1994). Thus, assuming $f(x;\beta)$ constitutes a martingale difference sequence, I can compute the CU-GMM criterion by regressing 1 on $f(x;\beta)$ with an OLS routine which is robust to potential singularities in the covariance matrix of the influence functions, as in Peñaranda and Sentana (2012). Although CUE is computationally more demanding, it is numerically invariant to normalisations, reparametrisations and parameter-dependent linear transformations of the moment conditions, which proves particularly useful in the context of underidentified models. In principle, though, other single-step GMM methods such as Empirical Likelihood or Exponentially Tilted could also be entertained.

3 Non-linear in parameter models

3.1 Theoretical discussion

As I mentioned in the previous section, these models are fully characterised by the fact that the influence functions are

$$f(x,\beta) = \Psi(x)\phi(\beta),$$

where $\Psi(x)$ contains $p \times (r+1)$ jointly Borel measurable functions of the observations and $\phi(\beta)$ is a non-linear, continuously differentiable function mapping $\beta : \mathbb{P} \to \mathbb{R}^{r+1}$ such

that $E[|f(x,\beta)|] < \infty$ for all β in the compact parameter space $\mathbb{P} \subseteq \mathbb{R}^k$. For simplicity of exposition, I assume that the observed sample is drawn from a stationary and ergodic stochastic process $\{x_t\}$.[4] In order for standard asymptotic results to apply, I also assume that the following high level regularity conditions hold as the sample size T goes to infinity:[5]

Assumption 1

$$\Psi_T = T^{-1} \sum\nolimits_{t=1}^{T} \Psi(x_t) \xrightarrow{a.s.} \bar{\Psi},$$

where $\bar{\Psi} = E[\Psi(x)]$ *is a non-stochastic* $p \times (r+1)$ *matrix, and*

$$\sqrt{T} vec(\Psi_T - \bar{\Psi}) \xrightarrow{d} N(\mathbf{0}, \mathcal{C}),$$

where $\mathcal{C}$ *is a non-stochastic* $p(r+1) \times p(r+1)$ *positive (semi)definite matrix.*

In this context, identification is only meaningful if $\phi(.)$ is an injective (i.e. one-to-one) function, for if there are two distinct parameter values β and β^* for which $\phi(\beta) = \phi(\beta^*)$, then it is clear *a priori* that one cannot identify β.

If the interest centred on the unrestricted estimation of $\alpha = \phi(\beta)$ instead of the restricted estimation of β, then the condition $\text{rank}(\bar{\Psi}) = r$ would be necessary and sufficient to identify $\alpha = \phi(\beta)$ up to a proportionality factor. Hence, identification problems may only arise if $\text{rank}(\bar{\Psi}) < r$. For obvious reasons, I rule out trivial problems by maintaining the assumption that $p \geq k$, so that the order condition is satisfied, but I also make the following stronger assumption:

Assumption 2 *For any two values of the parameter vector* $\beta \neq \beta^*$ *in* $\mathbb{P}$, $\phi(\beta) \neq \mathsf{c}\phi(\beta^*)$ *for some* $\mathsf{c} \in \mathbb{R}$.

While injectivity already rules out $\mathsf{c} = 1$, this assumption requires an implicit or explicit normalisation of the non-linear function $\phi(\beta)$ to eliminate scale multiples for $\mathsf{c} \neq 1$.

Suppose that theoretical considerations or previous empirical studies lead one to suspect that β may be first-order underidentified. Following Sargan (1983a), I initially simplify the presentation by assuming that the rank failure of the expected Jacobian is of

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

[5] See Newey and McFadden (1994) for an extensive discussion of more primitive conditions.

order one, and postpone the extension to situations in which its nullity is higher to the end of this subsection. For non-linear in parameters models, this amounts to

$$\bar{\Psi}\frac{\partial\phi(\beta)}{\partial\beta'}\gamma = 0, \tag{2}$$

at $\beta = \beta_0$, where $\gamma \in \mathbb{R}^k$ effectively determines the directional derivative along which the expected Jacobian is 0. On this basis, I can estimate both β and γ by optimally combining (2) with the original moment conditions (1) subject to a normalisation on γ such as $\gamma'\gamma = 1$. Thus, I can not only estimate the parameters of interest but also the "direction of weak identification". In some examples, though, the first-order underidentification problem may only affect a specific parameter, so I could restrict γ to be the corresponding canonical vector. In other cases, there may be a priori arguments for considering other pre-specified directional derivatives.

Given that the expected Jacobian of the joint set of moment conditions (1) and (2) is

$$\left\{ \begin{array}{cc} \bar{\Psi}[\partial\phi(\beta)/\partial\beta'] & 0 \\ \frac{\partial}{\partial\beta'}vec\left\{\bar{\Psi}[\partial\phi(\beta)/\partial\beta']\gamma\right\} & \frac{\partial}{\partial\gamma^{\dagger\prime}}vec\left\{\bar{\Psi}[\partial\phi(\beta)/\partial\beta']\gamma\right\} \end{array} \right\}, \tag{3}$$

where $\gamma^\dagger$ are the free elements of γ, this matrix must have full rank in a neighbourhood of the true values when $\gamma^\dagger$ is simultaneously estimated for standard GMM asymptotic theory to work. Similarly, when γ is fixed a priori, the first block of k columns of the above matrix must have full rank. But those rank conditions are precisely the second-order identification conditions mentioned in the introduction. Although in principle I could also consider second-order underidentified models, etc., in many locally identified but first-order underidentified examples the required rank condition on (3) holds.

After estimating β and possibly $\gamma^\dagger$ by optimal GMM, I can use the overidentification test of the augmented system (1) and (2) as a first-order underidentification test of the original moment conditions (1). The resulting test will have an asymptotic chi-square distribution with $2(p-k)+1$ degrees of freedom when the only restriction on γ affects its scale. If on the other hand γ is fixed a priori, then the number of degrees of freedom will be $2p-k$. I refer to both those tests as **first-order** I tests, because they provide an indication of the extent to which rank deficiency of the Jacobian should be a concern.

Suppose instead the original moment conditions (1) hold for β_0 and $\beta_0^* \neq \beta_0$. Then both $\phi(\beta_0)$ and $\phi(\beta_0^*)$ must belong to the null space of the matrix $\bar{\Psi}$, so that the system

of moment conditions

$$\bar{\Psi}[\phi(\beta), \phi(\beta^*)] = [0, 0] \tag{4}$$

evaluated at those two parameter values simultaneously holds. This system allows the joint estimation of the two observationally equivalent solutions. In particular, the optimal GMM estimators based on (4) will be asymptotically normal at the usual $\sqrt{T}$ rate subject to the first-order identifiability of β and β^*.[6] Finally, the usual overidentification test obtained after estimating β and β^* from (4) provides a test for the finite underidentification of (1). The rationale is straightforward. If one can find $\beta^* \neq \beta$ without statistical rejection, then the natural conclusion is that the identified set does indeed contain two points. But if the attempt fails statistically, then one may conclude β is globally identified. I refer to the resulting test as the **finite** I test. Standard GMM asymptotic theory implies that this I test will have an asymptotic chi-square distribution with $2(p-k)$ degrees of freedom if both β and β^* are first-order identified.

From a practical point of view, though, the main difficulty is ensuring that $\beta \neq \beta^*$, so that the duplicated moment conditions (4) do not effectively collapse to (1). Following Arellano, Hansen and Sentana (2012), in these non-linear in parameters models I can proceed as follows. I define the parameter space

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

and write each set of moment conditions as $\bar{\Psi}\alpha = 0$ for $\alpha \in \mathbb{Q}$. By assumption 2, the vectors $\phi(\beta)$ and $\phi(\beta^*)$ are not proportional. In addition, any linear combination of $\phi(\beta)$ and $\phi(\beta^*)$ must also belong to the null space of the matrix $\bar{\Psi}$.

I can then define the extended "linear subspace"

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

By playing around with c_1 and c_2, I can guarantee that the dimension of $\mathbb{Q}^*$ is always two. I discuss more practical details in the panel data example in section 3.2.

[6] Moreover, the joint estimator of β so obtained will be more efficient than a hypothetical GMM estimator based on the original moment conditions (1) which would somehow manage to restrict β to lie on a small neighbourhood of β_0, and the same applies to β^* (see section 2 of Arellano, Hansen and Sentana (2012) for a more formal argument). Given the block diagonality of the Jacobian matrix of (4), the hypothetical estimator based on (1) would only be as efficient as the joint estimator based on (4) in the highly unlikely situation in which the sample averages of the duplicated influence functions were asymptotically independent.

Importantly, if I reparametrise the model in terms of

$$\gamma^{\dagger} = \delta / \sqrt{\delta' \delta}$$

and

$$\eta = \sqrt{\delta' \delta},$$

where

$$\delta = \beta^{*} - \beta,$$

then I can equivalently re-write the duplicated moment conditions as

$$\bar{\Psi}\{\phi(\beta), \eta^{-1}[\phi(\beta + \eta\gamma^{\dagger}) - \phi(\beta)]\} = (0, 0)$$

for $\eta \geq \bar{\eta} > 0$. By l'Hôpital's rule, the limit as $\eta \to 0^{+}$ of a CU-GMM criterion based on these duplicated moment conditions will be the continuously updated GMM criterion based on

$$\bar{\Psi}\left\{\phi(\beta), [\partial\phi(\beta)/\partial\beta']\gamma^{\dagger}\right\} = 0$$

when β^{*} and β get closer and closer to each other in such a way that the dimension of the null space of $\bar{\Psi}$ remains two. Thus, first-order underidentified models can be formally interpreted as the limiting case of finite underidentified ones. The gain of one degree of freedom in the overidentifying test statistic simply reflects the fact that the CU-GMM criterion of the latter system is numerically invariant to the value of η.

Extensions to three or more isolated observationally equivalent solutions are straightforward by simply replicating the number of moment conditions in (4). Similarly, extensions to situations in which the nullity of the expected Jacobian is higher than 1 are also straightforward by replicating (2) in such a way that the corresponding $\gamma's$ provide a unique basis for its nullspace. Moreover, second-order underidentification situations could also be related to the limit of finite underidentification situations with three points when those three points become arbitrarily close, but I will not discuss those cases further in the interest of space.

3.2 Dynamic panel data

Consider the following univariate AR(2) model with individual specific intercepts

$$(Y_{it+2} - \eta_i) - \alpha_1(Y_{it+1} - \eta_i) - \alpha_2(Y_{it} - \eta_i) = v_{it+2}, \tag{5}$$

with

$$\begin{aligned} E(v_{it+2}|Y_{i1},...,Y_{it+1};\eta_i) &= 0, \\ V(v_{it+2}|Y_{i1},...,Y_{it+1};\eta_i) &= \sigma^2_{t+2}, \end{aligned} \tag{6}$$

where the expectations are taken by averaging across individuals, and (Y_{i1}, Y_{i2}, η_i) is a cross-sectionally *i.i.d.* random vector with bounded second moments, but no restrictions on the covariance between the unobserved effect η_i and the initial observations. I also assume the availability of a random sample of size N on $(Y_{i1},...,Y_{iT})$, with N large and $T \geq 4$ but negligible relative to N, leaving unspecified the temporal evolution of σ^2_{t+2}.[7]

The Arellano and Bond (1991) linear influence functions that eliminate the individual effects are

$$Y_{it-j}(\Delta Y_{it} - \alpha_1 \Delta Y_{it-1} - \alpha_2 \Delta Y_{it-2}) \quad j \geq 2, \quad t \geq 4. \tag{7}$$

This gives rise to a system of $T(T-3)/2$ moment conditions with two common coefficients and an increasing sequence of instruments, whose reduced form is non-standard.

If there are 5 or more time series observations, underidentification arises if and only if $\alpha_1 + \alpha_2 = 1$, so that the AR polynomial contains a unit root and ΔY_{it} follows an AR(1).[8] In that case, there will be an uncountable set of observationally equivalent solutions, all lying on the straight line $\alpha_2 = \gamma^2 - \gamma\alpha_1$, $\gamma \in \mathbb{R}$. Arellano, Hansen and Sentana (2012) show that this identified set can be efficiently estimated by applying optimal GMM to the AR(1) moment conditions

$$E[Y_{it-j}(\Delta Y_{it} - \gamma \Delta Y_{it-1})] = 0 \quad j \geq 1, \quad t \geq 2 \tag{8}$$

to infer γ. Moreover, the overidentification test of this system provides a **linear** I test.

But the Arellano and Bond (1991) conditions do not exploit all of the model restrictions. For that reason, Ahn and Schmidt (1995) proposed to combine (7) with the additional influence functions

$$(Y_{it+2} - \alpha_1 Y_{it+1} - \alpha_2 Y_{it})(\Delta Y_{it+1} - \alpha_1 \Delta Y_{it} - \alpha_2 \Delta Y_{it-1}) \tag{9}$$

[7] As Álvarez and Arellano (2004) forcefully argue, the dispersion of the cross-sectional distribution of errors at each period may change over time because of nonstationarity at the individual level or as a result of aggregate effects.

[8] When $T = 4$ identification problems may also arise even though no unit root exists (see Arellano, Hansen and Sentana (2012)).

to obtain more efficient estimators of α_1 and α_2 when the roots of the characteristic equations associated to (5) lie inside the unit circle. The question is whether these non-linear influence functions can rescue point identification in the unit root case. Given that γ will be uniquely identified from the Arellano, Hansen and Sentana (2012) moment conditions (8), it is convenient to express (9) in terms of α_1 and γ by replacing α_2 by $\gamma(\gamma-\alpha_1)$ so as to focus on the identification of α_1. In this way, I can write

$$\begin{aligned}
&[Y_{it}-\alpha_1(Y_{it-1}-\gamma Y_{it-2})-\gamma^2 Y_{it-2}][\Delta Y_{it-1}-\alpha_1(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})-\gamma^2\Delta Y_{it-3}]\\
&\qquad=\alpha_1^2(Y_{it-1}-\gamma Y_{it-2})(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})\\
&-\alpha_1[(Y_{it-1}-\gamma Y_{it-2})(\Delta Y_{it-1}-\gamma^2\Delta Y_{it-3})+(Y_{it}-\gamma^2 Y_{it-2})(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})]\\
&\qquad+(Y_{it}-\gamma^2 Y_{it-2})(\Delta Y_{it-1}-\gamma^2\Delta Y_{it-3})\quad t\geq 5. \qquad (10)
\end{aligned}$$

It turns out that heteroskedasticity matters, even though (6) is an aspect of the DGP deliberately left unspecified. In particular, if the cross-sectional variance of the innovations σ_t^2 varies freely over time, then the addition of the Ahn and Schmidt (1995) influence functions (10) to the moment conditions (8) will render α_1 first-order and therefore locally identified for $T\geq 6$. To understand why, it is convenient to compute the expected value of (10), which is given by

$$\sigma_{t-2}^2\alpha_1^2-[\sigma_{t-1}^2+(1+2\gamma)\sigma_{t-2}^2]\alpha_1+[(1+\gamma)\sigma_{t-1}^2+\gamma(1+\gamma)\sigma_{t-2}^2=0\ (t=5,\ldots,T),$$

where I have taken into account the unrestricted assumption about the initial conditions of the stochastic process for Y_{it}. This quadratic equation is clearly satisfied by $\alpha_1=1+\gamma$ for all t. For any specific t, though, there is a second solution given by

$$\alpha_{1,t}^*=\frac{\sigma_{t-1}^2}{\sigma_{t-2}^2}+\gamma.$$

However, when $T\geq 6$ this alternative solution is incompatible for different $t's$ unless

$$\frac{\sigma_{t-1}^2}{\sigma_{t-2}^2}=\frac{\sigma_{t-2}^2}{\sigma_{t-3}^2}=\alpha_1^*-\gamma\ (t=6,...,T). \qquad (11)$$

Therefore, if $T=5$ or if the cross-sectional variance of the innovations either grows or decreases exponentially over time, then α_1 will generally be first-order identified, but not globally identified, because there is a second solution

$$\begin{aligned}
\alpha_1^* &= \kappa+\gamma,\\
\kappa &= \frac{\sigma_{t+1}^2}{\sigma_{t+2}^2},
\end{aligned}$$

which satisfies the same moment conditions.

Further, given that the partial derivative of (10) with respect to α_1 will be[9]

$$2\alpha_1(Y_{it-1} - \gamma Y_{it-2})(\Delta Y_{it-2} - \gamma \Delta Y_{it-3})$$
$$-[(Y_{it-1} - \gamma Y_{it-2})(\Delta Y_{it-1} - \gamma^2 \Delta Y_{it-3}) + (Y_{it} - \gamma^2 Y_{it-2})(\Delta Y_{it-2} - \gamma \Delta Y_{it-3})], \qquad (12)$$

the expected Jacobian with respect to α_1 will be equal to 0 for

$$\alpha_{1,t}^{\circ} = \frac{1}{2}\left(\frac{\sigma_{t-1}^2}{\sigma_{t-2}^2} + 1 + 2\gamma\right)$$

under any form of time series heteroskedasticity, including (11), even though α_1° does not generally set to 0 the expected value of the Ahn and Schmidt (1995) influence functions (10). In fact, it is easy to see that $\alpha_{1,t}^{\circ} = .5(\alpha_1 + \alpha_{1,t}^{*})$, so that in the finite underidentified case the Jacobian rank deficiency will occur at the mid point between the two solutions.

Importantly, both $\alpha_{1,t}^{*}$ and $\alpha_{1,t}^{\circ}$ will converge to α_1 as $\sigma_{t-1}^2/\sigma_{t-2}^2 \to 1$, which means that if there is time series homoskedasticity (i.e. $\sigma_t^2 = \sigma^2 \ \forall t$), then α_1 will be globally identified as $1 + \gamma$, but it will become first-order underidentified.

In all cases, though, there is second-order identification because the quadratic nature of the $\phi(.)$ mapping implies that the Jacobian of the Jacobian of (10) with respect to α_1 will be proportional to

$$(Y_{it-1} - \gamma Y_{it-2})(\Delta Y_{it-2} - \gamma \Delta Y_{it-3})$$

for all α_1, whose expected value equals $2\sigma_{t-2}^2$ when the process contains a unit root.[10]

By combining the influence functions (10) and (12) with the moment conditions (8), I can efficiently estimate α_1 and γ (and therefore α_2), and obtain a **first-order** I test.

To deal with the finite underidentified case, I start by duplicating the Ahn and Schmidt (1995) influence function written in terms of α_1 and γ, which I then evaluate at α_1^{*}. To simplify the presentation, imagine γ is known. I can write the resulting system as

$$\bar{\Psi}[\phi(\alpha_1), \phi(\alpha_1^{*})] = [0, 0],$$

[9] In effect, this corresponds to a directional derivative along the line $\alpha_2 = \gamma^2 - \gamma\alpha_1$ in the original (α_1, α_2) space.

[10] Álvarez and Arellano (2004) state exactly the same underidentifiability conditions in the AR(1) version of model (5), while Bun and Kleibergen (2013) study the asymptotic distribution of the Ahn and Schmidt (1995) estimator in that case.

with

$$\begin{aligned}\Psi'(x) &= \left[\begin{array}{c}(Y_{it-1}-\gamma Y_{it-2})(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})\\ (Y_{it-1}-\gamma Y_{it-2})(\Delta Y_{it-1}-\gamma^2\Delta Y_{it-3})+(Y_{it}-\gamma^2 Y_{it-2})(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})\\ (Y_{it}-\gamma^2 Y_{it-2})(\Delta Y_{it-1}-\gamma^2\Delta Y_{it-3})\end{array}\right],\\ \phi'(\alpha_1) &= (\alpha_1^2,-\alpha_1,1).\end{aligned}$$

To keep the moments associated to α_1 and α_1^* apart, I generate the extended "linear subspace" $\mathbb{Q}^*$ by postmultiplying $[\phi(\alpha_1),\phi(\alpha_1^*)]$ by a 2×2 matrix $\mathbf{C}(\alpha_1,\alpha_1^*)$ in order to ensure that the nullity of $\bar{\Psi}$ is 2. In particular, if I choose $c_{11}=c_{22}=1$, $c_{12}=-1$ and $c_{21}=-\alpha_1/\alpha_1^*$ to avoid scale and rotation indeterminacies, I end up with

$$\bar{\Psi}\left(\begin{array}{cc}\alpha_1(\alpha_1-\alpha_1^*) & (\alpha_1^*-\alpha_1)(\alpha_1+\alpha_1^*)\\ 0 & \alpha_1-\alpha_1^*\\ (\alpha_1^*-\alpha_1)/\alpha_1^* & 0\end{array}\right)=0.$$

Dividing the first column by $(\alpha_1-\alpha_1^*)/\alpha_1^*$ and the second column by $(\alpha_1^*-\alpha_1)$ in an attempt to make sure $\alpha_1\neq\alpha_1^*$, I end up with the transformed influence functions

$$\begin{aligned}&\alpha_1\alpha_1^*(Y_{it-1}-\gamma Y_{it-2})(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})-(Y_{it}-\gamma^2 Y_{it-2})(\Delta Y_{it-1}-\gamma^2\Delta Y_{it-3}),\\ &(\alpha_1+\alpha_1^*)(Y_{it-1}-\gamma Y_{it-2})(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})\\ &-[(Y_{it-1}-\gamma Y_{it-2})(\Delta Y_{it-1}-\gamma^2\Delta Y_{it-3})+(Y_{it}-\gamma^2 Y_{it-2})(\Delta Y_{it-2}-\gamma\Delta Y_{it-3})],\end{aligned}\qquad(13)$$

which depend on the sum and product of the two solutions. In this context, I could estimate $\varsigma=\alpha_1+\alpha_1^*$ and $\pi=\alpha_1\alpha_1^*$, and then solve a simple quadratic equation to recover α_1 and α_1^*. Then, I could use the overidentification test of this system as a **finite** underidentification test. Such a test will reject with power equal to size for $T=5$ in the presence of a unit root because the relevant moment conditions will be jointly satisfied by $\alpha_1=1+\gamma_0$ and $\alpha_{1,5}^*=\gamma_0+\sigma_4^2/\sigma_3^2$. Exactly the same will happen for $T\geq 6$ if in addition (11) holds.

But this indirect procedure would occasionally lead to complex conjugate solutions for α_1 and α_1^*, in which case I should re-estimate subject to $\alpha_1=\alpha_1^*$. Although asymptotically this will happen with vanishing probability, in finite samples there is likely to be a "pile-up" problem, with a positive fraction of the samples yielding identical estimates for α_1 and α_1^*. As a result, the finite sample distribution of the **finite** I test may be somewhat distorted.

As expected from the discussion in section 3.1, I trivially recover from (13) the influence functions (10) and (12) associated to the first-order underidentified case when the two

separate solutions α_1 and α_1^* converge. The only difference is that there is an extra degree of freedom in the first-order underidentification test because of the restriction $\alpha_1 = \alpha_1^*$.

3.2.1 Simulation evidence

In this section I report the results of a limited Monte Carlo exercise with 2,500 replications of a Gaussian version of the AR(2) model with individual effects in (5) for a short panel of $T = 5$ time series observations and $N = 5,000$ cross-sectional units. Initially, I set the true values of the autoregressive parameters α_1 and α_2 to .3 and .7, respectively, so that the true value of γ is $-.7$. As for (6), I considered two values for the time-series heteroskedasticity "inflation" parameter $\kappa = \sigma_{t+1}^2/\sigma_{t+2}^2$:

1. $\kappa = 1$ (time-series homoskedasticity)

2. $\kappa = 1.1$ (time-series heteroskedasticity)

Importantly, I used the same underlying pseudo-random numbers in the different designs to minimise experimental error.

Starting with the homoskedastic case, the first thing to note is that the CU-GMM versions of the Arellano and Bond (1991) estimator and overidentifying restrictions test based on (7) are very unreliable in the presence of a unit root. Figure 1a displays the scatter plot of the CUEs of α_1 and α_2, which tend to lie along the line $\alpha_2 = .49 + .7\alpha_1$ but with a huge range of variation due to the lack of identification of the parameters (see Hillier (1990) for a discussion of the behaviour of symmetrically normalised estimators in underidentified single equation linear instrumental variable models). In turn, the size properties of the associated J test are summarised in Figure 1b using Davidson and MacKinnon's (1998) p-value discrepancy plot, which shows the difference between actual and nominal test sizes for every possible nominal size. In line with the theoretical results in Cragg and Donald (1993), underidentification in a linear in parameter model leads to substantial under-rejections for the overidentifying restriction test.

Figure 2a displays "bicorne" plots of the CUEs of α_1 and α_2 once I add the Ahn and Schmidt (1995) moment conditions (9).[11] As expected, the first-order underiden-

[11] These plots, which were introduced by Peñaranda and Sentana (2015) to characterise potentially asymmetric distributions with extreme tails, combine a doubly truncated non-parametric density estimate on top of a box plot. Therefore, the vertical lines describe the median and the first and third quartiles, while the length of the tails is one interquartile range.

tification of those parameters under time-series homoskedasticity leads to non-Gaussian distributions, with clearly visible but lower additional modes. In this case, though, the overidentification test, whose p-value plot is displayed in Figure 2b, shows substantial over-rejections, as expected from the results in Dovonon and Renault (2013).

In contrast, Figure 3 clearly indicates that the size of the first-order I test based on (8), (10) and (12) is very reliable.

Turning now to the design with time-series heteroskedasticity, Figure 4a confirms that the CUEs of α_1 and α_1^* based on (8) augmented with the influence functions (13) suffer from a pile-up problem, as there is a small fraction of them for which the two values coincide. In turn, this problem leads to some finite sample size distortions in the finite I test, as illustrated in Figure 4b. Those distortions disappear, though, as soon as I estimate the model in terms of $\varsigma = \alpha_1 + \alpha_1^*$ and $\pi = \alpha_1\alpha_1^*$, as shown in Figure 4c.[12]

To study the power of my proposed tests, I considered two additional designs. The first one is a persistent but covariance stationary homoskedastic model for $T = 5$ in which $\alpha_1 = .25$, $\alpha_2 = .7$, so that both the original Arellano and Bond (1991) and Ahn and Schmidt (1995) moment conditions are satisfied, but the expected values of the linear underidentification influence functions (10), the finite underidentification influence functions (13) or the first-order underidentification ones (12) are all different from 0 because the two AR roots are strictly inside the unit circle. Given that the linear I test in Arellano Hansen and Sentana (2012) rejected the null hypothesis with very high probability, it is perhaps not surprising that power remained close to one when I included either of the additional moment conditions that I have proposed.

In turn, the second design corresponds to a heterokesdatic unit root model for $T = 6$ in which the cross-sectional variance is $\sigma_t^2 = 1$ for all t except for $\sigma_5^2 = 1.2$, so that (11) does not hold. Unlike the previous design, in this first-order identified case the expected value of (10) is 0 but those of (12) and (13) are not. Once again, the power of both the first-order I test and especially the finite I test was also very high even though the linear I test has power equal to size in this case.

[12] All these Monte Carlo results may well extend to the ML estimators of panel data models in Álvarez and Arellano (2004), as well as to alternative GMM estimators which add the cross-sectional variances as additional exactly identified parameters. Validating such conjectures is left for further research.

4 Fundamental non-linearities

4.1 Theoretical discussion

Let $f(x,\beta)$ contain p influence functions jointly Borel measurable and twice continuously differentiable in their second argument for each value of x such that $E[|f(x,\beta)|] < \infty$ for every $\beta \in \mathbb{P}$, where β is a vector of $k \leq p$ unknown parameters that lie on the compact parameter space $\mathbb{P} \subseteq \mathbb{R}^k$.

The same basic approach I described in the previous section for non-linear in parameter but linear in variables models applies to fundamentally non-linear ones too. In the first-order underidentified case, inference will be based on the augmented set of moment conditions:

$$E\left[\begin{array}{c} f(x;\beta) \\ g(x;\beta,\gamma^{\dagger}) \end{array}\right] = 0, \tag{14}$$

where

$$g(x;\beta,\gamma^{\dagger}) = D(x;\beta)\gamma(\gamma^{\dagger}) = \frac{\partial f(x;\beta)}{\partial \beta'}\gamma(\gamma^{\dagger}),$$

and the free parameters that must be estimated are β and the "direction of weak identification" $\gamma^{\dagger}$, which corresponds to a basis of the null space of the expected Jacobian subject to some normalisation such as $\gamma'(\gamma^{\dagger})\gamma(\gamma^{\dagger}) = 1$.[13]

If I assume that

Assumption 3

$$\left[\begin{array}{c} \bar{f}_T(\beta_0) \\ \bar{g}_T(\beta_0,\gamma_0^{\dagger}) \end{array}\right] = T^{-1}\sum\nolimits_{t=1}^{T}\left[\begin{array}{c} f(x_t;\beta_0) \\ g(x;\beta_0,\gamma_0^{\theta\dagger}) \end{array}\right] \overset{a.s.}{\to} \left(\begin{array}{c} 0 \\ 0 \end{array}\right),$$

$$T^{-1}\sum\nolimits_{t=1}^{T}\left[\begin{array}{cc} \partial f(x;\beta_j^*)/\partial\beta' & 0 \\ \partial g(x;\beta_j^*,\gamma_j^{\dagger})/\partial\beta' & \partial g(x;\beta_j^*,\gamma_j^{\dagger})/\partial\gamma^{\dagger\prime} \end{array}\right] \overset{a.s.}{\to} \mathcal{J}_0$$

$$= E\left[\begin{array}{cc} \partial f(x;\beta_0)/\partial\beta' & 0 \\ \partial g(x;\beta_0,\gamma_0^{\dagger})/\partial\beta' & \partial g(x;\beta_0,\gamma_0^{\dagger})/\partial\gamma^{\dagger\prime} \end{array}\right]$$

for any sequence such that $(\beta_j^*,\gamma_j^{\dagger}) - (\beta_0^*,\gamma_0^{\dagger}) = o_p(1)$,

$$rank\left\{E\left[\begin{array}{cc} \partial f(x;\beta)/\partial\beta' & 0 \\ \partial g(x;\beta,\gamma^{\dagger})/\partial\beta' & \partial g(x;\beta,\gamma^{\dagger})/\partial\gamma^{\dagger\prime} \end{array}\right]\right\} = 2k-1$$

in an open neighbourhood of β_0 *and* $\gamma_0^{\dagger}$, *and*

$$\sqrt{T}\left[\begin{array}{c} \bar{f}_T(\beta_0) \\ \bar{g}_T(\beta_0,\gamma_0) \end{array}\right] \overset{d}{\to} N(\mathbf{0},\mathcal{I}_0),$$

where $\mathcal{I}_0$ *is a non-stochastic* $(p+r)\times(p+r)$ *positive definite matrix.*

[13] Once more, in some examples it may make sense to pre-specify the singularity direction γ.

then the optimal GMM estimators of based on β and $\gamma^\dagger$ will be consistently and asymptotically normal at the usual $\sqrt{T}$ rate.[14] Furthermore, the overidentification test associated to (14) will provide an asymptotically chi-square distributed test for first-order underidentification.

Similarly, in the finite underidentified case, inference will be based on the duplicated set of moment conditions

$$\begin{array}{c} E[f(x;\beta)] = 0, \\ E[f(x;\beta^*)] = 0. \end{array} \tag{15}$$

In this second instance, though, the main practical difficulty will be once more to keep β and β^* apart so that the duplicated moment conditions (15) do not collapse to (1). My proposed numerical device, which might be understood as a numerical implementation of L'Hôpital's rule, is as follows:

1. Reparametrise the model in terms of β, $\gamma^\dagger$ and η, where

$$\begin{aligned} \eta &= \sqrt{\delta'\delta}, \\ \delta &= \beta^* - \beta \end{aligned}$$

   and

$$\gamma^\dagger = \delta/\eta.$$

2. Replace the second influence function in (15) by

$$\eta^{-1}[f(x;\beta + \eta\gamma^\dagger) - f(x;\beta)].$$

3. Minimise a CU-GMM criterion function with respect to those new parameters subject to the restrictions $\eta \geq 0$ and $\gamma'(\gamma^\dagger)\gamma(\gamma^\dagger) = 1$.

4. If at any point during the minimisation algorithm η becomes smaller than some appropriately chosen threshold value $\bar{\eta}$, then replace the second estimating function by its first order approximation

$$g(x;\beta,\gamma^\dagger) = \frac{\partial f(x;\beta)}{\partial \beta'}\gamma(\gamma^\dagger).$$

   In practice, one should choose $\bar{\eta}$ so that the two CU-GMM criterion functions are numerically very close at $\bar{\eta}$ (see section 4.2.1 for further details).

[14] Similar assumptions are made by Kleibergen (2005), as well as by Dovonon and Gonçalves (2017) and Lee and Liao (2018) for the special case in which the expected Jacobian is equal to 0.

Importantly, step 4 is simply a trick to keep the two solutions apart in finite samples, but which becomes irrelevant asymptotically. The rationale is as follows. If the model is first-order identified at β and β^*, then the sample average of $\partial f(x;\beta)/\partial\beta' \cdot \gamma(\gamma^\dagger)$ will be numerically different from 0 with probability approaching 1 as the sample size increases. As a result, the GMM criterion function that uses the first-order approximation above will be large, and the optimisation routine will move away from the manifold $\beta = \beta^*$.

Once again, the procedure above confirms that first-order underidentified models may be regarded as limiting cases of finite underidentified ones because the criterion function associated to (15) converges to the GMM criterion for the original moment conditions augmented with their first derivatives, as in section 3.1. Nevertheless, one should always use the J statistic associated to (14) to test for first-order underidentification, which has one degree of freedom more than their finite underidentification counterparts because it sets η to 0 by construction.

4.2 A non-linear dynamic regression model for discrete data

Consider a Markov chain taking three different values: x_l, x_m and x_h. Suppose those values are of interest on their own. For example, x_t could be the dose of a drug taken by an addict at t. A researcher interested in predicting future drug consumption specifies the following fundamentally non-linear model

$$E(x_t^\beta | x_{t-1}) = \upsilon + \rho x_{t-1}, \tag{16}$$

where υ and ρ have the usual interpretation of intercept and slope of an autoregressive model, but they predict instead some power β of the observed variable.

For estimation purposes, the unconditional moment restrictions

$$E\left\{ (x_t^\beta - \upsilon - \rho x_{t-1}) \begin{bmatrix} \mathbf{1}(x_{t-1} = x_l) \\ \mathbf{1}(x_{t-1} = x_m) \\ \mathbf{1}(x_{t-1} = x_h) \end{bmatrix} \right\} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{17}$$

where $\mathbf{1}(.)$ is the usual indicator function, effectively contain the same information as the conditional moment restriction (16) because of the discrete, first-order Markovian nature of x_t.

In general, one would expect β to be point identified from those moment conditions. Nevertheless, as explained in Appendix A, it is possible to choose the transition matrix,

which is not an explicit part of model (16), so that these conditional moment restrictions, and therefore the unconditional moment restrictions (17), be satisfied by both β and $\beta^* \neq \beta$.

But even when there is a unique value of β that satisfies the original conditional moment restrictions (16), it is also possible to come up with transition matrices for which β_0 also satisfies

$$E(x_t^{\beta} \ln x_t | x_{t-1}) = 0, \tag{18}$$

which is the (conditional) expected value of the derivative of (16) with respect to β, so that the non-linearity parameter is first-order underidentified even though it is locally identified. In fact, those transition matrices naturally arise in the limiting case of $\beta^* = \beta$, exactly as in the dynamic panel data model (see again Appendix A for details).

The unconditional moment counterpart to (18) are

$$E\left\{ x_t^{\beta} \ln x_t \left[\begin{array}{c} \mathbf{1}(x_{t-1} = x_l) \\ \mathbf{1}(x_{t-1} = x_m) \\ \mathbf{1}(x_{t-1} = x_h) \end{array} \right] \right\} = \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right). \tag{19}$$

These moment conditions correspond to the ones Wright (2003) suggested to test for underidentification at a given point. Importantly, though, they must be combined with (17) to avoid estimating uninteresting values of β for which (19) holds but (17) does not (see Kleibergen (2005) for a related discussion in the case of LM tests). Specifically, it is easy to show that as in the panel data example, the expected Jacobian will become 0 at some intermediate point between β and β^* in finite underidentified cases, but those intermediate values will nevertheless fail to satisfy the original moment condition (16).

As previously explained, to keep β and β^* apart it is numerically convenient to combine the original unconditional moment conditions (17) with

$$E\left\{ \left(\frac{x_t^{\beta+\delta} - x_t^{\beta}}{\delta} \right) \left[\begin{array}{c} \mathbf{1}(x_{t-1} = x_l) \\ \mathbf{1}(x_{t-1} = x_m) \\ \mathbf{1}(x_{t-1} = x_h) \end{array} \right] \right\} = \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right), \tag{20}$$

which can be interpreted as the expected value of the relative (discrete) increment of $x_t^{\beta} - \upsilon - \rho x_{t-1}$ when one moves from β to $\beta^* = \beta + \delta$. The advantage of CUE is that the criterion function is the same whether one uses these moments or the original ones (17) evaluated at β^*. When $\delta \leq \bar{\delta}$, where $\bar{\delta}$ is a carefully chosen small but positive threshold value, one can safely replace (20) by (19), which are the moment conditions associated to

the Jacobian. Therefore, one set of moment conditions is the limiting case of the other, as expected from the theoretical discussion in section 4.1.

4.2.1 Simulation evidence

In this section I report the results of a limited Monte Carlo exercise with 2,500 replications of the discrete Markov chain model described in Appendix A for $T = 10,000$. To concentrate on the non-linear component of the model, which is characterised by the exponent parameter β, I keep υ and ρ fixed at their true values of .75 and .1, respectively. I initially considered two designs compatible with (16):

1. $\beta = 1$ and $\beta^* = 1.5$ (Finite underidentification)

2. $\beta = \beta_1^* = 1.5$ but with a 0 expected Jacobian (First-order underidentification)

Once again, I used the same underlying pseudo-random numbers in the different designs to minimise experimental error.

Starting with the finite underidentified design, the first thing to note is that the finite sample distribution of the CU-GMM estimator of β obtained from (17) seems to be a mixture of two Gaussian distributions, with two modes approximately equal to the values of β and β^* (see Figure 5a). By increasing the sample size to $T = 100,000$, as in Figure 5b, the separation of the sampling distribution into two Gaussian components becomes far more evident.

Somewhat surprisingly, though, the corresponding J test shows hardly any size distortion, as illustrated by Figure 5c. Intuitively, the reason is that a standard chi-square asymptotic distribution for Hansen's (1982) overidentifying restriction test requires that the criterion function is well behaved in the vicinity of a solution to the moment conditions. Given that in this design both β and β^* are first-order identified, J tests computed around each of those values will share exactly the same chi-square distribution, and the same is obviously true of their mixture regardless of the mixing proportion. Incidentally, the same argument implies that the finite underidentification test that I propose will also preserve its asymptotic chi-square distribution when there are more than two isolated solutions to the original moment conditions (17). As a result, one can use the finite I test as a diagnostic without necessarily knowing the exact number of possible solutions. For

analogous reasons, the first-order I test that assumes the nullity of the expected Jacobian to be 1 will also remain valid regardless of the number of zero singular values.

The CUEs of β and δ displayed in Figure 6a, obtained by combining the moment conditions (17) and (20) for $\delta \geq \bar{\delta} = 10^{-7}$, are also well behaved, although there is again some evidence of a pile-up problem, which in this case manifests itself by a non-negligible fraction of zero δ estimates. In addition, there is a strong negative correlation between the estimates of β and δ, as illustrated by the scatter plot in Figure 6b. To a large extent, this negative correlation reflects the rather elongated shape of the contours of the CU-GMM criterion function around its minimum, which are shown for a particular simulation in Figure 6c. In contrast, the estimate of β that exclusively relies on the moment conditions (19), and therefore ignores the original moment conditions (17), turns out to be centred around a pseudo-true value which roughly lies half way between 1 and 1.5, as shown in Figure 7.

Turning now to the first-order underidentified design, Figure 8a confirms that the finite sample distribution of the CU-GMM estimator of β obtained from (17) is clearly non-normal, with a distinctive but lower second mode. Similarly, Figure 8b indicates that the associated overidentification test shows substantial over-rejections, which is once again expected from the results in Dovonon and Renault (2013). In contrast, Figure 9 suggests that the finite sample distribution of the CUE of β obtained by combining the moment conditions (17) and (19) is nicely behaved around the true value of 1.5.

To study the power of my proposed tests, I finally generated data using a transition matrix for which there is a single value of β that satisfies (17) but does not satisfy (19) (see again Appendix A for details). The simulation results, which are available on request, show that both the first-order I test and the finite I test had power close to 1 for small deviations from (19). Moreover, the usual CU version of the overidentification test based on the original moment conditions (17) closely followed a chi-square distribution with the right number of degrees of freedom.

5 Conclusions

In linear econometric models parameters are either point identified, or set identified, but in the latter case the set of observationally equivalent structures is necessarily un-

countable. In non-linear models, in contrast, it is possible that only a finite number of distinct parameter values satisfy the original moment conditions. Further, another possibility is that the parameters are globally identified but the expected Jacobian is of reduced rank.

In this paper, I consider the estimation of observationally equivalent parameters in the finite underidentified case. To do so, I map this situation into a standard GMM problem by replicating the original moment conditions and evaluating each of them at different values of β. Given that the Jacobian is block diagonal, I can rely on standard asymptotic theory for GMM under the maintained assumption that each of the points is first-order identified. The main difficulty consists in keeping the solutions apart. The approach simplifies considerably for non-linear in parameters models, in which the set of observationally equivalent structures must belong to some restricted "linear" subspace. But in general, I achieve separation by working with discrete counterparts to directional derivatives.

I also discuss the estimation of the locally identified parameters in the first-order underidentified case. Once again, I restore standard asymptotics by combining the original moment conditions with the moment conditions associated to the rank failure of the Jacobian. I consider two different possibilities of practical interest, depending on whether a basis of the null space of the expected Jacobian is known.

Associated with the asymptotically normal estimators that I propose, the usual GMM overidentification restriction statistics of the augmented moment conditions provide an indication of the extent to which either the existence of multiple solutions to the original moment conditions or rank deficiency of the Jacobian should be a concern. Somewhat surprisingly, the asymptotic distribution of the finite underidentification test that assumes there are two solutions to the original moment conditions remains valid regardless of the exact number of actual solutions for the same reason the usual overidentification test of the original moment conditions remains valid in the presence of more than one solution. Similarly, the first-order underidentification test that assumes the nullity of the expected Jacobian is 1 will also remain valid regardless of the number of zero singular values.

Importantly, I explicitly relate the finite and first-order underidentified cases by showing that as two solutions of the original moments converge to each other, their duplicated

influence functions become equivalent to an extended system which combines the original moment conditions with their directional Jacobian.

I illustrate the proposed procedures with two examples. The first one is the linear in variables but non-linear in parameters autoregressive dynamic panel data model studied by Arellano and Bond (1991), Ahn and Schmidt (1995) and many others. As shown by Arellano, Hansen and Sentana (2012), linear GMM estimators of this model can only estimate an uncountable set of observationally equivalent parameter configurations when the autoregressive polynomial contains a unit root. The inclusion of non-linear moment conditions dramatically changes the nature of the underidentification problem in those circumstances, rendering the model parameters either first-order underidentified, locally but not globally identified, or fully identified, depending on the temporal evolution of the cross-sectional variance of the innovations, which is not an explicit part of the model.

The second example is a non-linear dynamic regression model for discrete data in which there is no separation between variables and parameters. Again, depending on the properties of the transition matrix, which is not explicitly modelled, the regression parameters can be first-order underidentified, locally but not globally identified or fully identified.

In both cases, the simulation results share the following features:

1) There is a pile-up problem in finite unidentified models, whereby a positive fraction of the estimates end up collapsing to a single solution.

2) The sampling distribution of the estimators of first-order underidentified models is not unimodal, with additional lesser modes around alternative "false" parameter values.

This blended behaviour reflects the fact that the criterion function used to deal with the first-order underidentified case is the limit of the criterion function used to deal with the finite underidentified one. Therefore, one could argue that first-order underidentification is not just a situation in which a standard regularity condition fails, but more fundamentally, one in which identification is dubious.

An important topic for future research would be to derive finite sample results that confirm the close relationship between finite and first-order identification. It would also be very interesting to explore other more complex empirically relevant models for which finite underidentification and its first-order underidentification limit represent important

concerns in practice. One such example is the following stylised social interactions model for N individuals in a network:

$$y = \rho W y + \beta x + \gamma W x + e, \tag{21}$$

where y and x are the $N \times 1$ vectors that contain the values of the endogenous and exogeneous variables, respectively, for each of those individuals, W is the so-called adjacency matrix of the underlying network, ρ, β and γ are parameters of interest measuring direct and indirect effects, and e contains the N structural residuals. Bramoullé, Djebbari and Fortin (2009) study the local identification of this model when W is known, while de Paula, Rasul and Souza (2018) extend their results to the case in which the non-diagonal elements of W are also estimated. Specifically, theorem 1 in de Paula, Rasul and Souza (2018) shows that under certain conditions the model parameters are first-order identified but not necessarily globally identified, while their theorem 2 shows that there will be at most one solution with $\rho\beta + \gamma > 0$ and another one with $\rho\beta + \gamma < 0$. I am currently exploring the application of the tests I have proposed in this paper to model (21).

References

Ahn, S.C. and Schmidt, P. (1995): "Efficient estimation of models for dynamic panel data", *Journal of Econometrics* 68, 5-27.

Álvarez, J. and Arellano, M. (2004): "Robust likelihood estimation of dynamic panel data models", CEMFI Working Paper 0421.

Antoine, B. and Renault, E. (2010): "Efficient inference with poor instruments: a general framework", chapter 2 in A. Ullah and D. Giles (eds.) *Handbook of Empirical Economics and Finance*. Chapman and Hall.

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

Arellano, M. and Bond, S.R. (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., Hansen, L.P. and Sentana, E. (1999): "Underidentification?", mimeo, CEMFI.

Arellano, M., Hansen, L.P. and Sentana, E. (2012): "Underidentification?", *Journal of Econometrics* 170, 256–280.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009): "Identification of peer effects through social networks", *Journal of Econometrics* 150, 41-55.

Bun, M. and Kleibergen, F. (2013): "Identification and inference in moments based analysis of linear dynamic panel data models", mimeo, University of Amsterdam.

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

Cragg, J.G. and S.G. Donald (1993): "Testing identifiability and specification in instrumental variable models", *Econometric Theory* 9, 222-240.

Davidson, R. and MacKinnon, J.G. (1998): "Graphical methods for investigating the size and power of tests statistics", *The Manchester School* 66, 1-26.

de Paula, A., Rasul, I. and Souza, P. (2018): "Recovering social networks from panel data: identification, simulations and an application", CEPR Discussion Paper 12,792.

Dovonon, P. and Gonçalves, S. (2017): "Bootstrapping the GMM overidentification test under first-order underidentification", *Journal of Econometrics* 201, 43–71.

Dovonon, P. and Renault, E. (2013): "Testing for common conditionally heteroskedas-

tic factors", *Econometrica* 81, 2561-2586.

Dufour, J.-M. (2003): "Identification, weak instruments and statistical inference in econometrics", *Canadian Journal of Economics* 36, 767–808.

Fiorentini, G. and Sentana, E. (2016): "Neglected serial correlation tests in UCARIMA models", *SERIEs* 7, 121-178.

Fisher, F.M. (1966): *The Identification Problem in Econometrics*. McGraw-Hill.

Kleibergen, F. (2005): "Testing parameters in GMM without assuming that they are identified", *Econometrica* 73, 1103-1124.

Koopmans, T.C. and Hood, W.C. (1953): "The estimation of simultaneous linear economic relationships", chapter 6 in W.C. Hood and T.C. Koopmans (eds.) *Studies in Econometric Method*. Cowles Commission Monograph 14. Wiley.

Hansen, L.P. (1982): "Large sample properties of generalized method of moments estimators", *Econometrica* 50, 1029–1054.

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 University Press.

Hillier, G.H. (1990): "On the normalization of structural equations: properties of direct estimators", *Econometrica* 58, 1181–1194.

Lee, J.H. and Liao, Z. (2018): "On standard inference for GMM with local identification failure of known forms", *Econometric Theory* 34, 790-814.

Lewbel, A. (2012): "Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models", *Journal of Business and Economic Statistics* 30, 67-80.

Lewbel, A. (2018): "The identification zoo: meanings of identification in econometrics", forthcoming in the *Journal of Economic Literature*.

Newey, W.K. and McFadden, D.L. (1994): "Large sample estimation and hypothesis testing", chapter 36 in R.F. Engle and D.L. McFadden (eds.) *Handbook of Econometrics* IV. Elsevier.

Papadopoulos, G. and Santos Silva, J.M.C. (2012): "Identification issues in some double-index models for non-negative data", *Economics Letters* 117, 365–367.

Peñaranda, F. and E. Sentana (2012): "Spanning tests in portfolio and stochastic

discount factor mean-variance frontiers: a unifying approach", *Journal of Econometrics* 170, 303–324.

Peñaranda, F. and E. Sentana (2015): "A unifying approach to the empirical evaluation of asset pricing models", *Review of Economics and Statistics* 97, 412–435.

Rothenberg, T.J. (1971): "Identification in parametric models", *Econometrica* 39, 577-592.

Rotnitzky, A., Cox, D. R., Bottai, M. and Robins, J. (2000): "Likelihood-based inference with singular information matrix", *Bernoulli* 6, 243-284.

Sargan, J.D. (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, 1605-1634.

——— (1983b). "Identification in models with autoregressive errors," chapter 8 in S. Karlin, T. Amemiya, and L. A., Goodman (eds.) *Studies in Econometrics, Time Series, and Multivariate Statistics.* Academic Press.

Stock, J.H., Wright, J.H. and Yogo, M. (2002): "A survey of weak instruments and weak identification in generalized method of moments", *Journal of Business and Economic Statistics* 20, 518–29.

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

Yildiz, N. (2012): "Consistency of plug-in estimators of upper contour and level sets", *Econometric Theory* 28, 309-327.

Appendices

A Discrete Markov chain

As is well known, the transition matrix

$$\begin{array}{cccc} x_t \diagdown x_{t-1} & x_l & x_m & x_h \\ x_l & \pi_l(x_l) & \pi_l(x_m) & \pi_l(x_h) \\ x_m & 1-\pi_l(x_l)-\pi_h(x_l) & 1-\pi_l(x_m)-\pi_h(x_m) & 1-\pi_l(x_h)-\pi_h(x_m) \\ x_h & \pi_h(x_l) & \pi_h(x_m) & \pi_h(x_h) \end{array}$$

fully characterises the serial dependence of x_t assuming strict stationarity for the chain. Further, the unconditional probabilities (π_l, π_m, π_h) coincide with the eigenvector associated to the unit eigenvalue normalised so that its coefficients add up to 1.

In order for two different sets of parameter values to satisfy the conditional moment restrictions (16), it must be the case that $E(x_t^{\beta}|x_{t-1}=x_j) = E(x_t^{\beta^*}|x_{t-1}=x_j)$, $j=l,m,h$.

Assuming for scaling purposes that $x_m = 1$, these equalities are equivalent to

$$\begin{aligned} 1-\pi_l(x_l)-\pi_h(x_l)+x_l^{\beta}\pi_l(x_l)+x_h^{\beta}\pi_h(x_l) &= 1-\pi_l(x_l)-\pi_l(x_l)+x_l^{\beta^*}\pi_l(x_l)+x_h^{\beta^*}\pi_h(x_l), \\ 1-\pi_l(x_m)-\pi_h(x_m)+x_l^{\beta}\pi_l(x_m)+x_h^{\beta}\pi_h(x_m) &= 1-\pi_l(x_m)-\pi_l(x_m)+x_l^{\beta^*}\pi_l(x_m)+x_h^{\beta^*}\pi_h(x_m), \\ 1-\pi_l(x_h)-\pi_h(x_h)+x_l^{\beta}\pi_l(x_h)+x_h^{\beta}\pi_h(x_h) &= 1-\pi_l(x_h)-\pi_l(x_h)+x_l^{\beta^*}\pi_l(x_h)+x_h^{\beta^*}\pi_h(x_h). \end{aligned}$$

Straightforward algebra shows that these conditions will be simultaneously satisfied when

$$\frac{\pi_h(x_l)}{\pi_l(x_l)} = \frac{\pi_h(x_m)}{\pi_l(x_m)} = \frac{\pi_h(x_h)}{\pi_l(x_h)} = \frac{x_l^{\beta}-x_l^{\beta^*}}{x_h^{\beta^*}-x_h^{\beta}} = s(x_l, x_h, \beta, \beta^*) \geq 0. \tag{A1}$$

With this restriction, it is easy to see that the conditions

$$0 \leq \pi_l(x_l), \pi_l(x_m), \pi_l(x_h) \leq \frac{1}{1+s(x_l, x_h, \beta, \beta^*)} \leq 1 \tag{A2}$$

guarantee the admissibility of the conditional probabilities of $x_t = x_m$ because in that case

$$\pi_m(x_{t-1}) = 1-\pi_h(x_{t-1})-\pi_l(x_{t-1}) = 1-[1+s(x_l, x_h, \beta, \beta^*)]\pi_l(x_{t-1})$$

will be between 0 and 1 for all three values of x_{t-1}.

The last restriction to impose is precisely the conditional moment restriction (16). Given that (A1) implies that

$$E(x_t^{\beta}|x_{t-1}) = 1+[-1+x_l^{\beta}+(x_h^{\beta}-1)s(x_l, x_h, \beta, \beta^*)]\pi_l(x_{t-1}), \tag{A3}$$

by assuming that $\pi_l(x_{t-1}) = a + bx_{t-1}$ for values of a and b that satisfy (A2), then it is easy to check that (16) will hold with

$$\upsilon = 1 + [-1 + x_l^{\beta} + (x_h^{\beta} - 1)s(x_l, x_h, \beta, \beta^*)]a = 1 + [-1 + x_l^{\beta^*} + (x_h^{\beta^*} - 1)s(x_l, x_h, \beta, \beta^*)]a$$

and

$$\rho = [-1 + x_l^{\beta} + (x_h^{\beta} - 1)s(x_l, x_h, \beta, \beta^*)]b = [-1 + x_l^{\beta^*} + (x_h^{\beta^*} - 1)s(x_l, x_h, \beta, \beta^*)]b,$$

which remain point identified.

In contrast, the first-order underidentified case requires that $E(x_t^{\beta} \ln x_t | x_{t-1} = x_j) = 0$, $j = l, m, h$, which is equivalent to

$$\begin{aligned} x_l^{\beta} \ln(x_l)\pi_l(x_l) + x_h^{\beta} \ln(x_h)\pi_h(x_l) &= 0, \\ x_l^{\beta} \ln(x_l)\pi_l(x_m) + x_h^{\beta} \ln(x_h)\pi_m(x_m) &= 0, \\ x_l^{\beta} \ln(x_l)\pi_l(x_h) + x_h^{\beta} \ln(x_h)\pi_h(x_h) &= 0. \end{aligned}$$

But these conditions will also be simultaneously satisfied when (A1) holds with $\beta^* = \beta$ as long as $s(x_l, x_h, \beta, \beta^*)$ is replaced by $(x_l/x_h)^{\beta} \ln(x_l/x_h)$. Therefore, the first-order underidentified case can once again be understood as the limiting case of the finite underidentified case as $\beta^* \to \beta$.

Finally, I generate data of the standard situation with a single, first-order identified solution by slightly modifying the design of the first-order underidentified case. Specifically, I add a small value Δ to $\pi_l(x_m)$ so that when I impose that

$$\pi_h(x_m) = \frac{\upsilon + \rho x_m - x_m^{\beta} + (x_m^{\beta} - x_l^{\beta})[\pi_l(x_m) + \Delta]}{(x_m^{\beta} - x_h^{\beta})}$$

to guarantee that the original moment conditions (17) are satisfied, both $\pi_h(x_m)$ and $\pi_m(x_m) = 1 - \pi_l(x_m) - \Delta - \pi_h(x_m)$ remain between 0 and 1. By construction, this design will also converge to the first-order identified case when Δ goes to 0.

B Underidentification in 3D

B.1 Linear model

Consider a model characterised by the moment conditions

$$E(\Psi_t)\varphi = 0 \tag{B4}$$

where φ is a $k \times 1$ parameter vector and Ψ_t is a $r \times k$ data matrix, with $r \geq k$. The archetypal example is the single equation, linear instrumental variables model, in which a certain linear combination of the variables of interest, namely $x_t'\varphi$, is assumed orthogonal to the vector of r instruments z_t. As a result, $\Psi_t = z_t x_t'$.

To provide a graphical representation of the different identification possibilities of the model above, I assume that $k = 3$.

If $rank[E(\Psi_t)] = 3$, then there is no value of φ other than $\varphi = 0$ that can satisfy the above moment conditions. In that case, we say that (B4) is rejected.

If on the other hand $rank[E(\Psi_t)] = 2$, then there is a linear subspace of dimension one which satisfy those restrictions. Algebraically, this subspace coincides with the nullspace of the matrix $E(\Psi_t)$. Graphically, this nullspace is a straight line in $\mathbb{R}^3$ that goes through the origin.

I can choose a single point on that line by either fixing one coordinate of φ to 1, which works provided its true value is not 0, or by imposing the symmetric normalisation restriction $\varphi'\varphi = 1$, which effectively defines a direction by forcing φ to lie on the surface of the unit sphere in $\mathbb{R}^3$. However, since any line through the origin will intersect the unit sphere twice, I need an additional "sign" restriction. For example, I could impose that the third coordinate be non-negative, which effectively restricts the solution to be on the Northern hemisphere. This works in all cases except when the line intersects the sphere exactly at the equator, in which case I should require the second coordinate to be non-negative, which forces the solution to be on the Eastern hemisphere instead.

Despite this scale indeterminacy, we normally speak of underidentification only when $rank[E(\Psi_t)] = 1$. In this case, there will be a linear subspace of dimension two of observationally equivalent solutions for φ that will satisfy the moment conditions (B4), which again coincides with the nullspace of $E(\Psi_t)$. Graphically, this linear subspace is a plane in $\mathbb{R}^3$ that goes through the origin.

In this case, even if I normalise φ to be on the surface of the unit sphere, I will obtain a "great circle" of observationally equivalent parameter values. As is well known, the intersection of a plane with a sphere produces a circle. When the plane goes through the centre of the sphere, the resulting circle is termed a "great circle" or "orthodrome".[15]

Finally, in the extreme case in which $rank[E(\Psi_t)] = 0$, $\mathbb{R}^3$ itself constitutes the identified set, and so does the entire unit sphere after imposing the symmetric normalisation restriction $\varphi'\varphi = 1$. I shall not discuss this uninteresting case any further.

B.2 Non-linear in parameter model

Next, I abandon the linear world and replace the original moment conditions by

$$E(\Psi_t)\varphi(\alpha) = 0, \tag{B5}$$

where $\varphi(\alpha)$ is a continuously differentiable function of a single parameter α that maps the closure of an open set defined over the real line onto the three dimensional Euclidean space. I make two further assumptions:

1. $\varphi(\alpha)$ is injective so that $\partial\varphi(\alpha)/\partial\alpha$ is not identically 0.

2. $\varphi(\alpha)$ has been normalised in such a way that Assumption 2 holds.

The first assumption rules out trivial cases in which the underidentification of α has nothing to do with $E(\Psi_t)$. In turn, the second assumption rules out situations in which underidentification is related to a mere scaling issue. For simplicity, henceforth I assume the symmetric normalisation $\varphi(\alpha)'\varphi(\alpha) = 1$ for all α, although my subsequent analysis will carry through *mutatis mutandi* for alternative normalisations.

Obviously, when $rank[E(\Psi_t)] = 3$, the only solution for α would require $\varphi(\alpha) = 0$, which should not be feasible after normalising $\varphi(\alpha)$.

In turn, when $rank[E(\Psi_t)] = 2$, only two situations may happen: either there is a single value of α, say α_0, for which $\varphi(\alpha)$ belongs to the nullspace of $E(\Psi_t)$, in which case this point is locally and globally identified, or there is not, in which case model

[15] Assuming planet Earth were a perfect sphere, the equator and all combinations of a meridian with its opposite meridian would be great circles. In fact, any geodesic lies on a great circle because on can always fix the location of the North Pole so that a geodesic becomes part of a meridian. In contrast, all parallels except the equator would be "small circles".

(B5) is misspecified even though model (B4) is correctly specified. Importantly, the uniqueness of α_0 under correct specification follows from assumptions 1 and 2. In addition, this single point will also be generally first-order identified because the Jacobian of (B5) is $E(\Psi_t)d\varphi(\alpha)/d\alpha$, which cannot be zero when $rank[E(\Psi_t)] = 2$ if (B5) holds, unless $d\varphi(\alpha)/d\alpha$ evaluated at α_0 happened to be proportional to $\varphi(\alpha_0)$.

In contrast, the non-linearity of the $\varphi(\alpha)$ function introduces new meaningful underidentification possibilities when $rank[E(\Psi_t)] = 1$.

To illustrate those possibilities, I study first a simple example in which $\varphi(\alpha)$ generates a small circle on the unit sphere. Without loss of generality, I can then redefine the 3D coordinate system so that this small circle is actually a parallel. Thus, I obtain

$$\varphi(\alpha) = (\cos\bar{\beta}\sin\alpha, \cos\bar{\beta}\cos\alpha, \sin\bar{\beta}), \tag{B6}$$

where $\bar{\beta}$ provides the common latitude of all the points on the small circle and $\alpha \in (0, 2\pi]$ their longitude. Please note that $\varphi(\alpha)$ is not only injective but it also satisfies assumption 2 (lack of proportionality) because $\varphi'(\alpha)\varphi(\alpha) = 1$ for all α. Without loss of generality, I assume that $\beta \geq 0$, so that $\varphi(\alpha)$ lies on the Northern hemisphere.

In this context, the set of admissible solutions for α will be given by the intersection of $\varphi(\alpha)$ in (B6) and the great (semi) circle generated by the intersection of the plane of observationally equivalent value of φ in (B4) with the Northern hemisphere.

Several situations may happen:

1. $\bar{\beta} > 0$ but the great circle of observationally equivalent normalised solutions to (B4) is such that the maximum latitude that it can reach in this coordinate system is below $\bar{\beta}$. For example, the small circle is the Arctic circle but the great circle is the ecliptic, whose maximum latitude is 0.409048628 radians (23.43676°), as in Figure B1a. In this case, there is no value of α that can satisfy the non-linear moment conditions (B5), so they will be rejected.

2. $\bar{\beta} = 0$ and the great circle of observationally equivalent normalised solutions to (B4) coincides with the equator, as in Figure B1b. In this case, there is a continuum of values of α that satisfy the moment conditions, so that α is not locally identified.

3. $\bar{\beta} \geq 0$ and the great circle of observationally equivalent normalised solutions to (B4) intersects (B6) twice, as in Figure B1c. In this case, α will be first-order and

therefore locally identified at each of those two solutions, but not globally identified. In the special case of $\bar{\beta} = 0$, those solutions are antipodean.

4. $\bar{\beta} > 0$ and the great circle of observationally equivalent normalised solutions to (B4) intersects (B6) once, but in such a way that the arc distance from the intersection point to the equator along the great circle of observationally equivalent linear structures is exactly $\pi/2$ (see Figure B1d). In this case, α is locally identified, but first-order underidentified. The intuition is as follows. Geometrically, $\varphi'(\alpha) = (\cos\bar{\beta}\cos\alpha, -\cos\bar{\beta}\sin\alpha, \sin\bar{\beta})$ describes another point on the equator which is orthogonal to the original point on the parallel $\varphi(\alpha)$. Therefore, if the plane $E(\Psi_t)\varphi = 0$ goes through those two points, then

$$E(\Psi_t)d\varphi(\alpha)/d\alpha = 0 \tag{B7}$$

and (B5) will simultaneously hold.

Interestingly, I can show that this solution can be obtained as the limit of the previous one as the solutions get closer and closer to each other in such a way that (B7) holds.

5. $\bar{\beta} > 0$ and the great circle of observationally equivalent normalised solutions to (B4) intersects (B6) once, but the arc distance along the great circle from the intersection point to the equator is different from $\pi/2$ (see Figure B1e). In this ideal case, α is first-order, locally and globally identified.

Unfortunately, when $\varphi(\alpha)$ is given by (B6), it is impossible to generate an infinite but countable number of observationally equivalent values of α. Nevertheless, I can do so with the following alternative example. Suppose that $\varphi(\alpha)$ generates what is know as a "rhumb line" (or loxodrome). This is the trajectory along the sphere that a ship starting from the North pole will follow if it moved by maintaining a constant angle $\beta \neq \pi/2$ against the meridians.[16]

Formally, the Cartesian coordinates of a rhumb line are

$$\varphi(\alpha) = \left(\frac{\cos\alpha}{\sqrt{\alpha^2\gamma^2+1}}, \frac{\sin\alpha}{\sqrt{\alpha^2\gamma^2+1}}, -\frac{\alpha\gamma}{\sqrt{\alpha^2\gamma^2+1}}\right),$$

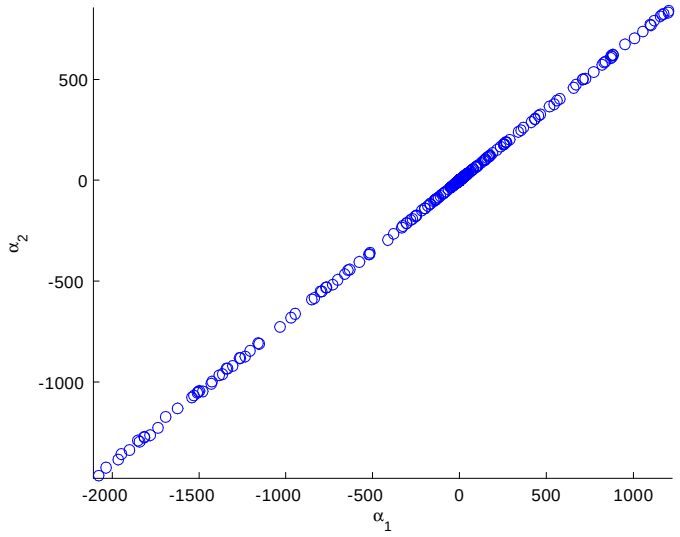
[16] Loxodromes became popular in sea navigation not only because they simply require that the compass is kept pointing in a constant direction or bearing, but also because they reduce to straight lines in Mercator's 2D projection.

where γ is a constant related to the bearing and $\alpha \in (-\pi, \pi]$.

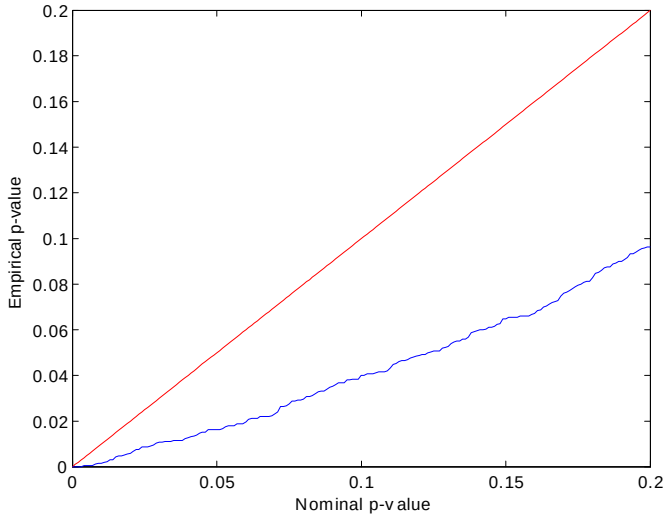
The number of points at which rhumb line will cross the great circle of observationally equivalent normalised solutions to (B4) depends on the inclination of this great circle with respect to the equator. The larger the inclination, the more crossings there will be, and consequently, the larger the number of observationally equivalent values of α compatible with (B5). In fact, a well known property of rhumb lines is that they are spherical spirals, which implies that if the great circle coincided with a meridian and $\beta \neq 0$, as in Figure B1f, then there would be a countable but infinite number of crossings very close to the North and South poles despite the distance from those two poles along the rhumb line being finite.

Figure 1: The effects of underidentification on Arellano and Bond (1991)

a: Scatter plot of CUEs



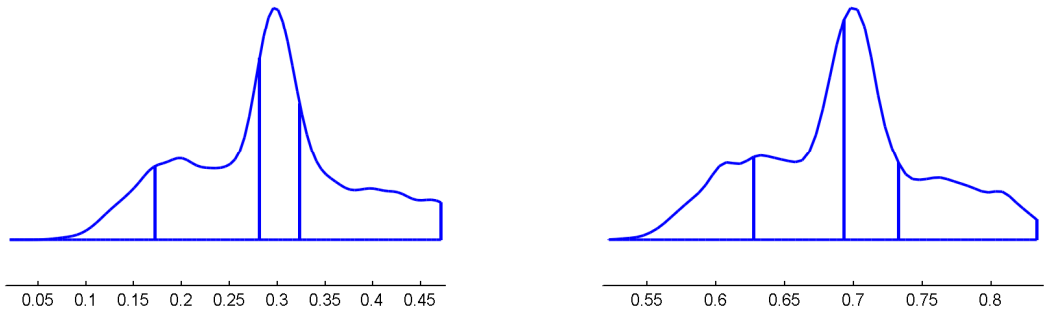
b: p-value plot of the J test



Notes: CUEs of α_1 and α_2 and associated J test based on the moment conditions (7) for $N = 5,000$ and $T = 5$ under time-series homoskedasticity (see section 3.2.1 for details).

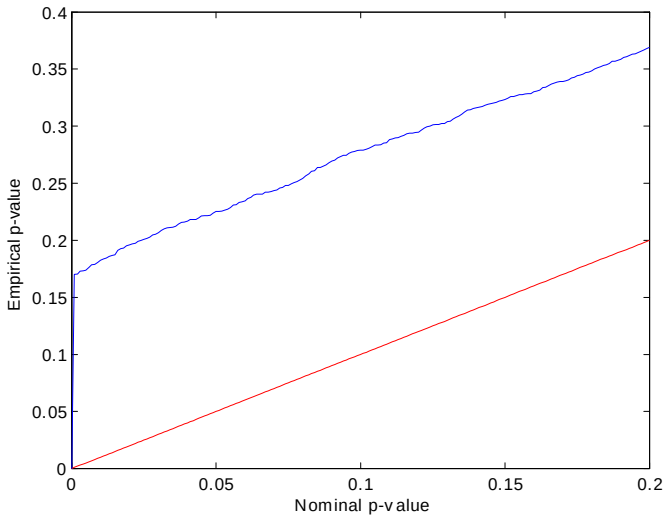
Figure 2: The effects of first-order underidentification on Ahn and Schmidt (1995)

a: Sampling distributions of CUEs



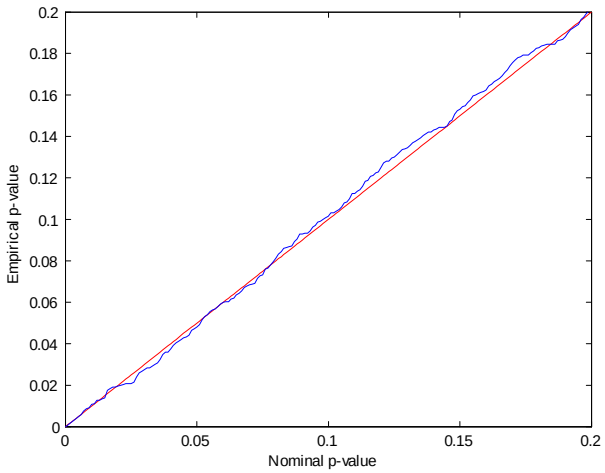
α_1 α_2

b: p-value plot for J test



Notes: CUEs of α_1 and α_2 and associated J test based on the moment conditions (7) and (9) for $N = 5,000$ and $T = 5$ under time-series homoskedasticity (see section 3.2.1 for details).

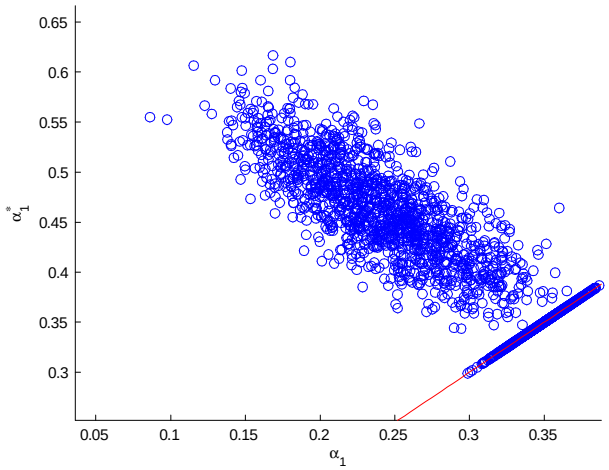
Figure 3: p-value plot for the first-order I test



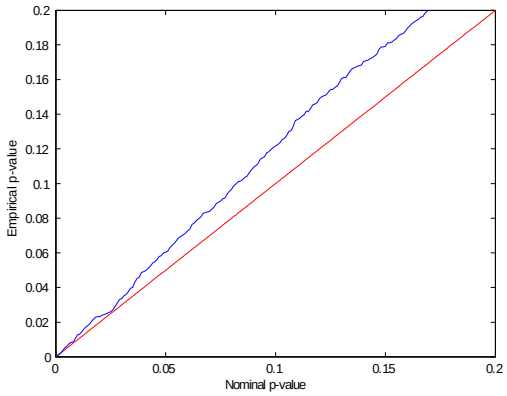
Notes: J test associated to the moment conditions (8), (10) and (12) for $N = 5,000$ and $T = 5$ under time-series homoskedasticity (see section 3.2.1 for details).

Figure 4: Pile-up problem with finite underidentification

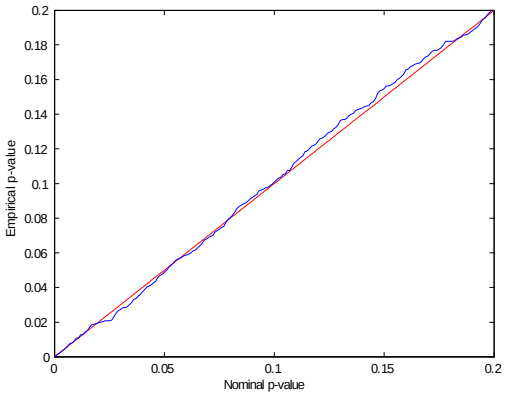
a: Scatter plot of the CUEs of α_1 and α_1^*



b: p-value plot of the non-linear finite I test



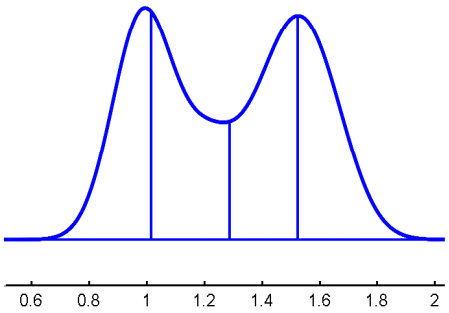
c: p-value plot of the linear finite I test



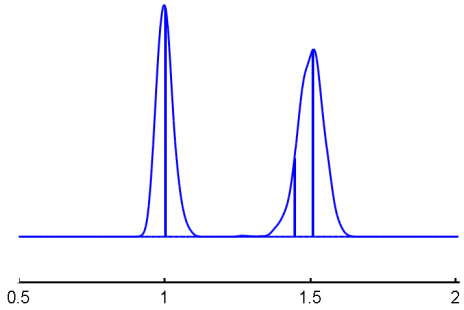
Notes: CUEs of α and α^* based on the moment conditions (8) and (13) for $N = 5,000$ and $T = 5$ under time-series heteroskedasticity (see section 3.2.1 for details).

Figure 5: The effects of finite underidentification on Hansen (1982)

Sampling distributions of CUEs

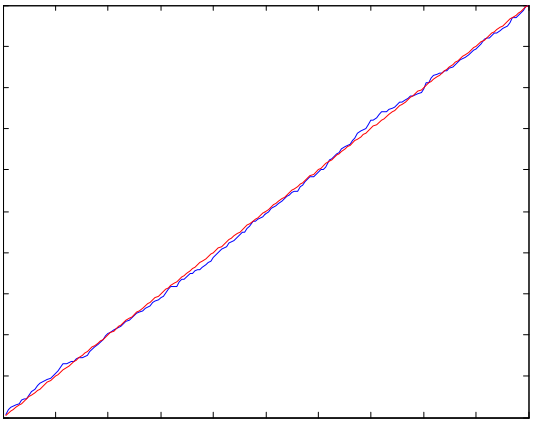


a: $T = 10,000$



b: $T = 100,000$

c: p-value plot for J test

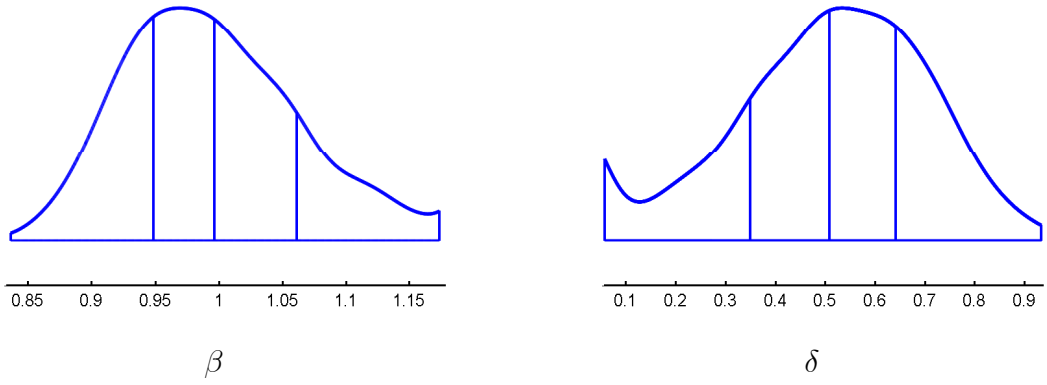


$T = 10,000$

Notes: CUE of β and associated J test based on the moment conditions (17) under finite underidentification (see section 4.2.1 for details).

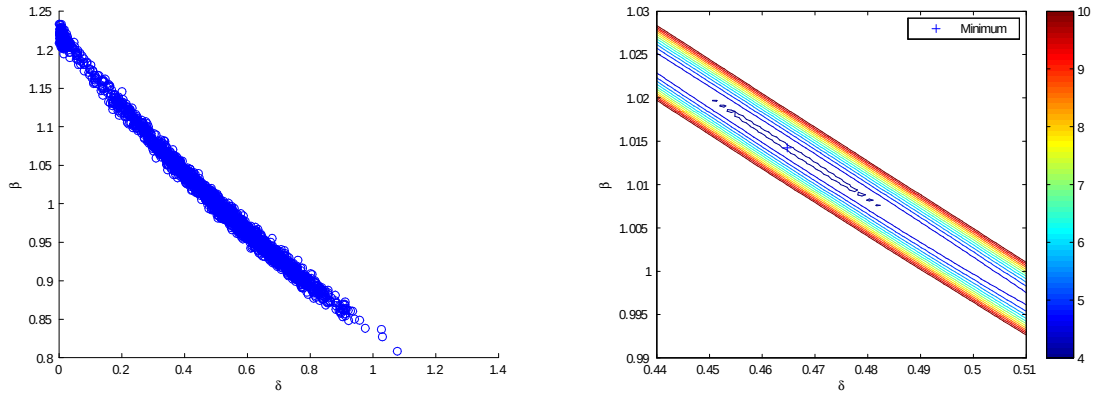
Figure 6: Finite set estimators

a: Sampling distributions of CUEs



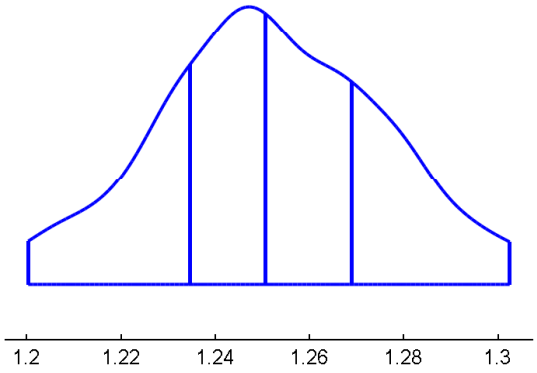
b: scatter plot

c: objective function



Notes: CUEs of β and δ based on the moment conditions (17) and (20) for $T = 10,000$ under finite underidentification (see section 4.2.1 for details).

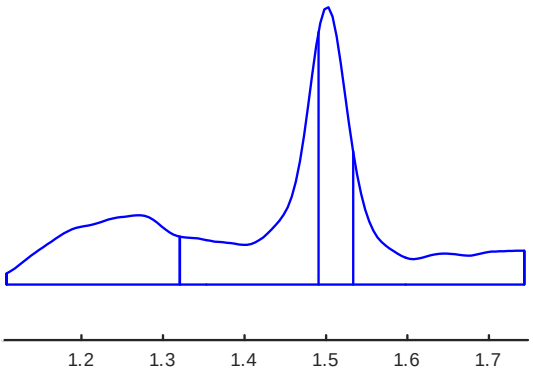
Figure 7: CUE based on expected Jacobian moments



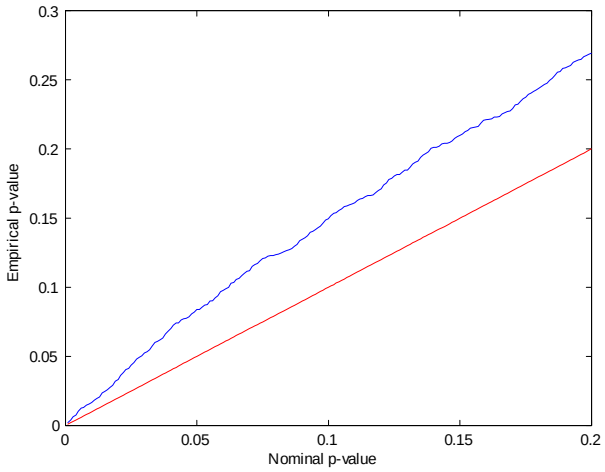
Notes: CUE of β based on the moment conditions (19) only for $T = 10,000$ under finite underidentification (see section 4.2.1 for details).

Figure 8: Effects of first-order identification on Hansen (1982)

a: Sampling distribution

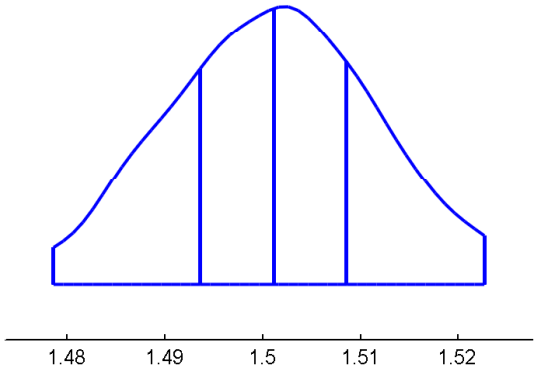


b: p-value plot of J test



Notes: CUE of β and associated J test based on the original moment conditions (17) for $T = 10,000$ under first-order underidentification (see section 4.2.1 for details).

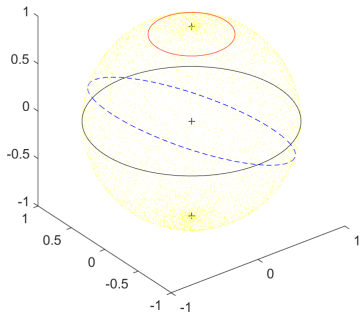
Figure 9: Restoring standard distributions under first-order identification



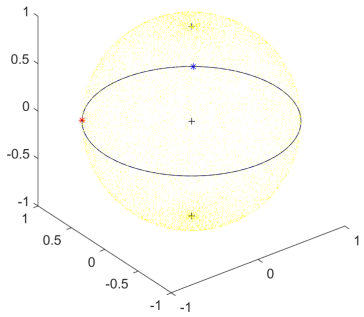
Notes: CUE of β based on the moment conditions (17) and (19) for $T = 10,000$ under first-order underidentification (see section 4.2.1 for details).

Figure B1: Underidentification in 3D

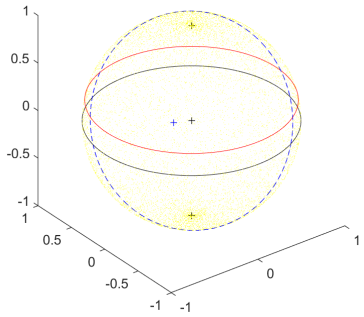
(a) Misspecification



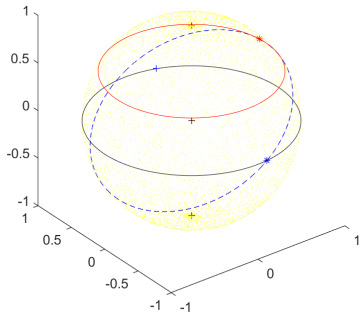
(b) Uncountable underidentification



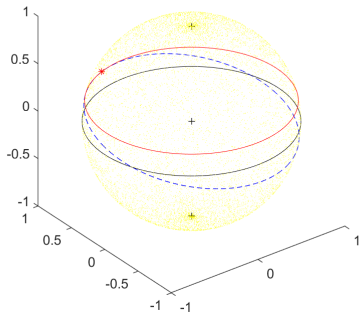
(c) Finite underidentification



(d) First-order underidentification



(e) First-order identification



(f) Countably infinite underidentification