Journal of Econometrics 170 (2012) 256–280

Contents lists available at SciVerse ScienceDirect

Journal of Econometrics

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

Underidentification?☆

Manuel Arellano [a,*], Lars Peter Hansen [b], Enrique Sentana [a]

[a] CEMFI, Casado del Alisal 5, E-28014 Madrid, Spain

[b] Department of Economics, University of Chicago, 1126 East 59th St. Chicago, IL 60637, USA

ARTICLE INFO

Article history:
Available online 2 June 2012

ABSTRACT

We develop methods for testing that an econometric model is underidentified and for estimating the nature of the failed identification. We adopt a generalized-method-of moments perspective in a possibly non-linear econometric specification. If, after attempting to replicate the structural relation, we find substantial evidence against the overidentifying restrictions of an augmented model, this is evidence against underidentification of the original model. To diagnose how identification might fail, we study the estimation of a one-dimensional curve that gives the parameter configurations that provide the greatest challenge to identification, and we illustrate this calculation in an empirical example.

1. Introduction

> "It is …natural to abandon without further computation the set of restrictions strongly rejected by the (likelihood ratio) test. Similarly, it is natural to apply a test of identifiability before proceeding with the computation of the sampling variance of estimates …and to forego any use of the estimates, if the indication of nonidentifiability is strong". Koopmans and Hood (1953) (see p. 184).

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

We find it productive to pose this as an estimation problem where we seek to identify the location of the identification problem. We adopt a generalized-method-of-moments (GMM) perspective and suppose that (under the null hypothesis) we may identify a curve or a function of a scalar variable that represents such a curve. Thus the target of estimation is this curve or the corresponding function. For models that are sufficiently linear, this curve represents a one-dimensional subspace, but translated. The resulting set can be easily parameterized and estimated using a standard GMM approach. In effect we build an augmented structural model in which the moment conditions are satisfied by a curve instead of a point. This estimation is of direct interest because it isolates the dimension along which identification of the original model is problematic. The familiar J test from the work of Sargan (1958) or Hansen (1982a) for overidentification of the augmented model now becomes a test for "underidentification" of the original model. If we can identify a curve or function representing that curve without statistical rejection, then the original model is not well identified and we refer to this phenomenon as underidentification. We refer to such a test as an I test. In contrast, a statistical rejection provides evidence that the parameter vector in the original model is indeed point identified, unless of course the familiar J test continues to reject its over-identifying restrictions.

The idea of identifying a curve extends to estimation environments in which we may not be able to represent the curve with a finite-dimensional parameterization. Thus we also suggest a more general estimation approach, study the resulting statistical efficiency and discuss implementation. Inferential methods are necessarily altered, and we are led to build on the work of Carrasco and Florens (2000) in designing a GMM approach to this problem.

We consider in progression three different estimation environments: models that are linear in parameters (Section 3), models

☆ This is a substantially revised and extended version of Arellano et al. (1999). We thank Javier Alvarez, Raquel Carrasco, Rui Cui, Jesús Carro, Alessandro Galesi, and especially Elena Manresa, Javier Mencía and Francisco Peñaranda for able research assistance at various stages of this project. Also, we thank Andres Santos and Azeem Shaikh for helpful conversations about our general approach to this problem. Finally, we thank Marine Carrasco and three referees for valuable feedback on earlier drafts of this paper. Moreover, we benefited from valuable comments from multiple seminar audiences. MA and ES gratefully acknowledge financial support from the Spanish Ministry of Science and Innovation through grant ECO 2008-00280 and LPH gratefully acknowledges support from the National Science Foundation through grant SES0519372.

* Corresponding author.
E-mail addresses: arellano@cemfi.es (M. Arellano), lhansen@uchicago.edu (L.P. Hansen), sentana@cemfi.es (E. Sentana).

0304-4076/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2012.05.005

with nonlinear restrictions on the parameters (Section 4), and finally, models with more fundamental nonlinearities (Section 5). Throughout we develop specific examples to illustrate the nature and the applicability of our methods. In Section 6 we show how to apply these methods to a consumption-based asset pricing model fit to microeconomic data. What follows is a more detailed overview of the paper.

2. Overview

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

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

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

to identify a parameter vector β_0. When β_0 is identified, it is the unique solution to (1), otherwise there will be multiple solutions. To relate to standard analyses of identification and develop tests for underidentification, we suppose that $p\geq k$. In discussing the lack of identification in non-linear models in those circumstances, it is important to distinguish the different situations that may arise. We say that $\beta^*\neq\beta_0$ is observationally equivalent to β_0 if and only if $E[f(x_t;\beta^*)]=0$. The true value β_0 is locally identifiable if there is a neighborhood of β_0 such that in this neighborhood $E[f(x_t;\beta)]=0$ only if $\beta=\beta_0$ (Fisher, 1966). 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 $E\left[\partial f(x,\beta)/\partial\beta'\right]$ is continuous at β_0, and $\operatorname{rank}\{E\left[\partial f(x,\beta_0)/\partial\beta'\right]\}=k$, then β_0 is locally identified (Fisher, 1966; Rothenberg, 1971). In contrast to the order condition, this condition is only sufficient. But if $\operatorname{rank}\{E\left[\partial f(x,\beta)/\partial\beta'\right]\}$ is also constant in a neighborhood of β_0, then the above rank condition becomes necessary too. However, as argued in Sargan (1983a,b), there are non-linear models in which the rank condition fails, and yet β_0 is locally identified (see Wright, 2003, for tests of the Jacobian rank condition in non-linear models).

In this paper we will take a decidedly global approach. Global identification requires that β_0 be the unique solution on $\mathbb{P}$ to the system of Eqs. (1). To study underidentification, we follow Sargan (1959) by imposing an explicit structure on the lack of identification. This leads us to study an alternative estimation problem. Specifically, we consider a parameterization of the form $\beta=\pi(\theta)$, where π is a continuous function with range $\mathbb{P}$ and $\theta\in\Theta\subset\mathbb{R}$, which is some conveniently chosen domain. For example, suppose that

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

so that θ is the first component of the parameter vector. We then explore a set of such functions that is restricted appropriately.[2] As an alternative identification condition, we require $\bar f[\pi(\theta)]=0$ for all $\theta\in\Theta$ if, and only if $\pi=\pi_0$. If we can successfully identify a nonconstant function π_0 that realizes alternative values in the parameter space, then we cannot uniquely identify a single parameter vector β_0 from the moment conditions (1). Thus the parameter vector β_0 is underidentified.

Our investigation of underidentification leads naturally to the question of how to estimate π_0 efficiently. One approach would be to use one of the standard GMM objective functions and try to construct an estimator of π_0 as the set of approximate minimizers of that objective. In this paper we explore a rather different approach, one that depicts the identification failure in the construction of π_0. It leads us naturally to ask what the efficiency gains are to estimating jointly alternative points along a curve, say $\pi_0(\theta)$ for alternative values of θ. Initially we illustrate these efficiency gains using a standard formulation of the GMM efficiency bounds.

In what follows we use parameterization (2). For each value of θ in a finite set $\Theta_n=\{\theta_1,\theta_2,\ldots,\theta_n\}$ we estimate the $(k-1)$-dimensional parameter vector $\tau(\theta)$. We can map this into a standard GMM problem where we simply replicate the original moment conditions. Later we will extend this discussion to the case in which the set of θ's that are of interest is an interval, but this finite set construction will set the stage for a more general treatment. As posed, this is a standard GMM estimation problem, albeit one with a special structure. To analyze the gains to joint efficiency, we presume the following central limit approximation:

Assumption 2.1. $\frac{1}{\sqrt T}\sum_{t=1}^T f[x_t,\pi_0(\theta)]$ converges to a Gaussian random vector $\{g(\theta):\theta\in\Theta_n\}$ with mean zero for all $\theta\in\Theta_n$ and covariance function:

$$K(\theta,\vartheta)=E\left[g(\theta)g(\vartheta)'\right]=\lim_{T\to\infty}\frac{1}{T}E\left(\sum_{t=1}^T f[x_t,\pi_0(\theta)]\sum_{s=1}^T f[x_s,\pi_0(\vartheta)]'\right).$$

In stating this assumption and in what follows we will abuse the E notation by using it both for the original probability space and for the probability space used in constructing the Gaussian process used in the central limit approximation.

Construct

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

where 0_{k-1} is a row vector of zeros and I_{k-1} is an identity matrix of dimension $k-1$.

Assumption 2.2. $D(\theta)'D(\theta)$ is nonsingular for each $\theta\in\Theta_n$.

With these ingredients, we apply directly the analysis in Hansen (1982a) and the earlier analysis in Sargan (1958, 1959), which involves reducing the moment conditions by introducing a $(k-1)\times n$ by $p\times n$ selection matrix A that picks among the possible moment conditions:

$$Ef[x_t,\pi(\theta_j)]=0$$

for $j=1,2,\ldots,n$. The moment restrictions are thus broken into n blocks and the parameter vector $\pi(\theta_j)$ only appears in block j. Our characterization of the GMM efficiency bound exploits this block structure. A partitioned selection matrix reduces the moments to be the same as the number of free parameters and has the estimation problem focus on:

$$\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1n}\\ A_{21}&A_{22}&\cdots&A_{2n}\\ -&-&-&-\\ A_{n1}&A_{n2}&\cdots&A_{nn}\end{bmatrix}\begin{bmatrix}Ef[x_t,\pi(\theta_1)]\\ Ef[x_t,\pi(\theta_2)]\\ -\\ Ef[x_t,\pi(\theta_n)]\end{bmatrix}=\begin{bmatrix}0\\0\\-\\0\end{bmatrix},$$

where A_{ij} has dimension $(k-1)\times p$. The choice of selection matrix $A=[A_{ij}]$ alters the resulting statistical efficiency of a

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

[2] See Section 5 for further details.

GMM estimation. Some collections of selection matrices imply the same asymptotic efficiency, however. For instance, forming a new selection matrix by premultiplying a given selection matrix by a nonsingular matrix in effect uses the same moment conditions for estimation and hence does not alter the asymptotic efficiency of the corresponding GMM estimator. As a consequence, if we define D as a block diagonal matrix of dimension $p \times n$ by $(k-1) \times n$ with $D(\theta_j)$ in the jth diagonal position, selection matrices for our estimation problem can always be restricted to satisfy (at least asymptotically):

$$AD = I \tag{3}$$

without altering the efficiency bound. By imposing this restriction we simplify the formula for the asymptotic covariance matrix as we will see. Consider now the first row block of (3):

$$A_{11}D(\theta_1) = I,$$
$$A_{1j}D(\theta_j) = 0 \quad j = 2, 3, \ldots, n. \tag{4}$$

It follows from Hansen (1982a) that the asymptotic covariance matrix for the resulting GMM estimator of $\tau(\theta_1)$ is the covariance matrix of

$$\sum_{j=1}^{n} A_{1j} g(\theta_j).$$

The corresponding efficiency bound is solved by "minimizing" this covariance matrix by choice of $A_{11}, A_{12}, \ldots, A_{1n}$. While covariance matrices are only partially ordered, this minimization problem turns out to have a well defined minimum. The zero restrictions imposed in (4) control for the fact that $\tau(\theta_j)$ for $j = 2, \ldots, n$ are estimated at the same time as $\tau(\theta_1)$ and hence limit the construction of the selection matrix. While we have focused on the efficiency of $\tau(\theta_1)$, an analogous argument applies for $\tau(\theta_j)$ for $j = 2, \ldots, n$.

It will be convenient for us to represent this minimization problem differently. We consider random variables of the form

$$\sum_{j=1}^{n} B_j' g(\theta_j),$$

where

$$D(\theta_j)' B_j = 0$$

for $j = 1, 2, \ldots, n$. The zero restriction limits the basis random variables $B_j' g(\theta_j)$ that we will use in our construction of the bound. Form

$$\mathcal{F}_n = \left\{ y = \sum_{j=1}^{n} B_j' g(\theta_j) : D(\theta_j)' B_j = 0,\ j = 1, 2, \ldots, n \right\}.$$

Next transform the random vectors $\{g(\theta) : \theta \in \Theta_n\}$ into:

$$h(\theta) = [D(\theta)' D(\theta)]^{-1} D(\theta)' g(\theta).$$

Then the efficiency bound for a given n is obtained by solving:

$$\min_{f \in \mathcal{F}_n} E\left(\left[\gamma' h(\theta_1) - f \right]^2 \right)$$

for any vector γ. The solution is

$$\text{Proj}\left[\gamma' h(\theta_1) | \mathcal{F}_n \right],$$

where Proj is the least squares projection operator and the minimized objective is the second moment of the least squares projection error.[3]

While the finite parameter GMM bound is well known, what follows gives a formal representation of that bound that will have direct extension as we expand the number of moment conditions used in estimation.

[3] In what follows when we use the notation Proj applied to a random vector we mean the vector of projections obtained by projecting each coordinate on the relevant closed linear space of random variables that is being projected on.

Proposition 2.1. *The GMM efficiency bound for estimating $\tau(\theta_j)$ is:*

$$E\left[\left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}_n] \right) \left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}_n] \right)' \right]$$

for $j = 1, 2, \ldots, n$.

Proof. Write

$$h(\theta_1) - \text{Proj}\left[h(\theta_1) | F_n \right] = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \cdots & \tilde{A}_{1n} \end{bmatrix} \begin{bmatrix} g(\theta_1) \\ g(\theta_2) \\ - \\ g(\theta_n) \end{bmatrix},$$

where

$$\tilde{A}_{1j} D(\theta_j) = 0$$

for $j = 1, 2, \ldots, n$. Form

$$A_1^* = \begin{bmatrix} -\tilde{A}_{11} + [D(\theta_1)' D(\theta_1)]^{-1} D(\theta)' & -\tilde{A}_{12} & \cdots & -\tilde{A}_{1n} \end{bmatrix}.$$

Notice that

$$A_1^* D = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}.$$

Let A_1^* be the first row block of a selection matrix A^* that satisfies (3). Consider some other matrix A_1 such that

$$A_1 D = \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix},$$

which is the first row block of a selection matrix A that satisfies (3). Then the entries of the random vector

$$[A_1 - A_1^*] \begin{bmatrix} g(\theta_1) \\ g(\theta_2) \\ - \\ g(\theta_n) \end{bmatrix}$$

are in $\mathcal{F}_n$. Hence

$$(A_1 - A_1^*) E \left(\begin{bmatrix} g(\theta_1) \\ g(\theta_2) \\ - \\ g(\theta_n) \end{bmatrix} \begin{bmatrix} g(\theta_1)' & g(\theta_2)' & \cdots & g(\theta_n)' \end{bmatrix} \right) (A_1^*)' = 0$$

since a vector of projection errors is orthogonal to the space that is being projected onto. Thus

$$A_1 E \left(\begin{bmatrix} g(\theta_1) \\ g(\theta_2) \\ - \\ g(\theta_n) \end{bmatrix} \begin{bmatrix} g(\theta_1)' & g(\theta_2)' & \cdots & g(\theta_n)' \end{bmatrix} \right) (A_1)'$$
$$\geq A_1^* E \left(\begin{bmatrix} g(\theta_1) \\ g(\theta_2) \\ - \\ g(\theta_n) \end{bmatrix} \begin{bmatrix} g(\theta_1)' & g(\theta_2)' & \cdots & g(\theta_n)' \end{bmatrix} \right) (A_1^*)'$$

where $\geq$ is the usual inequality for comparing positive semidefinite matrices. □

Remark 2.2. In this representation the covariance of $h(\theta_j)$ is the asymptotic covariance matrix for a GMM estimator that uses only moment conditions based on

$$[D(\theta_j)' D(\theta_j)]^{-1} D(\theta_j)' E f[x_t, \pi(\theta_j)] = 0,$$

but ignores the possible efficiency gains from joint estimation. It even fails to solve the second-best problem of efficiently estimating $\tau(\theta_j)$ using linear combinations of the moment restrictions:

$$E f[x_t, \pi(\theta_j)] = 0$$

except in very special circumstances.

Remark 2.3. Our choice of using the selection matrix $[D(\theta_j)' D(\theta_j)]^{-1}D(\theta_j)'$ in the construction of $h(\theta)$ is only one possibility. Notice that if we had chosen an alternative selection matrix $\tilde{D}_j$ such that $\tilde{D}_j D(\theta_j) = I$, then

$$\begin{aligned}\text{Proj}[\tilde{D}_j g(\theta_j)|\mathcal{F}_n] &= \text{Proj}\left([D(\theta_j)'D(\theta_j)]^{-1}D(\theta_j)'g(\theta_j)|\mathcal{F}_n\right)\\ &\quad\times\left[\tilde{D}_j - [D(\theta_j)'D(\theta_j)]^{-1}D(\theta_j)'\right]g(\theta_j).\end{aligned}$$

Thus

$$\tilde{D}_j g(\theta_j) - \text{Proj}[\tilde{D}_j g(\theta_j)|\mathcal{F}_n] = h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}_n]$$

resulting in the same least squares projection error. As expected our choice of starting point is inconsequential to our final calculation.

Focusing on a finite set Θ_n, while pedagogically revealing, is too restrictive for some of our analysis. Given our regression-error characterization of the efficiency bound, there is a direct extension to estimating curves (Section 5). Nevertheless, for some important examples that we consider in the next two sections, in which we consider models that are linear in the parameters (Section 3) and models in which the nonlinearity is concentrated in the parameters (Section 4), it suffices to focus on a finite number of θ's.

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

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

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

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

has reduced rank for any θ^* in the interior of Θ. In contrast the weak instruments literature considers the reduced rank as the limit of a sequence of data generating models indexed by the sample size.[4] In our analysis such a sequence could be interesting as a local specification under the alternative hypothesis of identification. We seek to infer the specific manner in which identification may fail whereas the weak instrument literature focuses on developing reliable standard errors and tests of hypotheses about a unique true value of β.

The other strand is the set estimation literature (see e.g. Chernozhukov et al., 2007 or more recently Yildiz, 2012), which often allows for $E[f(x;\beta)] = 0$ for set values of β and whose objective is to make inferences about this set.[5] In contrast, in this paper we explore the precise nature of the underidentification. Given this focus, we are led to add structure to the potential underidentification that is considered. By adding this structure to the possible identification failure, we are led to alter the usual GMM objective in order to estimate efficiently the one-dimensional function π that parameterizes the potential lack of identification.

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

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

3. Linearity in the parameters

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

$$E(\Psi_t)\begin{bmatrix}1\\-\beta\end{bmatrix} = 0, \tag{5}$$

where β is a k-dimensional unknown parameter vector and Ψ_t is an p by $k+1$ matrix of random variables constructed from data.[6] Suppose that there are two solutions to equation (5), say $\beta^{[1]}$ and $\beta^{[2]}$ where the first entry of $\beta^{[1]}$ is restricted to be one and the first entry of $\beta^{[2]}$ to zero.[7] Then

$$E(\Psi_t)\begin{bmatrix}1\\-\pi_0(\theta)\end{bmatrix} = 0,$$

for all $\theta \in \mathbb{R}$, where

$$\pi_0(\theta) = \theta\pi_0(0) + (1-\theta)\pi_0(1). \tag{6}$$

A feature of the linearity in parameters is that we can identify two distinct values of β that satisfy (5), in this case $\pi_0(0) = \beta_0^{[1]}$, $\pi_0(1) = \beta_0^{[2]}$. Thus to study underidentification, we focus on identifying $\beta_0^{[1]}$ and $\beta_0^{[2]}$ that solve the duplicated moment conditions

$$\left.\begin{aligned}E(\Psi_t)\begin{bmatrix}1\\-\beta^{[1]}\end{bmatrix} = 0\\ E(\Psi_t)\begin{bmatrix}1\\-\beta^{[2]}\end{bmatrix} = 0\end{aligned}\right\}. \tag{7}$$

In this problem we envision $\beta_0^{[1]}$ and $\beta_0^{[2]}$ as the target of a GMM estimation problem subject to the restrictions on the first entries that we mentioned previously. This leads us to estimating $2\times(k-1)$ free parameters. Given estimates of $\beta_0^{[1]}$ and $\beta_0^{[2]}$, we then infer a one-dimensional curve (actually a line) using formula (6).

In posing the above estimation problem, we imposed a "normalization" in the original equation (5). In what follows we will adopt a different and slightly more general starting point by considering:

$$E(\Psi_t)\alpha = 0, \tag{8}$$

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

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

which we effectively imposed in (5). Alternatively, we could choose $q = \alpha$ so that $|\alpha| = 1$, together with a sign restriction on one of the nonzero coefficients as in:

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

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

[7] We adopt these restrictions for convenience. Normalizing a coefficient to unity is common practice, and normalizing the second one to have zero coefficient rules out the possibility that the resulting coefficient vectors are proportional. Other normalizations are possible.

where $|\beta| \leq 1$. Neither of these approaches can be employed without loss of generality, however. The particular application dictates how to select the parameters of interest from this direction.[8]

Suppose now that instead of a one-dimensional subspace, we can actually infer a two-dimensional subspace of α's that satisfies (8). This leads us to efficiently estimate those $\alpha_0^{[1]}$ and $\alpha_0^{[2]}$ for which

$$\left.\begin{array}{l} E(\Psi_t)\alpha^{[1]} = 0 \\ E(\Psi_t)\alpha^{[2]} = 0 \end{array}\right\}. \tag{9}$$

Our parameterization π_0 given in (6) gives us one way to parameterize this two-dimensional subspace. It is the space spanned by the two vectors: $\begin{bmatrix} 1 & -\pi_0(0) \end{bmatrix}'$ and by $\begin{bmatrix} 1 & -\pi_0(1) \end{bmatrix}'$ as required by (7).

The duplicated moment conditions in (7) or (9) give us a direct link to the rank condition familiar in the econometrics literature. Suppose the order condition ($p \geq k$) is satisfied, but not necessarily the rank condition. Thus the maximal possible rank of the matrix $E(\Psi_t)$ is $\min\{p, k+1\}$. Model (8) is said to be *identified* when $E(\Psi_t)$ has rank k, in which case its null space is precisely one-dimensional. When $p > k$ and the model is identified, it is said to be *overidentified* because the rank of the matrix $E(\Psi_t)$ now must not be full. Instead of having maximal rank $k+1$, $E(\Psi_t)$ has reduced rank k. This implication is known to be testable and statistical tests of overidentification are often conducted in practice.

In contrast, model (8) is said to be *underidentified* when the rank of $E(\Psi_t)$ is less than k. In this case the null space of $E(\Psi_t)$ will have more than one dimension. A single normalization will no longer select a unique element from the parameter space. By focusing on (6), our approach puts an explicit structure on the lack of identification, as illustrated by (9). Thus, we initially make the following assumption (see Section 3.1.2 for other possibilities):

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

Under this hypothesis the set of solutions to Eq. (8) is two-dimensional. To test for this lack of identification, we think of (9) as a new *augmented model*. We attempt to determine $(\alpha^{[1]}, \alpha^{[2]})$ simultaneously and ask whether they satisfy the combined overidentifying moment restrictions (9). If they do, then we may conclude that the original econometric relation is *not identified* or equivalently is underidentified. Thus by building an augmented equation system, we may pose the null hypothesis of underidentification as a hypothesis that the augmented equation system is overidentified. Rejections of the overidentifying restrictions for the augmented model provide evidence that the original model is indeed identified. Posed in this way, underidentification can be tested simply by applying appropriately an existing test for overidentification. For instance, a standard J test for overidentification, such as those of Sargan (1958) and Hansen (1982a), is potentially applicable to the augmented model. This test will be our I test.

The following example illustrates our formulation.

Example 3.1. Suppose that $p = 1$ and $k = 1$. Write

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

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

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

3.1. Single equation IV

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

$$y_t'\alpha = u_t,$$

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

$$E(z_t u_t) = 0. \tag{10}$$

Form:

$$\Psi_t = z_t y_t'.$$

Then orthogonality condition (10) is equivalent to α satisfying the moment relation (8).

For this example we duplicate the moment conditions as in (9), and study the simultaneous overidentification of those $2p$ moment conditions. To proceed with the construction of a test, we have to rule out the possibility that $\alpha^{[1]}$ and $\alpha^{[2]}$ are proportional. One strategy is to restrict $\alpha^{[2]}$ to be orthogonal to $\alpha^{[1]}$. Two orthogonal directions can be parameterized with $2k-1$ parameters, k parameters for one direction and $k-1$ for the orthogonal direction. However, there is not a unique choice of orthogonal directions to represent a two-dimensional space. There is an additional degree of flexibility. A new direction can be formed by taking linear combinations of the original two directions and a corresponding orthogonal second direction. Thus the number of required parameters is reduced to $2k-2$, and the number of overidentifying restrictions for the I test of underidentification is $2p-2k+2$.

In practice, we can impose the normalizing restrictions $|\alpha^{[1]}| = |\alpha^{[2]}| = 1$ by using spherical coordinates, forcing $\alpha^{[1]\prime}\alpha^{[2]} = 0$, and setting the first entry of $\alpha^{[2]}$ to zero. This works provided that all vectors in the null space of $E(z_t y_t')$ do not have zeros in the first entry. Alternatively, we could restrict the top two rows $(\alpha^{[1]}, \alpha^{[2]})$ to equal an identity matrix of order two. This rules out the possibility of a vector in the null space that is identically zero in its first two entries, but this may be of little concern for some applications.[9] When $k = 1$, both approaches boil down to setting $(\alpha^{[1]}, \alpha^{[2]}) = I_2$ so that the $2p$ moment conditions:

$$E\left(z_t y_t'\right) = 0$$

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

Example 3.1 could emerge as a special case of Example 3.2 with $p = 1$ and $k = 1$. Notice that our underidentification test in this case tests simultaneously the restriction that $a_1 = 0$ and $a_2 = 0$. More generally, when $p \geq 2$ our test considers simultaneously $E(z_t y_{1,t}) = 0$ and $E(z_t y_{2,t}) = 0$. The resulting I test is different from the test for the relevance of instruments

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

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

in a model with a normalization restriction on one variable to be estimated by, say, two-stage least squares. Such a test would examine only $E\left(z_t y_{2,t}\right) = 0$.

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

$$E\left[z_t\left(y_{0,t} - \beta_1 y_{1,t} - \beta_2 y_{2,t}\right)\right] = 0.$$

Now z_t need not be uncorrelated with all three endogenous variables for there to be underidentification. Lack of correlation with two linear combinations of them is enough.[10] For example, we may write the null of underidentification as

$$H_0 : E\begin{bmatrix} z_t\left(y_{0,t} - \gamma_1 y_{2,t}\right) \\ z_t\left(y_{1,t} - \gamma_2 y_{2,t}\right)\end{bmatrix} = 0.$$

If H_0 holds, for any β_1^*

$$E\left\{z_t\left[y_{0,t} - \beta_1^* y_{1,t} - \left(\gamma_1 - \gamma_2\beta_1^*\right) y_{2,t}\right]\right\} = 0,$$

so that the observationally equivalent values $\left(\beta_1^*, \beta_2^*\right)$ are contained in the line $\beta_2^* = \gamma_1 - \gamma_2\beta_1^*$.

A time series example is a forward-looking Phillips curve as in Galí et al. (2001), where the components of y denote current inflation, future inflation, and a measure of aggregate demand, whereas the components of z consist of lags of the previous variables, and of other variables such as the output gap and wage inflation. There are theoretical and empirical considerations to suggest that a null like H_0 is plausible in this context. For example, lack of higher-order dynamics in a new Keynesian macro model has been shown to be a source of underidentification of a hybrid Phillips curve with lagged inflation (see Mavroeidis, 2005 and Nason and Smith, 2008). Relatedly, Cochrane (2011) also raises similar concerns regarding the identification of Taylor rules by Clarida et al. (2000) and others.

3.1.1. Related literature

Tests of underidentification in a single structural equation were first considered by Koopmans and Hood (1953) and Sargan (1958). When the model is correctly specified and identified, the rank of $E(z_t y_t')$ is k. Under the additional assumptions that the error term u_t is a conditionally homoskedastic martingale difference, an asymptotic chi-square test statistic of overidentifying restrictions with $p - k$ degrees of freedom is given by $T\lambda_1$, where

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

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

Koopmans and Hood (1953) and Sargan (1958) indicated that when the rank of $E(z_t y_t')$ is $k - 1$ instead, if λ_2 is the second smallest characteristic root, $T(\lambda_1 + \lambda_2)$ has an asymptotic chi-square distribution with $2(p - k) + 2$ degrees of freedom. These authors suggested that this result could be used as a test of the hypothesis that the equation is underidentified and that any possible equation has an i.i.d. error term.

The statistic $T(\lambda_1 + \lambda_2)$ has a straightforward interpretation in terms of our approach. Indeed, it can be regarded as a continuously-updated GMM test of overidentifying restrictions of the augmented model (9), subject to the additional restrictions on the error terms mentioned previously. To see this, let $A = \begin{bmatrix}\alpha^{[1]} & \alpha^{[2]}\end{bmatrix}$ and consider the minimizer of

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

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

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

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

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

$$L = E(\Psi_t)'\left[E(z_t z_t')\right]^{-1} = E\left(y_t z_t'\right)\left[E(z_t z_t')\right]^{-1}.$$

This is the matrix of coefficients of the reduced form system of population regressions of the entries of y_t onto z_t. Suppose the second component of y_t is the first component of z_t. Partition L as:

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

The nullity of L and hence $E(\Psi_t)$ is the same as the nullity of Π_2. Cragg and Donald (1993) construct a minimum chi-square test statistic that enforces the rank restriction in Π_2.[11] Their statistic can also be related to our approach. As we show in Appendix A, under the assumption that u_t is a conditionally homoskedastic martingale difference, the Cragg–Donald statistic minimizes

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

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

While the Cragg and Donald (1993) approach is straightforward to implement in the single-equation case, it is more difficult to implement in some models with cross-equation restrictions. This difficulty can emerge because we must simultaneously impose the

[10] Phillips (1989) and Choi and Phillips (1992) study the IV estimator of β_1 and β_2 in the presence of identification failure.

[11] Cragg and Donald (1993) also considered an alternative null of no identifiability in an equation with the coefficient of one of the endogenous variables normalized to unity. This is a rank restriction in the submatrix of Π_2 that excludes the row corresponding to the normalized entry.

restrictions on the reduced form together with the rank deficiency. In Example 3.2, this is easy to do, and it is also feasible in the applications to linear observable factor pricing models of asset returns carried out by Cragg and Donald (1997) and Burnside (2007), but not in more general models as we will illustrate in Sections 3.2 and 3.3.

3.1.2. Underidentification of a higher dimension

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

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

If we impose $(j+1)^2$ normalizing restrictions on $(\alpha^{[1]}, \alpha^{[2]}, \ldots, \alpha^{[j+1]})$ to avoid indeterminacy,[12] the effective number of parameters is $(j+1)(k+1) - (j+1)^2 = (j+1)(k-j)$ and the number of moment conditions is $(j+1)p$ under the assumption that there are no redundancies. Therefore, by testing the $(j+1)(p-k+j)$ overidentifying restrictions in (13) we test the null that α is underidentified of dimension j against the alternative of underidentification of dimension less than j or identification. Henceforth, we shall refer to those tests as I_j tests.

3.2. Multiple equations with cross-equation linear restrictions

We next consider examples with multiple equations with common parameters.[13]

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

$$\alpha' \begin{bmatrix} y_{1,t} \\ y_{3,t} \end{bmatrix} = u_{1,t},$$

$$\alpha' \begin{bmatrix} y_{2,t} \\ y_{3,t} \end{bmatrix} = u_{2,t},$$

where $y_{1,t}, y_{2,t}$ are scalars. Let z_t denote a p^*-dimensional vector of common instrumental variables appropriate for both equations, so that

$$E\left(z_t u_{1,t}\right) = 0,$$

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

Form:

$$\Psi_t = \begin{bmatrix} z_t y_{1,t} & z_t y_{3,t} \\ z_t y_{2,t} & z_t y_{3,t} \end{bmatrix},$$

so that $p = 2p^*$. We transform this equation system to obtain an equivalent one by forming:

$$\Psi_t^* = \begin{bmatrix} z_t(y_{1,t} - y_{2,t}) & 0 \\ z_t y_{1,t} & z_t y_{3,t} \end{bmatrix} \tag{14}$$

implying that

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

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

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

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

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

$$y_{j,t} = \beta_{j,1} p_{1,t} + \beta_{j,2} p_{2,t} + \beta_{j,3} p_{3,t} + v_{j,t} \quad (j = 1, 2, 3),$$

where $y_{j,t}$ denotes the cost share of input j, and $p_{j,t}$ is the log price of input j relative to the omitted input.[14] The underlying cost function implies the following three cross-equation symmetry constraints

$$\beta_{j,k} = \beta_{k,j} \quad j \neq k.$$

Moreover, prices are endogenous (possibly due to data aggregation) and a p-dimensional vector of instruments z_t is available, so that:

$$E(z_t v_{j,t}) = 0 \quad (j = 1, 2, 3). \tag{16}$$

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

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

$$E[z_t(y_{j,t} - \gamma_{j,2} p_{2,t} - \gamma_{j,3} p_{3,t})] = 0, \quad (j = 1, 2, 3) \tag{17}$$

$$E[z_t(p_{1,t} - \gamma_{0,2} p_{2,t} - \gamma_{0,3} p_{3,t})] = 0. \tag{18}$$

Since there are $4p$ orthogonality conditions and 8 parameters, with $p = 2$ the augmented set of moments does not introduce any overidentifying restrictions. For arbitrary p, (17)–(18) imply that (16) is satisfied for any $\beta_{j,1}^*$, and for $\beta_{j,2}^*$, $\beta_{j,3}^*$ $(j = 1, 2, 3)$ such that

$$\beta_{j,2}^* = \gamma_{j,2} - \beta_{j,1}^* \gamma_{0,2} \qquad \beta_{j,3}^* = \gamma_{j,3} - \beta_{j,1}^* \gamma_{0,3}. \tag{19}$$

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

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

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

Thus, if we do not impose symmetry, the identified set will be of dimension three $(\beta^*_{1,1}, \beta^*_{2,1}, \beta^*_{3,1})$ and will be characterized by the eight γ parameters in (17)–(18). However, one restriction must be imposed on those parameters for the augmented model to characterize observationally equivalent values of the original β parameters satisfying the symmetry constraints. To see this, note that, subject to the cross-restrictions, (17)–(18) imply that (16) are satisfied as before for any $\beta^*_{1,1}$ (and for $\beta^*_{1,2}$ and $\beta^*_{1,3}$ as in (19)), but only for $\beta^*_{2,1} = \beta^*_{1,2}$ so that

$$\beta^*_{2,1} = \gamma_{1,2} - \beta^*_{1,1}\gamma_{0,2},$$

and for $\beta^*_{2,2}$ and $\beta^*_{2,3}$ such that

$$\beta^*_{2,2} = \gamma_{2,2} - (\gamma_{1,2} - \beta^*_{1,1}\gamma_{0,2})\gamma_{0,2},$$

$$\beta^*_{2,3} = \gamma_{2,3} - (\gamma_{1,2} - \beta^*_{1,1}\gamma_{0,2})\gamma_{0,3}.$$

Equally, they are satisfied only for $\beta^*_{3,1} = \beta^*_{1,3}$ so that

$$\beta^*_{3,1} = \gamma_{1,3} - \beta^*_{1,1}\gamma_{0,3},$$

and for $\beta^*_{3,2}$ and $\beta^*_{3,3}$ such that

$$\beta^*_{3,2} = \gamma_{3,2} - (\gamma_{1,3} - \beta^*_{1,1}\gamma_{0,3})\gamma_{0,2}$$

$$\beta^*_{3,3} = \gamma_{3,3} - (\gamma_{1,3} - \beta^*_{1,1}\gamma_{0,3})\gamma_{0,3}.$$

Moreover, the restriction $\beta^*_{3,2} = \beta^*_{2,3}$ implies that the admissible values of the coefficients in the augmented model must satisfy for any $\beta^*_{1,1}$:

$$\gamma_{3,2} - (\gamma_{1,3} - \beta^*_{1,1}\gamma_{0,3})\gamma_{0,2} = \gamma_{2,3} - (\gamma_{1,2} - \beta^*_{1,1}\gamma_{0,2})\gamma_{0,3},$$

or

$$\gamma_{3,2} - \gamma_{2,3} = \gamma_{1,3}\gamma_{0,2} - \gamma_{1,2}\gamma_{0,3}. \tag{20}$$

Thus, after enforcing symmetry the identified set is of dimension one $(\beta^*_{1,1})$ and depends on seven parameters only. The I test for this problem is a test of overidentifying restrictions based on the moments (17)–(18) subject to (20). Enforcing (20) reduces the set of observationally equivalent parameters under the null, but this is the right way to proceed since the existence of other β's that satisfy the instrumental-variable conditions but not the symmetry conditions should not be taken as evidence of underidentification of the model.[15]

3.3. Sequential moment conditions

Consider next an example with an explicit time series structure. The expectations are taken by averaging across individuals (over i).

Example 3.5. Suppose that

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

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

$$\alpha' y_{i,t+2} = u_{i,t+2},$$

where

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

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

$$E\left[z_{i,t}{y_{i,t+2}}'\right]\alpha = 0. \tag{21}$$

The dimension of the vector $z_{i,t}$ varies with t. This dependence is relevant in a panel data setting in which the number of time periods is small relative to the number of individuals.[16] Assume that there is no redundancy among the entries of $z_{i,t}$. That is, $E\left(z_{i,t}z'_{i,t}\right)$ is nonsingular. Moreover, assume that the entries of $z_{i,t-1}$ are among the entries of $z_{i,t}$.

For this model to be underidentified, we must be able to find an $\alpha^* \neq \alpha$, both distinct from zero, such that α^* also satisfies equation system (21). Since α and α^* are distinct and linear combinations of α and α^* must satisfy (21), it follows that

$$E\left[z_{i,t}y^{*\prime}_{i,t+1}\right]\gamma = 0 \tag{22}$$

for $t = 1, 2, \ldots$, where

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

and γ is not degenerate and has k entries.

Conversely, suppose that moment conditions (22) are in fact satisfied. Notice that

$$E\left[z_{i,t}y^{*\prime}_{i,t+2}\right]\gamma = 0$$

because

$$E\left[z_{i,t+1}y^{*\prime}_{i,t+2}\right]\gamma = 0,$$

where this latter equation is just (22) shifted one time period forward. As a consequence, both

$$\alpha = \begin{bmatrix} \gamma' & 0 \end{bmatrix}',$$

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

necessarily satisfy (21). Thus the I test for underidentification naturally leads us to test an alternative set of moment conditions with one less free parameter given by (22). Identification of the parameter vector α from (21) up to scale requires that we reject moment (22) up to scale.

In a panel data setting, the I test is built from the moment conditions (22) for $t = 1, 2, \ldots, T$ and large N. This construction of the I does not simply duplicate moment conditions, as this would lead to a degeneracy or repetition of moment conditions. Instead, the time series structure naturally leads to an alternative equation system to be studied. Also we could construct a collection of reduced form equations by projecting $y_{i,t+2}$ onto $z_{i,t}$ for each i and explore the restrictions imposed on coefficients. The reduced-form coefficients would necessarily be time dependent, and they would include some implicit redundancies. For this example, it is particularly convenient to work directly with the original structural equation system.

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

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

$$\alpha_1(v_{i,t+2} - \eta_i) = -\alpha_2(v_{i,t+1} - \eta_i) - \alpha_3(v_{i,t} - \eta_i) + u_{i,t+2} \quad (t = 3, \ldots, T), \tag{23}$$

and

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

Taking the first differences of Eq. (23) eliminates the fixed effect. Following Arellano and Bond (1991), consider GMM estimation of α_1 and α_2 based on a random sample $\{v_{i,1}, \ldots, v_{i,T} : i = 1, \ldots, N\}$ and the unconditional moment restriction:

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

Thus, we have a system of $T - 3$ equations with a set of admissible "instruments" that increases with T, but a common parameter

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

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

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

Returning to our original specification (23), suppose that $\alpha_1 + \alpha_2 + \alpha_3 = 0$. Then

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

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

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

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

Up until now we have considered only models that are linear in the variables. We extend this discussion to include models with nonlinearities. In this discussion, it is important to distinguish two cases. In the first case there is a separation between variables and parameters, and hence the nonlinearity is confined to the parameters. In the second case, the nonlinearities between variables and parameters interact in a more essential way.

4. Nonlinearity in the parameters

We first extend our previous analysis by replacing the parameter vector α by a nonlinear, continuously differentiable function $\phi : \mathbb{P} \to \mathbb{R}^{k+1}$ where $\mathbb{P}$ is the closure of an open set in $\mathbb{R}^{\ell}$. We study the nonlinear equation:

Assumption 4.1.

$$E(\Psi_t)\,\phi(\beta) = 0 \tag{24}$$

for some $\beta \in \mathbb{P}$.

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

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

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

We find it fruitful to think of the function ϕ as imposing restrictions on a parameter vector α through the mapping $\phi(\beta) = \alpha$. By thinking of α as the parameter to be estimated, we can use aspects of the approach described previously. Since ϕ is one-to-one, we can uncover a unique β for each α. This leads us to construct the parameter space:

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

Suppose now that two values $\beta^{[1]}$ and $\beta^{[2]}$ satisfy Assumption 4.1 and are distinct. Thus both $\phi(\beta^{[1]})$ and $\phi(\beta^{[2]})$ are in the null space of the matrix $E(\Psi_t)$. By Assumption 4.2, the vectors $\phi(\beta^{[1]})$ and $\phi(\beta^{[2]})$ are not proportional. Any two linear combinations of $\phi(\beta^{[1]})$ and $\phi(\beta^{[2]})$ must also be in the null space of $E(\Psi_i)$.

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

$$\mathbb{Q}^* \doteq \{\alpha : \alpha = \mathrm{c}_1\alpha_1 + \mathrm{c}_2\alpha_2, \alpha_1 \in \mathbb{Q}, \alpha_2 \in \mathbb{Q}, \mathrm{c}_1, \mathrm{c}_2 \in \mathbb{R}\}. \tag{25}$$

Notice that if

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

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

To illustrate how nonlinearity in parameters can alter the analysis, we use an example that is closely related to the non-linear IV model with serially correlated errors considered by Sargan (1959). Nevertheless, it differs in an important way because in our case the valid instrumental variables are predetermined but not necessarily strictly exogenous.[17]

Example 4.1. Consider a time series example:

$$x_t \cdot \beta_1 = u_t + \gamma_1' w_t,$$
$$u_t = \beta_2 u_{t-1} + \gamma_2' w_t, \tag{26}$$

where $\{w_t\}$ is a multivariate martingale difference sequence. Suppose also that z_{t-1} is a linear function of $w_{t-1}, w_{t-2}, \ldots$. The process $\{u_t\}$ is unobservable to the econometrician, but

$$x_t \cdot \beta_1 - \beta_2(x_{t-1} \cdot \beta_1) = (\gamma_1 + \gamma_2)'\, w_t - \beta_2\gamma_1' w_{t-1}.$$

Let

$$\Psi_t = \begin{bmatrix} z_{t-2}x_t' & -z_{t-2}x_{t-1}' \end{bmatrix},$$

and consider identification of β based on:

$$E(\Psi_t)\phi(\beta) = 0,$$

where

$$\phi(\beta) = \begin{bmatrix} \beta_1 \\ \beta_2\beta_1 \end{bmatrix}. \tag{27}$$

To achieve identification requires that we impose an additional normalization, say $|\beta_1| = 1$. We may wish to restrict $|\beta_2| < 1$. Since we have not restricted $\gamma_2' w_t$ to be uncorrelated with u_{t-1}, the unobserved (to the econometrician) process $\{u_t\}$ can be stationary and still satisfy Eq. (26). Thus when $|\beta_2| > 1$,

$$u_t = -\sum_{j=1}^{\infty} (\beta_2)^{-j}\, \gamma_2' w_{t+j}$$

is a stationary process that satisfies (26). Notice, however, in this case $u_t + \gamma_1' w_t$ is orthogonal to z_{t-1} so there is an additional moment restriction at our disposal. As is well known the case of $|\beta_2| = 1$ requires special treatment.

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

$$\beta_1^* = \mathrm{c}\beta_1 + \mathrm{d}\eta_1 \tag{28}$$

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

In line with the linear case assume that $\mathrm{rank}[E(\Psi_t)] = k - 1$ so that it has a two-dimensional null space. This means that if there are other observationally equivalent structures, they must satisfy

$$E(\Psi_t)\begin{bmatrix} \mathrm{c}\beta_1 + \mathrm{d}\eta_1 \\ \mathrm{c}\beta_2^*\beta_1 + \mathrm{d}\beta_2^*\eta_1 \end{bmatrix} = 0. \tag{29}$$

Given the partly linear and partly non-linear structure of the model, underidentification emerges in three ways that we now consider.

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

4.1. Only β_1 identified:

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

$$E\left(z_{t-2}x_{t-1}'\right)\beta_1 = 0,$$

and hence

$$E\left(z_{t-1}x_t'\right)\beta_1 = 0 \tag{30}$$

for some β_1. This phenomenon can occur for one of two reasons. First perhaps the choice z_{t-2} is unfortunate. Alternatively, $x_t \cdot \beta_1$ may depend only on current and possibly future values of the martingale difference sequence $\{w_t\}$. As we have seen, this may happen when $|\beta_2| > 1$ or in the degenerate case when u_t is identically zero ($\gamma_1 = 0$).[18] For this same β_1, it is also required that

$$E\left(z_{t-2}x_t'\right)\beta_1 = 0.$$

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

$$E\left(z_{t-1}^* x_t'\right)\beta_1 = 0.$$

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

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

$$\pi(\theta) = \begin{bmatrix} \beta_1 \\ \theta \end{bmatrix}.$$

This test is closely related but not identical to the underidentification test proposed by Sargan (1959) for the non-linear in parameters model that he studied. The augmented set of moment conditions that he considered were (30) and

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

where he implicitly chose β_2^* so that the sample covariance matrix of $x_t'\beta_1$ and $(x_t' - \beta_2^* x_{t-1}')\beta_1^*$ were 0. Apart from our emphasis on symmetric normalization and robustness to serial correlation and heteroskedasticity, the main difference with his approach is that we impose the restriction $\beta_1 = \beta_1^*$, which, in parallel with a gain in estimation efficiency, leads to a reduction in the number of degrees of freedom and the resulting gain in power, and also eliminates the need to choose two arbitrary values for β_2.

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

4.2. Only β_2 is identified:

As another alternative suppose there is a vector $\beta_1^* \neq \beta_1$ such that

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

satisfies the moment conditions:

$$E(\Psi_t)\alpha^* = 0.$$

Since any linear combination of α and α^* must satisfy moment conditions, we can choose $c = 0$ in (28) so that

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

should also satisfy the moment conditions (29). This gives rise to a second I test. We parameterize two orthonormal directions η_1 and β_1 along with a *single* parameter β_2. When β_1 has only two components, we are free to set β_1 and η_1 equal to the two coordinate vectors and freely estimate only the parameter β_2. In that case the moment conditions of the I test can be expressed as

$$E[z_{t-2}(x_{i,t} - \beta_2 x_{i,t-1})] = 0, \quad i = 1, 2.$$

More generally, under the null hypothesis associated with this I test there is a two-dimensional plane of (non-normalized) values of the original parameter vector β_1 that satisfy the moment conditions, but only one value of β_2. After normalization, the manifold of observationally equivalent structures will be given by (28), and hence we may represent the observationally equivalent β's via:

$$\pi(\theta) = \begin{bmatrix} \theta\beta_1 + \sqrt{1-\theta^2}\eta_1 \\ \beta_2 \end{bmatrix}$$

for $-1 \leq \theta \leq 1$ where we introduce additional restrictions that permit us to identify β_1 and η_1 used to represent π. Note that if $E[z_{t-2}(x_{i,t} - \beta_2 x_{i,t-1})] = 0$ for some i, then all the β_1 coefficients will be identified except the one corresponding to $x_{i,t}$.

Importantly, this test is different from a linear test of $\text{rank}[E(\Psi_t)] = k - 1$ derived along the lines of Section 3.1, since such a test would not impose that the observationally equivalent structures must satisfy (27).

Once again, as a by-product of our procedure we will obtain efficient GMM estimators of β_2, and the parameters β_1 and η_1 that characterize the identified set through (28).

4.3. Another possibility

In the two previous cases, we constructed functions π with realized values that satisfied the moment conditions. Another possibility is that the $\text{rank}[E(\Psi_t)] = k - 1$ but that there are only two distinct parameter values in $\mathbb{P}$, say $\beta^{[1]}$ and $\beta^{[2]}$ that satisfy:

$$E(\Psi_t)\phi(\beta) = 0.$$

In this case there is still a two-dimensional subspace of $\mathbb{Q}^*$ constructed in (25). With an additional normalization, obtained, say, by restricting the magnitude of the vector β_1 to have a norm equal to one, the curve is reduced to one dimension.[19]

5. Fundamental nonlinearity

In this section we explore the underidentification problem when there is a more fundamental nonlinearity of the parameters in the moment conditions. Recall that in the linear model discussed

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

[19] In the first-order underidentified case studied by Sargan (1983a), there is only one β_0 that satisfies the moment conditions (24) even though the rank of the matrix $\{E[\Psi_t]\partial\phi(\beta_0)/\partial\beta'\}$ is less than ℓ. This case can be regarded as the limit of the isolated two-points case in which $\beta^{[1]}$ and $\beta^{[2]}$ get closer and closer to each other in such a way that the dimension of the nullspace of $E(\Psi_t)$ remains two.

in Section 3, underidentification implies that we can estimate a line, which we chose to implicitly parameterize by means of two parameter vectors. Similarly, in the non-linear in parameters model discussed in Section 4, we also implicitly parameterize a curve as a function of a finite number of parameters. The natural extension for a fully nonlinear model is to estimate a one-dimensional curve.

What follows are two distinct and largely self contained subsections. The first one extends the efficiency problem posed in Section 2. The second subsection explores the consistent estimation of one-dimensional curves and suggests a resulting test for under-identification.

5.1. Efficiency reconsidered

In this subsection we establish a pointwise (in θ) efficiency bound for GMM using an approach similar to Hansen (1985) that extends the analysis of Section 2, where we constructed the GMM efficiency bound for a finite set Θ_n.

Let Θ be a compact interval. We impose the following extensions to Assumptions 2.1 and 2.2:

Assumption 5.1. $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f[x_t, \pi_0(\theta)]$ converges in distribution to a Gaussian process $\{g(\theta) : \theta \in \Theta\}$ with mean zero for all $\theta \in \Theta$ and covariance function:

$$K(\theta, \vartheta) = \lim_{T\to\infty} \frac{1}{T} E\left(\sum_{t=1}^{T}\sum_{s=1}^{T} f[x_t, \pi_0(\theta)] f[x_s, \pi_0(\vartheta)]'\right)$$

where the covariance function K is continuous in its two arguments.

A sufficient condition for this assumption is that $\left\{\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[f(x_t, \beta) - \bar{f}(\beta)\right] : \beta \in \mathbb{P}\right\}$ converges in distribution to a Gaussian random element.

Assumption 5.2. $D(\theta)$ is continuous on Θ and $D(\theta)'D(\theta)$ is non-singular for all $\theta \in \Theta$.

As before, we construct:

$$h(\theta) = [D(\theta)'D(\theta)]^{-1}D(\theta)'g(\theta),$$

for all $\theta \in \Theta$. Let $\{\theta_j : j = 1, 2, \ldots.\}$ be a dense sequence in Θ. Construct $\mathcal{F}$ as the mean square closure of $\bigcup_{n=1}^{\infty} \mathcal{F}_n$, where as before

$$\mathcal{F}_n = \left\{ y = \sum_{j=1}^{n} B_j' g(\theta_j) : D(\theta_j)'B_j = 0,\ j = 1, 2, \ldots, n \right\}.$$

The sequence approach gives us one means of approximating the efficiency bound.

Proposition 5.1. *The GMM efficiency bound for estimating $h(\theta_j)$ is*

$$E\left[\left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}]\right)\left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}]\right)'\right].$$

Proof. From least squares theory

$$\lim_{n\to\infty} \text{Proj}\left[h(\theta_j)|\mathcal{F}_n\right] \to \text{Proj}\left[h(\theta_j)|\mathcal{F}\right],$$

where the convergence is in mean-square. It follows that

$$\lim_{n\to\infty} E\left[\left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}_n]\right)\left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}_n]\right)'\right]$$
$$= E\left[\left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}]\right)\left(h(\theta_j) - \text{Proj}[h(\theta_j)|\mathcal{F}]\right)'\right]. \quad \square$$

The precise choice of the sequencing is important in practice, but it does not alter $\mathcal{F}$.

Lemma 5.2. *For any $\theta \in \Theta$ and vector B such that $D(\theta)'B = 0$, $B'g(\theta)$ is in $\mathcal{F}$.*

Proof. This result follows immediately for θ a member of the sequence $\{\theta_j : j = 1, 2, \ldots.\}$. More generally, consider a subsequence $\left\{\theta_{j(k)} : k = 1, 2, \ldots\right\}$ that converges to θ, and compute

$$E\left(\left|g[\theta_{j(k)}] - g(\theta)\right|^2\right) = \text{trace}\Big[K\left(\theta_{j(k)}, \theta_{j(k)}\right) - K\left(\theta_{j(k)}, \theta\right) - K\left(\theta, \theta_{j(k)}\right) + K\left(\theta, \theta\right)\Big].$$

The right-hand side converges to zero since the covariance function K is continuous in both arguments, and hence so does the left-hand side. Let

$$B_k = B - D\left(\theta_{j(k)}\right)\left[D\left(\theta_{j(k)}\right)' D\left(\theta_{j(k)}\right)\right]^{-1} D\left(\theta_{j(k)}\right)' B.$$

Notice that $D\left(\theta_{j(k)}\right)' B_k = 0$ and that $\{B_k : k = 1, 2, \ldots\}$ converges to B since $\left\{D\left(\theta_{j(k)}\right) : k = 1, 2, \ldots\right\}$ converges to $D(\theta)$ and $D(\theta)'B = 0$. It follows that

$$\lim_{k\to\infty} E\left(|B_k' g\left(\theta_{j(k)}\right) - B'g(\theta)|^2\right) = 0$$

since

$$\left[E\left(\left|B_k' g\left[\theta_{j(k)}\right] - B'g(\theta)\right|^2\right)\right]^{1/2}$$
$$\le |B_k - B|\left[E\left(\left|g\left(\theta_{j(k)}\right)\right|^2\right)\right]^{1/2}$$
$$+ |B|\left[E\left(\left|g\left(\theta_{j(k)}\right) - g(\theta)\right|^2\right)\right]^{1/2},$$

$\{g\left(\theta_j\right) : j = 1, 2, \ldots.\}$ converges in mean-square to $g(\theta)$ and $\{B_k : k = 1, 2, \ldots\}$ converges to B. $\square$

As a direct extension of Proposition 5.1 and Lemma 5.2, our calculations apply jointly to any finite number of θ's. The construction of $\mathcal{F}$ remains the same.

Theorem 5.3. *The GMM efficiency bound for any finite collection of θ's, $\theta_1, \theta_2, \ldots, \theta_m$ is given by $E(YY')$ where*

$$Y = \begin{bmatrix} h(\theta_1) - \text{Proj}\left[h(\theta_1)|\mathcal{F}\right] \\ h(\theta_2) - \text{Proj}\left[h(\theta_2)|\mathcal{F}\right] \\ \cdots \\ h(\theta_m) - \text{Proj}\left[h(\theta_m)|\mathcal{F}\right] \end{bmatrix}.$$

While our GMM efficiency bound applies to any finite collection of points along a curve, it is suggestive of a more general result. As an immediate corollary to the previous theorem, consider the efficiency bound for estimating

$$\sum_{j=1}^{m} \gamma_j' \tau(\theta_j). \tag{31}$$

This bound is given by the variance in the forecast error:

$$\sum_{j=1}^{m} \gamma_j' h(\theta_j) - \text{Proj}\left[\sum_{j=1}^{m} \gamma_j' h(\theta_j)|\mathcal{F}\right].$$

Consider now the efficiency bound for estimating

$$\int_{\Theta} \gamma(\theta) \cdot \tau(\theta) d\theta.$$

Reimann sum approximations to this integral can be represented in the form given in (31). Provided that we can construct an

approximating sequence

$$E\left[\left|\sum_{j=1}^{m}\gamma_j'\tau(\theta_j)-\int_{\Theta}\gamma(\theta)\cdot\tau(\theta)d\theta\right|^2\right]$$

that converges to zero, we conjecture that the efficiency bound is the variance of

$$\int_{\Theta}\gamma(\theta)'g(\theta)d\theta-\text{Proj}\left[\int_{\Theta}\gamma(\theta)\cdot g(\theta)d\theta\,\middle|\,\mathcal{F}\right].$$

This result is a nontrivial extension of our "pointwise" calculations, but a formal proof of it is beyond the scope of this paper.

5.2. Estimation of curves

We now adopt a more global perspective by studying the statistical consistency of GMM estimators of the function $\pi:\Theta\to\mathbb{P}$. Conveniently, this estimation problem looks like a standard problem except that we seek to estimate a function instead of a finite-dimensional parameter vector.

Assumption 5.3. Let $\mathbb{P}$ be a compact subset of $\mathbb{R}^k$.

As in Section 2, introduce a function $f(x,\cdot):\mathbb{P}\to\mathbb{R}^p$ for each x. The function f is jointly Borel measurable and at the very least continuous in its second argument for each value of x. Thus $f(x_t,\cdot)$ is a p-dimensional random function on $\mathbb{P}$ or a random element.

Assumption 5.4. $E|f(x_t,\beta)|<\infty$ for each $\beta\in\mathbb{P}$.

This assumption justifies the definition of $\bar f(\beta)=E[f(x_t,\beta)]$ for each $\beta\in\mathbb{P}$.
As in Hansen (1982a), we also assume:

Assumption 5.5. $f(x_t,\cdot)$ is first-moment continuous for each $\beta\in\mathbb{P}$.

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

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

Assumption 5.6. Π is a compact set of admissible functions defined using the supnorm.

From the Arzelà–Ascoli Theorem it suffices that there be a uniform bound on the functions in Π and that the functions be equicontinuous. The uniform bound comes from the compactness of $\mathbb{P}$ (Assumption 5.3).

Consider next first-moment continuity. Notice that

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

provided that $\|\pi-\tilde\pi\|\le\epsilon$. This simple inequality implies that the first-moment continuity restriction given in Assumption 5.5 extends to the parameter space Π.

Let $\mathbb{C}$ denote the space of all continuous functions from Θ into $\mathbb{R}^p$ equipped with the sup-norm. Let g be a continuous function mapping $\mathbb{P}$ into $\mathbb{R}^p$. Then for any $\pi\in\Pi$, the composition $g\circ\pi$ is in $\mathbb{C}$. Thus we may view $g\circ\pi$ as a continuous function mapping Π into $\mathbb{C}$. In particular, the functions $f[x,\pi(\cdot)]$ for each x and $f[\pi(\cdot)]$ are continuous in π on the parameter space Π. Since Π is compact, these functions are in fact uniformly continuous.

Proposition 5.4. *Suppose that Assumptions 5.3–5.6 are satisfied. Then*

$$\sup_{\pi\in\Pi}\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T}f[x_t,\pi(\theta)]-\bar f[\pi(\theta)]\right|$$

converges to zero almost surely.

Proof. See Hansen (1982b). □

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

Assumption 5.7. $\bar f[\pi(\theta)]=0$ for all $\theta\in\Theta$ if and only if $\pi=\pi_0$.

This assumption rules out the possibility that there exists $\tilde\pi$ such that

$$\{\tilde\pi(\theta)|\theta\in\Theta\}\subset\{\pi_0(\theta)|\theta\in\Theta\}\tag{32}$$

for some $\tilde\pi\neq\pi_0$, in which case there would exist two functions in this closure for which the image of one function is a proper subset of the other. Note that (32) is ruled out *a priori* if we use parameterization (2) given by:

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

because the first coordinate of π is allowed to vary.

To obtain a consistent estimator of π_0, we introduce a positive definite quadratic form $\mathcal{W}$ which plays the role of the "weighting matrix" in GMM estimation. We let $\mathbb{L}^2$ denote the space of Borel measurable functions ϕ that map Θ into $\mathbb{R}^p$, where the coordinate functions of ϕ are restricted to be square integrable using a conveniently chosen measure on Θ. We use a Lebesgue measure on Θ and $(\int_\Theta|\phi(\theta)|^2\,d\theta)^{1/2}$ as norm, but in some applications other measures may turn out to be more convenient. The quadratic form, $\mathcal{W}$, maps $\mathbb{L}^2\times\mathbb{L}^2$ into $\mathbb{R}$. Quadratic forms satisfy:

(i) $\mathcal{W}(\phi,\psi)=\mathcal{W}(\psi,\phi)$ for $\phi,\psi\in\mathbb{L}^2$;
(ii) $\mathcal{W}(\mathrm{r}\phi,\psi)=\mathrm{r}\mathcal{W}(\phi,\psi)$ for $\phi,\psi\in\mathbb{L}^2$, $\mathrm{r}\in\mathbb{R}$;
(iii) $\mathcal{W}(\phi_1+\phi_2,\psi)=\mathcal{W}(\phi_1,\psi)+\mathcal{W}(\phi_2,\psi)$ for $\phi_1,\phi_2,\psi\in\mathbb{L}^2$.

A positive semidefinite form satisfies $\mathcal{W}(\phi,\phi)\ge0$. We strengthen this restriction by imposing positive definiteness and boundedness:

Assumption 5.8. For any $\phi\neq0$ in $\mathbb{L}^2$

$$\mathcal{W}(\phi,\phi)>0,$$

and for some positive number $\bar{\mathrm{b}}$

$$\mathcal{W}(\phi,\phi)\le\bar{\mathrm{b}}\int_{\Theta}|\phi(\theta)|^2d\theta$$

for all $\phi\in\mathbb{L}^2$.

A positive definite form defines an alternative norm on $\mathbb{L}^2$ constructed as $[\mathcal{W}(\phi,\phi)]^{1/2}$ and the form itself defines an alternative inner product. The upper bound in this assumption guarantees that $\mathcal{W}$ is $\mathbb{L}^2$ continuous. To see this observe that

$$|\mathcal{W}(\psi,\psi)-\mathcal{W}(\phi,\phi)|\le\mathcal{W}(\psi-\phi,\psi-\phi)+2|\mathcal{W}(\phi,\psi-\phi)|.$$

Positive semidefinite forms satisfy the Cauchy–Schwarz Inequality and hence

$$|\mathcal{W}(\phi,\psi-\phi)|\le[\mathcal{W}(\phi,\phi)]^{1/2}\,[\mathcal{W}(\psi-\phi,\psi-\phi)]^{1/2}.$$

The $\mathbb{L}^2$ continuity of $\mathcal{W}$ now follows from

$$\mathcal{W}(\psi-\phi,\psi-\phi)\le\bar{\mathrm{b}}\int_{\Theta}|\psi(\theta)-\phi(\theta)|^2d\theta.$$

Consider an estimator π_T that solves:

Problem 5.5. Let π_T be a solution to:

$$\min_{\pi\in\Pi} \mathcal{W}\left(\frac{1}{T}\sum_{t=1}^{T} f[x_t,\pi(\cdot)], \frac{1}{T}\sum_{t=1}^{T} f[x_t,\pi(\cdot)]\right).$$

The following result establishes the consistency of π_T:

Theorem 5.6. *Suppose that Assumptions 5.3–5.8 are satisfied. Then*

$$\sup_{\theta\in\Theta} |\pi_T(\theta) - \pi_0(\theta)|$$

converges to zero almost surely as T gets large.

Proof. Let F be a continuous function mapping $\mathbb{P}$ into $\mathbb{R}^p$. Recall that we may view $F[\pi(\cdot)]$ as a continuous function mapping Π into $\mathbb{C}$. In addition we may view $\mathcal{W}(F[\pi(\cdot)], F[\pi(\cdot)])$ as a continuous function mapping Π into the nonnegative real numbers. From Proposition 5.4 $\left\{\frac{1}{T}\sum_{t=1}^{T} f(x_t,\pi) : T = 1,2,\ldots\right\}$ converges uniformly in π to the continuous function $\bar{f}(\pi)$. Thus $\mathcal{W}\left(\frac{1}{T}\sum_{t=1}^{T} f[x_t,\pi(\cdot)], \frac{1}{T}\sum_{t=1}^{T} f[x_t,\pi(\cdot)]\right)$ converges to $\mathcal{W}\left(\bar{f}[\pi(\cdot)], \bar{f}[\pi(\cdot)]\right)$ uniformly in π almost surely. It follows that the set of minimizers of $\mathcal{W}\left(\frac{1}{T}\sum_{t=1}^{T} f[x_t,\pi(\cdot)], \frac{1}{T}\sum_{t=1}^{T} f[x_t,\pi(\cdot)]\right)$ converges in the Hausdorff metric over compact subsets of Π to the unique minimizer of $\mathcal{W}\left(\bar{f}[\pi(\cdot)], \bar{f}[\pi(\cdot)]\right)$, which is π_0. □

This consistency argument presumes that we have full flexibility in constructing the function π. In practice we have to approximate this function, for instance by selecting grid points and building a smooth function that passes through these points via splines. Formally our consistency argument allows for this construction to be arbitrarily refined. Once we explore ways to approximate the efficiency bound, important implementation issues come into play that require more care be paid to this approximation. As this is an important practical issue, we will suggest an approach without fully analyzing the implications for inference.

What follows are examples of forms that interest us.

Example 5.1. Consider

$$\mathcal{W}(\phi,\psi) = \int_\Theta \phi(\theta)' W(\theta)\psi(\theta)d\theta$$

where $W(\theta)$ is continuous and positive definite for each $\theta \in \Theta$ and $W(\theta) \leq \bar{b}I$. In this case the GMM minimization Problem 5.5 is separable in θ and the minimization can be done θ by θ. One possibility is to let

$$W(\theta) = K(\theta,\theta)^{-1}$$

which is known to achieve the familiar GMM efficiency bound, but this bound ignores the cross θ restrictions.

Example 5.2. Let $\{\varphi_j : j = 1,2,\ldots\}$ denote an orthonormal sequence of functions in $\mathbb{L}^2$ with a span given by the entire space. Form

$$\mathcal{W}(\phi,\psi) = \sum_{j=1}^{\infty} \zeta_j \int_\Theta \varphi_j(\theta)\cdot\phi(\theta)d\theta \int_\Theta \varphi_j(\theta)\cdot\psi(\theta)d\theta,$$

and $\underline{b} \leq \zeta_j \leq \bar{b}$ for $j = 1,2,\ldots$. Then

$$\phi = \sum_{j=1}^{\infty} \varphi_j \int_\Theta \phi(\theta)\cdot\varphi_j(\theta)d\theta$$

where the convergence of the infinite sum is in the $\mathbb{L}^2$ norm given by $(\int_\Theta |\phi(\theta)|^2\, d\theta)^{1/2}$, and hence the Parceval formula implies that

$$\int_\Theta \phi(\theta)\cdot\phi(\theta)d\theta = \sum_{j=1}^{\infty}\left[\int_\Theta \phi(\theta)\cdot\varphi_j(\theta)d\theta\right]^2.$$

As a consequence, $\mathcal{W}$ satisfies Assumption 5.8.

This second construction allows for efficiency gains via joint estimation. One possibility is to follow Carrasco and Florens (2000) by using the form:

$$\mathcal{K}(\phi,\psi) = \int_\Theta\int_\Theta \phi(\theta)' K(\theta,\vartheta)\psi(\vartheta)d\theta d\vartheta.$$

Construct the functions ω_j and numbers λ_j as eigenfunctions and eigenvalues of the form $\mathcal{K}$

$$\mathcal{K}(\omega_j,\phi) = \lambda_j \int_\Theta \omega_j\cdot\phi$$

for all $\phi \in \mathbb{L}^2$. For a different estimation problem, Carrasco and Florens (2000) suggest constructing a form that "inverts" the eigenvalues as a counterpart to inverting a weighting matrix. This leads to the GMM objective:

$$\sum_{j=1}^{\infty} \zeta_j \left(\int_\Theta \omega_j(\theta)\cdot\frac{1}{T}\sum_{t=1}^{T} f[x_t,\pi(\theta)]d\theta\right)^2 \tag{33}$$

where $\zeta_j = \frac{1}{\lambda_j}$. As they note this particular choice is problematic because the ζ_j's are unbounded since the λ_j's converge to zero. This phenomenon leads them to alternative choices based on regularization such as:

$$\zeta_j = \frac{\lambda_j}{\lambda_j^2 + \varsigma}$$

or

$$\zeta_j = \frac{1}{\lambda_j + \varsigma},$$

where $\varsigma > 0$ is a regularization parameter. Notice that the ζ_j's are now bounded. If we hold fixed ς as a function of sample size we will distort the efficiency, but we conjecture that by making ς small we will approximate the efficiency bound we discussed previously.[20] Alternatively, we could follow Carrasco and Florens (2000) and consider a framework in which the regularization parameter diminishes as the sample size gets larger in such a manner as to achieve the GMM efficiency bound.[21]

It is interesting to relate the inferential problems in the previous sections with the one in this section. The main difference is that the structure of (8) and (24) implies that the resulting form $\mathcal{K}$ will only have a finite number of positive eigenvalues. Once we take this fact into account, though, the curves that we will estimate with the procedure that we have developed in this section will coincide with the curves that we implicitly estimated using the procedures developed in Sections 3 and 4.

To see why, consider for instance the linear in parameters model (8), and suppose that instead of (6) we seek to estimate a non-linear parametric curve with the following structure

$$\pi(\theta) = \theta\cdot\alpha^{[1]} + (1-\theta)\cdot\alpha^{[2]} + \begin{bmatrix}\theta\\ \upsilon(\theta)\end{bmatrix}. \tag{34}$$

Further, assume that $\pi(\theta)$ can be uniquely identified from the continuum of moment conditions (3). We know that for each possible υ the linear span of the image will be finite-dimensional. As we show in Appendix B, the method proposed in this section will select $\upsilon(\theta) = 0\ \forall\theta$ in order to keep the dimension of the linear span as small as possible, in this case two.

[20] This conjecture is not entirely obvious because we are now imposing compactness in the parameter space whereas previously we ignored this restriction.

[21] We cannot simply appeal to the results in Carrasco and Florens (2000) because they consider estimation of a finite-dimensional parameter vector with a continuum of moment conditions, which is a different estimation problem.

5.3. Testing

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

More generally, an overidentification test could be constructed analogously to that of Carrasco and Florens (2000) by scaling appropriately the minimized regularized sample counterpart to (33). The resulting test could produce a normal distribution as the limit of a sequence of appropriately centred and scaled (approximate) chi-square distributions with an arbitrarily large number of degrees of freedom. An attractive alternative approximation that incorporates the role of regularization leads instead to an approximate quadratic form in normal variables. Our experience suggests that this alternative Imhof (1961)-style approximation described in Appendix C is an improvement over the limiting normal distribution.

6. An empirical illustration with asset returns

In this section we will illustrate our methods using two versions of consumption-based capital asset pricing models. The representative agent (RA) version of these models with isoelastic preferences was originally estimated by GMM by Hansen and Singleton (1982) using aggregate consumption data.[22] From aggregate data we know that J tests are typically large except in cases studied by Hansen and Singleton (1996) that focus on aggregate equity returns (see also Hansen et al., 1996; Stock and Wright, 2000 and Kleibergen, 2005).[23] When stocks and bonds alone are used to construct two moment conditions, the model is at best exactly identified, but the coefficient of risk aversion is typically very large, which is evidence of the so-called *equity premium puzzle* coined by Mehra and Prescott (1985). To illustrate our procedures, we consider such a specification except that we use microeconomic data to construct an aggregate stochastic discount factor (SDF). Specifically, we follow a recent paper by Kocherlakota and Pistaferri (2009), which contrasts the implications of the RA model with one in which consumption and portfolio choices conform to a Pareto optimal allocation under private information (PIPO). When Kocherlakota and Pistaferri (2009) use the PIPO model, they find a substantially lower value for the coefficient of relative risk aversion.

For estimation purposes, they considered the following moment conditions:

$$E\left[f(x_t, \gamma, \rho)\right] = E\left[R_t k(x_t, \gamma) - \rho \iota_2\right] = 0, \tag{35}$$

where ι_2 is a two-dimensional vector of ones, R_t is a two-dimensional vector of gross returns on stocks and bonds denoted as:

$$R_t = \begin{bmatrix} R_{m,t} \\ R_{f,t} \end{bmatrix}$$

and $k(x_t, \gamma)$ is a model dependent kernel given by

$$k(x_t, \gamma) = \begin{cases} \left(N^{-1}\sum_{i=1}^{N} c_{i,t}\right)^{-\gamma} \Big/ \left(N^{-1}\sum_{i=1}^{N} c_{i,t-3}\right)^{-\gamma} & \textbf{RA} \\ \left[N^{-1}\sum_{i=1}^{N} (c_{i,t-3})^{\gamma}\right] \Big/ \left[N^{-1}\sum_{i=1}^{N} (c_{i,t})^{\gamma}\right] & \textbf{PIPO}. \end{cases}$$

In the formula for $k(x_t, \gamma)$, $c_{i,t}$ is consumption of household i $(1, \ldots, N_t)$ over "month" t $(1, \ldots, T)$. The parameters of interest are (γ, ρ), with ρ^{-1} corresponding to the subjective discount factor and γ to the coefficient of relative risk aversion. The implied SDFs used to represent asset prices are $\rho^{-1}k(x_t, \gamma)$ for the alternative specifications of $k(x_t, \gamma)$. These SDFs both discount the future and adjust for risk.

To study underidentification, we consider the functions $\rho_m(\gamma) = E[R_{m,t}k(x_t, \gamma)]$ and $\rho_f(\gamma) = E[R_{f,t}k(x_t, \gamma)]$. If $\rho_m(\gamma) \neq \rho_f(\gamma)$ for all admissible values of γ except γ_0, then γ_0 and ρ_0 are identified. Alternatively, if $\rho_m(\gamma) = \rho_f(\gamma) = \rho(\gamma)$ for γ's in an interval, then the model is underidentified, and we seek to identify the values of ρ that correspond to each value of γ. To justify this latter identification, we consider an overidentification test applied to the problem of estimating ρ as a function of γ for an interval of γ's.

It is instructive to contrast this approach with a method of constructing a confidence interval for γ using a GMM criterion function. Suppose the model is point identified and consider a GMM estimator of ρ for alternative values of γ in an interval. For each such value of γ, there are two moment conditions that can be used to estimate the corresponding ρ. Suppose now we consider the γ's for a continuously-weighted GMM objective function that are below a pre-specified threshold, where the GMM objectives are evaluated at the minimizing choices of γ.[24] The threshold is set using the appropriate quantile of a chi-square one distribution. Such a procedure leads not only to a confidence interval for γ but also to a curve that depicts the minimizing values of ρ corresponding to each value of γ in the confidence interval. This curve depicts a potentially interesting tradeoff in the empirically relevant parameter values.

In contrast, suppose the model is in fact underidentified. For the reasons given in this paper, it is more efficient to estimate this curve by stacking all of the relevant moment conditions, or more precisely by studying the resulting continuum of moment conditions simultaneously. Under the perspective that the model is underidentified, we have a more efficient way to estimate the curve depicting the tradeoff in parameters resulting from the empirical evidence. Under this perspective we could use the I test threshold instead of the J test threshold to help in the determination of the length of the interval.

6.1. Implementation

One simple way of implementing our approach is by means of the following discrete grid procedure.

1. Choose n values of γ, denoted $\gamma_1, \ldots, \gamma_n$ spaced within the interval of interest;
2. Replicate n times the moment conditions (35) evaluating them at γ_j and $\rho(\gamma_j)$;
3. Estimate the parameters $\rho(\gamma_1), \ldots, \rho(\gamma_n)$ using efficient GMM.

Since we restrict ourselves to a grid of points, the method only approximates the efficiency bound derived in Section 5 and only

[22] This application is closely related to Example 3.3.

[23] Hansen and Singleton (1996) use a recursive model of preferences with a unitary coefficient of risk aversion to motivate the identification of the elasticity of intertemporal substitution with the return on the wealth portfolio.

[24] The resulting γ's need not be a single interval, but suppose for sake of illustration we consider only the interval surrounding the GMM point estimate.

provides estimates of the curve evaluated at a finite number of points. For n fixed, however, the corresponding overidentification test will be valid, testing now the null hypothesis of underidentification. When the values of γ are close to each other, the optimal weighting matrix may be close to singular in finite samples. We find it best to apply some regularization procedure, such as "ridge" or Tikhonov pseudo inverses.[25] The overidentification test associated with a regularized procedure will also be asymptotically χ^2_n if the regularization parameter goes to zero at a suitable rate, but its finite sample distribution may be better approximated by a quadratic form in normal variables, as described in Appendix C.

An alternative way of approximating the efficiency bound discussed in Section 5 is as follows.

1. Choose $H+1$ values of γ, denoted $\gamma_1, \ldots, \gamma_{H+1}$ spaced within the interval of interest;
2. Parameterize $\rho(.)$ a continuous, twice differentiable function of γ using what is referred to as a *natural* cubic spline with knots at $\gamma_1, \ldots, \gamma_{H+1}$. The spline introduces $H+1$ parameters to be estimated. (See Appendix E for details.)
3. Use the Carrasco and Florens (2000) procedure to estimate the free parameters of the spline using (35) to construct a continuum of moment conditions indexed by γ.

Since the spline only approximates the curve, the number of knot points needs to increase with the sample size to avoid asymptotic bias in the curve estimation. As before, a quadratic form in normal variables might offer an attractive alternative to using a standard normal for the limiting distribution of the resulting test statistic.

In practice, there are several choices that one needs to make in order to implement the Carrasco and Florens (2000) procedure. In particular, one must choose the regularization scheme and an associated regularization parameter, an estimator of the covariance operator (e.g. centered or uncentered), and the estimator at which the covariance operator will be evaluated. See Appendix D for more details.

The discrete grid approach gives consistent estimators of a finite number of points along the curve while the spline approach gives an estimator of the curve. We defer formal analysis of both methods as n and H get large as a function of the sample size for future research.

In the next subsection we compare these two methods to one in which we estimate ρ as a function of γ for each γ using a pointwise optimal GMM estimator constructed from the two moment conditions in (35). By changing the value of γ we trace a curve. This estimator ignores efficiency gains that are available through joint estimation for alternative values of γ, but nevertheless it provides a convenient benchmark.

6.2. Empirical results

We follow Kocherlakota and Pistaferri (2009) in using $T = 288$ months of the rotating US Consumer Expenditure (CEX) panel over the period 1980–2004. We consider in turn results for the two models.

[25] Let M_T denote a consistent estimator of the long run covariance matrix of the $2n$ replicated moment conditions, and denote by $W\Delta W'$ its spectral decomposition. Tikhonov regularization involves replacing $M_T^{-1} = W\Delta^{-1}W'$ by $W\Delta^{1/2}[\varsigma_T I_{2n} + \Delta^2]^{-1}\Delta^{1/2}W'$, while "ridge" regularization uses $W[\upsilon_T I_{2n} + \Delta]^{-1}W'$ instead, where ς_T and υ_T are some small positive regularization parameters that should tend to 0 with the sample size. While Tikhonov replaces δ_i^{-1} by $\delta_i/(\varsigma_T + \delta_i^2)$, ridge uses $1/(\upsilon_T + \delta_i)$, both of which remain bounded as $\delta_i \to 0$. As a result, the damping factors are $\delta_i^2/(\varsigma_T + \delta_i^2)$ and $\delta_i/(\upsilon_T + \delta_i)$, respectively, which are monotonic functions of δ_i that are 0 for $\delta_i = 0$ and go to 1 when $\delta_i \to \infty$ or the regularization parameter goes to 0. One advantage of the ridge procedure, though, is that it preserves the ordering of the non-regularized weights. In contrast, the maximum Tikhonov weight occurs at $\delta = \sqrt{\varsigma_T}$. For $\delta's$ larger than $\sqrt{\varsigma_T}$ the Tikhonov weights are decreasing in δ, while they are increasing in δ when $\delta < \sqrt{\varsigma_T}$, which may lead to noticeable differences in parameter estimates in finite samples, especially if M_T is not too ill conditioned.

Table 1
Parameter estimates and standard errors. Discrete grid estimators and standard errors (s.e.) are based on an optimal weighting matrix that uses ridge regularization with parameter $\varsigma = 0.15$.

γ	Pointwise efficient		Discrete grid	
	ρ^{-1}	s.e.	ρ^{-1}	s.e.
50	0.435	0.044	0.442	0.032
51.25	0.419	0.044	0.424	0.033
52.5	0.404	0.044	0.406	0.032
53.75	0.389	0.043	0.388	0.028
55	0.375	0.043	0.370	0.020

6.2.1. Representative agent model

Fig. 1 depicts the estimation results for the RA model. Panel 1(a) gives the minimized GMM objective function when we estimate $\rho(\gamma)$ separately for each value of γ; and Panel 1(b) gives the estimates of $\rho(\gamma)$ using the three methods described in Section 6.1. All three provide similar estimates.

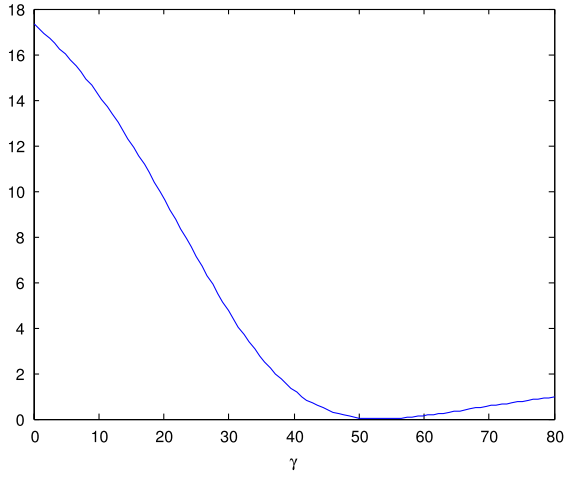
The minimized value of the GMM objective function in Panel 1(a) occurs at γ (=53.26) and ρ^{-1} (=0.395), as obtained by Kocherlakota and Pistaferri (2009) from the original moment conditions (35). Not surprisingly, the estimate of γ is very large because otherwise the RA SDF would not sufficiently co-vary with excess returns on stocks. In order to adjust for the large values of γ, the implied subjective discount factors are very low.

The relevant issue for our purposes, though, is underidentification. Can the population moment conditions be satisfied for a range of γ's? We consider an arguably narrow range using a discrete grid of 5 values around the minimizing value of γ from the plot in Panel 1(a). Specifically, we use grid points 50, 51.25, 52.5, 53.75, 55. Given that the covariance matrix of the moment conditions is close to being singular over that range mostly because of the large values of γ, we use ridge regularization. Panel 1(b) plots the corresponding estimated values of the subjective discount factor ρ^{-1}. The underidentification test is equal to 0.054, whose p-value using the Imhof approximation will be equal to 85.9%, so there is very little evidence against the null of underidentification. This statistic should be viewed with considerable caution when the concentrated GMM criterion depicted in Panel 1(a) is used to select the range of γ's. Conditioning on such a selection will alter the distribution of the I test. Alternatively, the I test could be used explicitly to infer a range, as we mentioned in Section 6.1. Formal analysis of these important issues is beyond the scope of this paper.

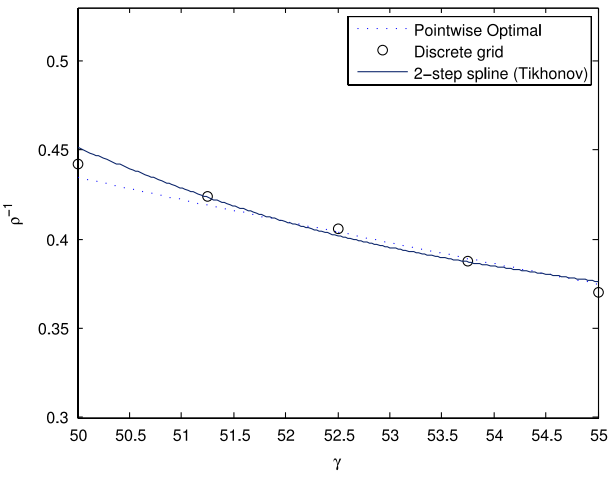
Table 1 presents point estimates of the subjective discount factor ρ^{-1} for the values of γ used in our discrete grid implementation, together with standard errors obtained from joint estimation and from separate estimation for each value of γ. The efficiency gains discussed in Proposition 2.1 are evident from a comparison of the standard errors.

We also implement the cubic spline approach described in Section 6.1. Given the smoothness of the curve obtained by estimating ρ^{-1} for each value of γ separately, we only consider three knots, namely 50, 52.5 and 55, and therefore two subintervals. Panel 1(b) presents the spline obtained with a two-step estimator with Tikhonov regularization of the second moment (uncentered) operator that uses the GMM estimator of ρ at the γ values at each knot point separately as first-stage estimator.[26] We obtain rather similar results with GMM estimators for a continuum of moment conditions that use ridge regularization applied to the uncentered covariance operator, as well as ridge and Tikhonov regularization applied to the centered covariance operator. See Appendix D for further details.

[26] For this approach the implied I test statistic is so close to zero that the resulting p-value is effectively one. Further work is required to dispel doubts about the finite sample reliability of these statistics.

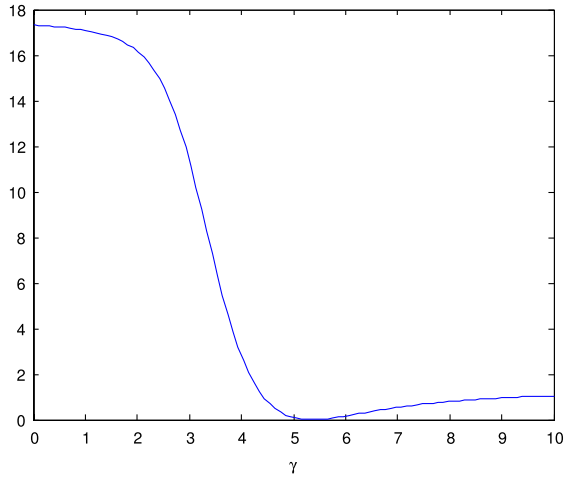


(a) Pointwise optimal GMM criterion function after minimizing with respect to ρ.

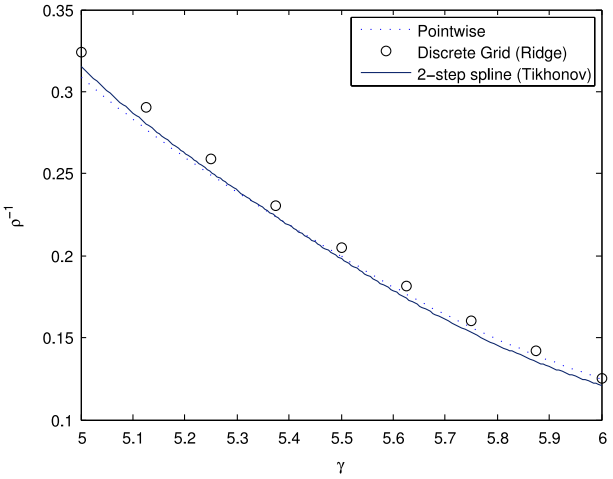


(b) Alternative estimators of the implied subjective discount factor for different values of γ.

Fig. 1. Results for the RA model.



(a) Pointwise optimal GMM criterion function for the PIPO model after minimizing with respect to ρ.



(b) Alternative estimators of the implied subjective discount factor for different values of γ.

Fig. 2. Results for the PIPO model.

6.2.2. Private information Pareto optimal model

Fig. 2 depicts the estimation results for the PIPO model. Panel 2(a) gives the minimized GMM objective function when we estimate $\rho(\gamma)$ separately for each value of γ; and Panel 2(b) gives the estimates of $\rho(\gamma)$ using the three methods described in Section 6.1. Again we find that the estimates are very similar for all three methods.

As argued by Kocherlakota and Pistaferri (2009), the PIPO SDF has the potential to fit the data with more reasonable values for γ by assigning an important role to consumption inequality and private information. Consistent with their evidence, the GMM objective function depicted in Panel 2(a) achieves a minimum at the more reasonable value of 5.33. The subjective discount factor, however, remains very low with an implied value of .232. To assess underidentification, once again we consider an equidistant discrete grid of 5 values of γ, but this time over the range 5 to 6. In this case, the underidentification test is equal to 0.536, whose Imhof-based p-value is equal to 61.9%, so once again there is little evidence against the null of underidentification.

Panel 2(b) also presents a three-knot spline estimated with the same approach that we used for the RA model. The criterion function-based underidentification test is 0.184, whose Imhof p-value is 0.96. Once again, further investigation is required to assess the reliability of these inferences.

7. Conclusions

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

In this paper we propose a method for constructing tests of underidentification based on the structural form of the equation system. We regard underidentification as a set of overidentifying

restrictions imposed on an augmented structural model. Therefore, our proposal is to test for underidentification by testing for overidentification in the augmented model using either standard overidentifying testing methods available in the literature, or some generalizations developed in this paper. To reveal the nature of the underidentification, we suggest estimating the direction or the curve that shows the parameter-tradeoffs that have comparable empirical implications.

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

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

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

1. we use the structural specification and exploit the fact that if β_0 is not identified, there will be a curve of $\beta's$ that will satisfy the original moment conditions;
2. we parameterize this curve, and write all the implied moment conditions as an extended system with either a finite or a continuum of moment conditions;
3. we obtain an estimator of the curve that is the counterpart to an efficient GMM estimator;
4. we compute the overidentification test of the extended system;

We do not provide an omnibus underidentification test. Instead our aim is to provide a test for underidentification and a measure of the nature of this underidentification in situations in which the characteristics of the identified set of interest are either theoretically or empirically motivated. We illustrated how to implement these methods with a consumption-based asset pricing model fit to microeconomic data.

Although we posed the target of the estimation to be a function π_0, we think of the object of interest to be the resulting curve. This approach to curve estimation requires that we know how to construct a valid parameterization of the curve *a priori*. It would be interesting to develop implementation methods that are insensitive to how we parameterize this curve. More generally, we could pose the estimation problem directly as one in which we infer a curve of maximal length. In addition, the most important aspects of our analysis should apply to manifolds of dimension higher than one.

In their study of observable factor models, Nagel and Singleton (2011) show that taking account of the conditioning information in an efficient way substantially alters the assessment of competing linear asset pricing models. Thus another important topic for future research is to incorporate conditional moment restrictions and to explore more generally the extent to which underidentification remains an important concern in practice.

Appendix A. The Cragg and Donald test of underidentification

Cragg and Donald (1993) considered single equation tests of underidentification based on the reduced form. Let us partition y_t into $(p+1)$- and r_1-dimensional vectors of endogenous and predetermined variables, respectively, $y_t = (y_{1t}', z_{1t}')'$, so that $k = p + r_1$ and $z_t = (z_{1t}', z_{2t}')'$, where z_{2t} is the vector of r_2 instruments excluded from the equation. Moreover, let Π and $\widehat{\Pi} = Y_1'Z(Z'Z)^{-1}$ be the $(p+1) \times r$ matrices of population and sample reduced form linear-projection coefficients, respectively. With this notation and the partition $\Pi = (\Pi_1, \Pi_2)$ conforming to that of z_t, α is identified up to scale if and only if the rank of Π_2 is p, but it is underidentified if the rank is $p-1$ or less.

To test for underidentification (Cragg and Donald, 1993) considered the minimizer of the minimum distance criterion

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

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

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

To relate (A.1) to our framework, write the augmented model

$$y_t'\alpha = u_t,$$
$$y_t'\alpha^* = v_t$$

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

$$By_{1t} + Cz_t = u_t^\dagger,$$

where $B = \left(B_1', B_2'\right)'$, $C = \left(C_1', C_2'\right)'$, $B_2 = (0_{p-1,2} \vdots I_{p-1})$, and $(B_1, C_{11}) = A'$, where $C_1 = (C_{11}, C_{12})$. To visualize the mapping between the structural parameters and the Cragg–Donald parameterization of the rank restriction, let us introduce the partitions $C_2 = (C_{21}, C_{22})$ and $B_1 = (B_{11}, B_{12})$. We then have that $\Pi_{22} = -C_{22}$ and $\Pi_{21} = B_{11}^{-1}B_{12}C_{22}$, so that $\Gamma_2 = -B_{11}^{-1}B_{12}$. Π_1 is unrestricted with $-B_{11}^{-1}(C_{11} - B_{12}C_{21})$ as the first component and $-C_{21}$ as the second.

Then noting that

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

so that

$$\mathrm{vec}(\widehat{\Pi} - \Pi) = (B \otimes Z'Z)^{-1}\sum_{i=1}^{T}(u_t^\dagger \otimes z_t),$$

(A.1) can be expressed as

$$\sum_{i=1}^{T}(u_t^\dagger \otimes z_t)'[(B \otimes Z'Z)V(B' \otimes Z'Z)]^{-1}\sum_{i=1}^{T}(u_t^\dagger \otimes z_t), \tag{A.2}$$

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

Specifically, using a standard result on the irrelevance of unrestricted moments (Arellano, 2003) (see pp. 196–197), criterion (A.2) concentrated with respect to C_2 can be shown to equal

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

An optimal weight matrix under classical errors is $V = Y_1' \times MY_1 \otimes \left(Z'Z\right)^{-1}$, where $M = I - Z\left(Z'Z\right)^{-1}Z'$, in which case the

concentrated criterion boils down to

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

Its minimizer subject to $A'Y'MYA = I$ coincides with the sum of the two smallest characteristic roots of $Y'Z\left(Z'Z\right)^{-1}Z'Y$ in the metric of $Y'MY$, which is one of the (non-robust) test statistics discussed by Cragg and Donald (1993).

Next, an optimal weight matrix under heteroskedastic errors is

$$V = \left(I \otimes Z'Z\right)^{-1}\sum_t (\widehat{\varepsilon}_t\widehat{\varepsilon}_t' \otimes z_t z_t')\left(I \otimes Z'Z\right)^{-1},$$

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

$$\left(\alpha' Y'Z, \alpha^{*\prime} Y'Z\right)\left(\sum_t A'\widetilde{y}_t\widetilde{y}_t'A \otimes z_t z_t'\right)^{-1}\begin{pmatrix} Z'Y\alpha \\ Z'Y\alpha^{*}\end{pmatrix},$$

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

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

Appendix B. Estimating finite-dimensional specifications of π

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

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

Let

$$\bar{f}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} f(x_t, \beta) = \begin{bmatrix} \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} f^{[1]}\left(x_t, \beta^{[1]}\right) \\ \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} f^{[2]}\left(x_t, \beta^{[1]}, \beta^{[2]}\right)\end{bmatrix}.$$

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

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

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

Lemma B.1.

$$\min_{\beta\in\mathbb{P}} \bar{f}_T(\beta)'[V_T(\beta)]^{-1}\bar{f}_T(\beta) \geq \min_{\beta\in\mathbb{P}} \bar{f}_T^{[1]}\left(\beta^{[1]}\right)' \times \left[V_T^{[11]}\left(\beta^{[1]}\right)\right]^{-1}\bar{f}_T^{[1]}\left(\beta^{[1]}\right).$$

Proof. Form

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

and construct similarly $\bar{f}_T^*(\beta, \gamma)$. The proof follows in three steps.

1.
$$\min_{\beta\in\mathbb{P}} \bar{f}_T(\beta)'[V_T(\beta)]^{-1}\bar{f}_T(\beta) \geq \min_{\beta\in\mathbb{P},\gamma} \bar{f}_T^*(\beta, \gamma)'[V_T(\beta)]^{-1}\bar{f}_T^*(\beta, \gamma).$$
The right-hand side minimization problem will not have a unique solution but this does not matter.

2.
$$\min_{\gamma} \bar{f}_T^*(\beta, \gamma)'[V_T(\beta)]^{-1}\bar{f}_T^*(\beta, \gamma) = \bar{f}_T^{[1]}(\beta^{[1]})'\left[V_T^{[11]}\left(\beta^{[1]}\right)\right]^{-1}\bar{f}_T^{[1]}(\beta^{[1]}). \quad \text{(B.3)}$$
This follows by using the first-order conditions for γ to show that
$$\bar{f}_T^{[2]}(\beta) - \gamma = \left(\begin{bmatrix}0 & I\end{bmatrix}[V_T(\beta)]^{-1}\begin{bmatrix}0 \\ I\end{bmatrix}\right)^{-1} \times \begin{bmatrix}0 & I\end{bmatrix}[V_T(\beta)]^{-1}\begin{bmatrix}I \\ 0\end{bmatrix}\bar{f}_T^{[1]}\left(\beta^{[1]}\right).$$
Substitute this outcome into the objective function on the left-hand side of (B.3) and apply the partitioned inverse formula to establish equality with the right-hand side of (B.3).

3. Finally,
$$\min_{\beta\in\mathbb{P},\gamma} \bar{f}_T^*(\beta, \gamma)'[V_T(\beta)]^{-1}\bar{f}_T^*(\beta, \gamma) = \min_{\beta\in\mathbb{P}}\min_{\gamma} \bar{f}_T^*(\beta, \gamma)'[V_T(\beta)]^{-1}\bar{f}_T^*(\beta, \gamma) = \min_{\beta\in\mathbb{P}} \bar{f}_T^{[1]}(\beta^{[1]})'\left[V_T^{[11]}\left(\beta^{[1]}\right)\right]^{-1}\bar{f}_T^{[1]}(\beta^{[1]}).$$

The conclusion follows from these three steps. □

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

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

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

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

As a result, if we use continuously-updated GMM and allow for explorations across alternative degrees of underidentification, then the objective will lead us to the smallest allowable degree of underidentification. In particular, if we allow for the estimation of nonlinear curves such as (34) in a model that is fundamentally linear, then the continuously-updated GMM objective will lead us to represent the underidentification by means of a line or at least the segment of a line.

Appendix C. Imhof-based approximation to the distribution of GMM tests

Let

$$\bar{f}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T} f(x_t, \beta),$$

and define

$$M = \lim_{T\to\infty} \mathrm{Var}\left[\sqrt{T}\bar{f}_T(\beta_o)\right].$$

Since the purpose of this appendix is to explain the application of Imhof (1961) results in our context, initially we will abstract from estimation issues by assuming that β_0 is known.

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

$$T\bar{f}_T(\beta_0)M^{-1}\bar{f}_T(\beta_0)$$

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

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

$$\begin{aligned}
&T\bar{f}_T'(\beta_0)M^{1/2}(\varsigma_T I_p + M^2)^{-1}M^{1/2}\bar{f}_T(\beta_0)\\
&\quad= \sqrt{T}\bar{f}_T(\beta_0)'W\Delta^{1/2}W'(\varsigma_T I_p + W\Delta^2 W')^{-1}\\
&\qquad\times W\Delta^{1/2}W'\sqrt{T}\bar{f}_T(\beta_0)\\
&\quad= \left[\sqrt{T}\bar{f}_T(\beta_0)'W\Delta^{-1/2}\right]\left[(\varsigma_T I_p + \Delta^2)^{-1}\Delta^2\right]\\
&\qquad\times\left[\Delta^{-1/2}W'\sqrt{T}\bar{f}_T(\beta_0)\right]\\
&\quad= \sum_{j=1}^{p}\frac{\delta_j^2}{\delta_j^2+\varsigma_T}\left[\sqrt{T}\varepsilon_{j,T}\right]^2,
\end{aligned}$$

where $W\Delta W'$ provides the spectral decomposition of M, $\varepsilon_{j,T}$ is the jth entry of the random vector $\varepsilon_T = \Delta^{-1/2}W'\bar{f}_T(\beta_0)$ and ς_T is a regularization parameter. Since $\sqrt{T}\varepsilon_T \to N(0, I_p)$, we will recover the chi-square limiting distribution under the null if we let ς_T go to 0 at a suitable rate. But given that for a fixed ς_T the above statistic will converge to a diagonal quadratic form in standard normal random variables as $T \to \infty$, we can use Koerts and Abrahamse (1969) implementation of the Imhof (1961) procedure for evaluating the probability that a quadratic form of normals is less than a given value (see also Farebrother, 1990). Although the smallest eigenvalue of M, $\delta_{\min}$ say, will generally be strictly positive, from a numerical point of view it makes sense to truncate the previous expression so that we only use those terms for which

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

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

In practice, β_0 will be replaced by its "optimal" GMM estimator, which in the case of Tikhonov regularization will approximately satisfy the first order conditions

$$D'W\Delta^{1/2}(\varsigma_T I_p + \Delta^2)^{-1}\Delta^{1/2}W'\bar{f}_T(\hat{\beta}) = G'E'\bar{f}_T(\hat{\beta}) = 0$$

in large samples, where D is the expected Jacobian matrix of the moment conditions, $E = W\Delta^{1/2}(\varsigma_T I_p + \Delta^2)^{-1/2}$ and $G = E'D$. Standard arguments then imply that

$$\sqrt{T}(\hat{\beta} - \beta_0) = -(G'G)^{-1}G'\sqrt{T}\bar{f}_T(\beta_0) + o_p(1).$$

Linearizing $\bar{f}_T(\hat{\beta})$ around β_0 allows us to write

$$\sqrt{T}E'\bar{f}_T(\hat{\beta}) = [I_p - G(G'G)^{-1}G']\sqrt{T}E'\bar{f}_T(\beta_0) + o_p(1),$$

where $[I_p - G(G'G)^{-1}G']$ is an idempotent matrix of rank $p-k$. As a result, the overidentification restriction test will be equal to

$$\begin{aligned}
&T\bar{f}_T'(\hat{\beta})M^{1/2}(\varsigma_T I_p + M^2)^{-1}M^{1/2}\bar{f}_T(\hat{\beta})\\
&\quad= \sqrt{T}\bar{f}_T'(\hat{\beta})EE'\sqrt{T}\bar{f}_T(\hat{\beta})\\
&\quad= \sqrt{T}\bar{f}_T'(\beta_0)E[I_p - G(G'G)^{-1}G']\sqrt{T}E'\bar{f}_T(\beta_0) + o_p(1)\\
&\quad= \sqrt{T}\bar{f}_T'(\beta_0)\Delta^{-1/2}W'\Big\{\Delta(\varsigma_T I_p + \Delta^2)^{-1/2}[I_p - G(G'G)^{-1}\\
&\qquad\times G'](\varsigma_T I_p + \Delta^2)^{-1/2}\Delta\Big\}\\
&\qquad\times\sqrt{T}\Delta^{-1/2}W'\bar{f}_T(\beta_0) + o_p(1),
\end{aligned}$$

whose finite sample distribution can also be approximated by a more complex quadratic form in standard normal random variables. As expected, this distribution will converge to the usual χ^2 with $p-k$ degrees of freedom when ς_T goes to 0 at a suitable rate.

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

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

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

where $C_{\cdot s}$ is the sth column of C and ι_T is a vector of T 1's. Consider the spectral decomposition $C = U\Lambda U'$. Then, it is possible to show that the continuum of moment conditions test studied by Carrasco and Florens (2000) is numerically identical to the following expression

$$\begin{aligned}
v'\left[\varsigma_T I_T + C^2\right]^{-1} v &= \iota_T' C\left[\varsigma_T I_T + C^2\right]^{-1} C\iota_T\\
&= \iota_T' U\begin{bmatrix}\frac{\lambda_1^2}{\varsigma_T+\lambda_1^2} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{\lambda_T^2}{\varsigma_T+\lambda_T^2}\end{bmatrix} U\iota_T.
\end{aligned}$$

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

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

where

$$p_T(\varsigma_T) = \sum_{j=1}^{T}\frac{\lambda_j^2}{\lambda_j^2+\varsigma_T}$$

$$q_T(\varsigma_T) = 2\sum_{j=1}^{T}\frac{\lambda_j^4}{\left(\lambda_j^2+\varsigma_T\right)^2}.$$

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

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

Appendix D. GMM estimators with a continuum of moments

D.1. Covariance operators

D.1.1. Uncentered covariance operator

First of all, it is worth noting that (Carrasco and Florens, 2000) only consider models with a single moment condition indexed by a parameter vector. Therefore, in order to use their expressions with two or more moment conditions one would need to use two indices: a discrete index that tracks the moments, and another continuous one that refers to the true index parameter γ (see Carrasco et al., 2007). In contrast, here we use a vector of moments indexed with respect to a scalar parameter, which should lead to identical results under the assumption that all moments are weighted equally in computing inner products. On this basis, we can define the sample second moment (matrix) integral operator $\bar{K}_T^u(\varrho)$ associated with the kernel

$$\bar{\kappa}_T^u(\gamma', \gamma''; \varrho) = \frac{1}{T}\sum_{t=1}^{T} f(x_t, \gamma'; \varrho) f'(x_t, \gamma''; \varrho)$$

to be such that

$$\begin{aligned}\bar{K}_T^u(\varrho)\varphi(\gamma') &= \int_{\gamma_1}^{\gamma_{H+1}} \bar{\kappa}_T^u(\gamma', \gamma''; \varrho)\varphi(\gamma'')d\Psi(\gamma'') \\ &= \frac{1}{T}\sum_{t=1}^{T} f(x_t, \gamma'; \varrho) \\ &\quad \times \int_{\gamma_1}^{\gamma_{H+1}} f'(x_t, \gamma''; \varrho)\varphi(\gamma'')d\Psi(\gamma''),\end{aligned}$$

where $\varphi(\gamma)$ is a 2×1 function of γ, $\Psi(\gamma)$ is typically the cdf of some random variable defined over (γ_1, γ_{H+1}), which for simplicity we take to be uniform hereinafter, and ϱ is a vector of knots in the cubic splines as explained in Appendix E.

In order to make this definition operational with the moment conditions (35), define $C^u(\dot{\varrho}, \ddot{\varrho})$ as the $T \times T$ matrix that appears in the objective function of the GMM estimators described in Appendix D.2, whose (t, s) element is given by

$$\begin{aligned}c_{ts}^u(\dot{\varrho}, \ddot{\varrho}) &= \frac{1}{T}\int_{\gamma_1}^{\gamma_{H+1}} f'(x_t, \gamma; \dot{\varrho}) f(x_s, \gamma; \ddot{\varrho})d\gamma \\ &= c_{ts}^h - [z_t(\dot{\varrho}) + z_s(\ddot{\varrho})] + d(\dot{\varrho}, \ddot{\varrho}),\end{aligned}$$

where

$$c_{ts}^h = \frac{1}{T}\int_{\gamma_1}^{\gamma_{H+1}} (R_{m,t}R_{s,t} + R_{f,t}R_{f,s})k(x_t, \gamma)k(x_s, \gamma)d\gamma,$$

$$z_t(\ddot{\varrho}) = \frac{1}{T}\int_{\gamma_1}^{\gamma_{H+1}} \rho(\gamma; \ddot{\varrho})(R_{m,t} + R_{f,t})k(x_t, \gamma)d\gamma$$

and

$$d(\dot{\varrho}, \ddot{\varrho}) = \frac{2}{T}\int_{\gamma_1}^{\gamma_{H+1}} \rho(\gamma; \dot{\varrho})\rho(\gamma; \ddot{\varrho})d\gamma.$$

Therefore, in matrix notation we can write

$$C^u(\dot{\varrho}, \ddot{\varrho}) = C^h - [z(\dot{\varrho})\iota_T' + \iota_T z'(\ddot{\varrho})] + d(\dot{\varrho}, \ddot{\varrho})\iota_T\iota_T',$$

with $z(\dot{\varrho}) = [z_1(\dot{\varrho}), \ldots, z_T(\dot{\varrho})]'$.

The moment conditions (35) are highly non-linear in γ. However, one result that we will use repeatedly is that although the cubic spline $\rho(\gamma; \varrho)$ is a non-linear function of γ for a given ϱ, it is possible to write it as a linear function of ϱ for a given γ (see Appendix E). Consequently, $f(x_t, \gamma; \varrho)$ will be affine in ϱ for a specific value of γ, and the same is true of $\bar{f}_T(\gamma; \varrho)$ *a fortiori*. More importantly, we can also prove that integrals of $f(x_t, \gamma; \varrho)$ are also linear in ϱ. As a result, we can write $z(\varrho) = \Upsilon\varrho$ and $d(\dot{\varrho}, \ddot{\varrho}) = \dot{\varrho}'\Omega\ddot{\varrho}$, where the computation of the required coefficient matrices can also be found in Appendix E.

Let $\omega_j^u(\gamma; \varrho)$ denote the T orthogonal eigenfunctions of the operator $\bar{K}_T^u(\varrho)$ associated to its T non-zero eigenvalues $\lambda_j(\varrho)$, which can be obtained from the conditions

$$\bar{K}_T^u(\varrho)\omega_j^u(\gamma'; \varrho) = \lambda_j^u(\varrho)\omega_j^u(\gamma'; \varrho).$$

For our purposes it is more convenient to work with the normalized eigenfunctions

$$\omega_j^{u+}(\gamma; \varrho) = \frac{1}{T}\sum_{t=1}^{T} \phi_{jt}^{u+}(\varrho) f(x_t, \gamma; \varrho),$$

which are such that $\left\langle \omega_j^{u+}(\gamma; \varrho), \omega_j^{u+}(\gamma; \varrho)\right\rangle = 1$. Following Carrasco et al. (2007), we can show that if we denote the spectral decomposition of the matrix $C^u(\varrho, \varrho)$ by

$$C^u(\varrho, \varrho) = U^u(\varrho)\Lambda^u(\varrho)U^{u\prime}(\varrho),$$

then we will have that

$$\phi_j^{u+}(\varrho) = \frac{\sqrt{T}}{\sqrt{\lambda_j^u(\varrho)}}u_j^u(\varrho),$$

where

$$U^u(\varrho) = [u_1^u(\varrho), \ldots, u_j^u(\varrho), \ldots, u_T^u(\varrho)].$$

D.1.2. Centered covariance operator

We can alternatively define the kernel of the integral operator $\bar{K}_T^c(\varrho)$ as

$$\begin{aligned}\bar{\kappa}_T^c(\gamma', \gamma''; \varrho) = \frac{1}{T}\sum_{t=1}^{T} &[f(x_t, \gamma'; \varrho) - \bar{f}_T(\gamma'; \varrho)] \\ &\times [f(x_t, \gamma''; \varrho) - \bar{f}_T(\gamma''; \varrho)]'.\end{aligned}$$

(see Carrasco and Kotchoni, 2010, for an application of this operator in characteristic function-based estimation). Let $C^c(\dot{\varrho}, \ddot{\varrho})$ be a $T \times T$ matrix whose (t, s) element is

$$\begin{aligned}c_{ts}^c(\dot{\varrho}, \ddot{\varrho}) = \frac{1}{T}\int_{\gamma_1}^{\gamma_{H+1}} &[f(x_t, \gamma; \varrho) - \bar{f}_T(\gamma; \dot{\varrho})]' \\ &\times [f(x_t, \gamma; \ddot{\varrho}) - \bar{f}_T(\gamma; \ddot{\varrho})]d\gamma.\end{aligned}$$

It immediately follows from this definition that

$$\begin{aligned}C^c(\dot{\varrho}, \ddot{\varrho}) &= (I_T - T^{-1}\iota_T\iota_T')C^u(\dot{\varrho}, \ddot{\varrho})(I_T - T^{-1}\iota_T\iota_T') \\ &= (I_T - T^{-1}\iota_T\iota_T')C^h(I_T - T^{-1}\iota_T\iota_T') = C^c,\end{aligned}$$

which does not depend on $\dot{\varrho}$ or $\ddot{\varrho}$. Another worthwhile feature of C^c is that it has rank $T - 1$ at most, with ι_T being the eigenvector associated to the zero eigenvalue.

Let the spectral decomposition of C^c be

$$C^c = U^c\Lambda^c U^{\prime c}.$$

Then, the weights of the orthonormalised eigenfunctions will be given by

$$\phi_j^{c+} = \frac{\sqrt{T}}{\sqrt{\lambda_j^c}}u_j^c.$$

D.2. Alternative estimators

D.2.1. Uncentered two-step estimators

Let $\bar{\varrho}_T$ denote some preliminary consistent estimator of ϱ. In this context, the analog to the two-step efficient GMM criterion function that uses Tikhonov regularization will be

$$Q^{2S}(\varrho;\bar{\varrho}_T,u,T)=\sum_{j=1}^{T}\frac{\lambda_j^u(\bar{\varrho}_T)}{[\lambda_j^u(\bar{\varrho}_T)]^2+\varsigma_T}\times\left\langle\omega_j^{u+}(\gamma;\bar{\varrho}_T),\bar{f}_T(\gamma;\varrho)\right\rangle^2.$$

Following Carrasco et al. (2007), we can write this expression in matrix notation as

$$\iota_T'C^{u\prime}(\bar{\varrho}_T,\varrho)[\varsigma_T I_T+(C^u(\bar{\varrho}_T,\bar{\varrho}_T))^2]^{-1}C^u(\bar{\varrho}_T,\varrho)\iota_T.$$

But since in our case $C^u(\bar{\varrho}_T,\varrho)$ is affine in ϱ, we can write

$$\begin{aligned}C^u(\bar{\varrho}_T,\varrho)\iota_T&=(C^h\iota_T-T\Upsilon\bar{\varrho}_T)+\iota_T(\iota_T'\Upsilon\varrho+T\bar{\varrho}_T\Omega\varrho)\\&=\pi(\bar{\varrho}_T)+\Pi(\bar{\varrho}_T)\varrho,\end{aligned}$$

so that the criterion function is a quadratic form in ϱ. As a result, we can obtain its optimum in closed form as

$$\begin{aligned}\hat{\varrho}_T^{2S}=&\{\Pi'(\bar{\varrho}_T)U^u(\bar{\varrho}_T)[\varsigma_T I_T+\Lambda^{u2}(\bar{\varrho}_T)]^{-1}U^{u\prime}(\bar{\varrho}_T)\Pi(\bar{\varrho}_T)^{-1}\\&\times\Pi'(\bar{\varrho}_T)U^u(\bar{\varrho}_T)[\varsigma_T I_T+\Lambda^{u2}(\bar{\varrho}_T)]^{-1}U^{u\prime}(\bar{\varrho}_T)\pi(\bar{\varrho}_T).\end{aligned}$$

Iterated estimators can be easily obtained by computing this expression recursively.

An alternative regularization scheme would use the criterion function

$$Q^{2S}(\varrho;\bar{\varrho}_T,u,r)=\sum_{j=1}^{T}\frac{1}{\lambda_j^u(\bar{\varrho}_T)+\varsigma_T}\left\langle\omega_j^{u+}(\gamma;\bar{\varrho}_T),\bar{f}_T(\gamma;\varrho)\right\rangle^2.$$

Given that

$$\begin{aligned}\left\langle\omega_j^{u+}(\gamma;\bar{\varrho}_T),\bar{f}_T(\gamma;\varrho)\right\rangle&=\frac{1}{T^2}e_j'\Phi^{+u\prime}(\bar{\varrho}_T)C^u(\bar{\varrho}_T,\varrho)\iota_T\\&=\frac{1}{T^2}e_j'B^{+u\prime}(\bar{\varrho}_T)[\pi(\bar{\varrho}_T)+\Pi(\bar{\varrho}_T)\varrho],\end{aligned}$$

where e_j' is the jth vector of the canonical basis, $\left\langle\omega_j^{u+}(\gamma;\bar{\varrho}_T),\bar{g}_T(\gamma;\varrho)\right\rangle^2$ will be quadratic in ϱ, which means that $\arg\min_\varrho Q^{2S}(\varrho;\bar{\varrho}_T,u,r)$ should also have a closed form. Nevertheless, since $\Phi^{+u}(\bar{\varrho}_T)=\sqrt{T}U^u(\bar{\varrho}_T)[\Lambda^u(\bar{\varrho}_T)]^{-1/2}$, the computation of $Q^{2S}(\varrho;\bar{\varrho}_T,u,r)$ will be problematic unless we also truncate the eigenfunctions.

D.2.2. Uncentered continuously updated estimators

A continuously updated version of the Tikhonov regularized criterion function will be

$$Q^{CU}(\varrho;u,T)=\sum_{j=1}^{T}\frac{\lambda_j(\varrho)}{\lambda_j^2(\varrho)+\varsigma_T}\left\langle\omega_j(\gamma;\varrho),\bar{f}_T(\gamma;\varrho)\right\rangle^2,$$

which in matrix notation becomes:

$$\iota_T'C^{u\prime}(\varrho,\varrho)[\varsigma_T I_T+(C^u(\varrho,\varrho))^2]^{-1}C^u(\varrho,\varrho)\iota_T.$$

Unfortunately, in this case there does not seem to be a simple closed expression for the optimal estimator.

For analogous reasons, the continuously updated version of the ridge regularization will not lead to closed-form expressions either, even though its computation should not be problematic since we could always do it as

$$Q^{CU}(\varrho;u,r)=\sum_{j=1}^{T}\frac{\lambda_j(\varrho)}{\lambda_j(\varrho)+\varsigma_T}\left\{\sum_{t=1}^{T}u_{jt}(\varrho)\right\}^2.$$

D.2.3. Centered estimators

Given that neither eigenvalues nor eigenfunctions require a preliminary estimator, in this case the efficient GMM criterion function that uses Tikhonov regularization will be given by

$$\begin{aligned}&Q(\varrho;c,T)\\&=\sum_{j=1}^{T-1}\frac{\lambda_j^c}{\lambda_j^{c2}+\varsigma_T}\left\langle\omega_j^{c+}(\gamma),\bar{f}_T(\gamma;\varrho)\right\rangle^2\\&=\sum_{j=1}^{T-1}\frac{1}{\lambda_j^{c2}+\varsigma_T}\left\{\int_{\gamma_1}^{\gamma_{H+1}}\left[\frac{\sqrt{T}}{T}\sum_{t=1}^{T}u_{jt}^c[R_tk(x_t,\gamma)\right.\right.\\&\quad\left.\left.\times-\overline{R_tk(x_t,\gamma)}]'\right]\bar{f}_T(\gamma;\varrho)d\gamma\right\}^2,\end{aligned}$$

where $\overline{R_tk(x_t,\gamma)}$ denotes the sample average of $R_tk(x_t,\gamma)$. Note that we have only included $T-1$ terms because the centred covariance operator has a zero eigenvalue. Straightforward algebra then implies that

$$\begin{aligned}&\int_{\gamma_1}^{\gamma_{H+1}}[R_tk(x_t,\gamma)-\overline{R_tk(x_t,\gamma)}]'\bar{f}_T(\gamma;\varrho)d\gamma\\&=\frac{1}{T}e_t'C^h\iota_T-\frac{1}{T^2}\iota_T'C^h\iota_T-e_t'z(\varrho)+\frac{1}{T}z'(\varrho)\iota_T,\end{aligned}$$

where e_t is the tth vector of the canonical basis. Hence,

$$\begin{aligned}&\frac{\sqrt{T}}{T}\sum_{t=1}^{T}u_{jt}^c\\&\quad\times\left[\frac{1}{T}\sum_{s=1}^{T}\int_{\gamma_1}^{\gamma_{H+1}}[R_tk(x_t,\gamma)-\overline{R_tk(x_t,\gamma)}]'f(x_s,\gamma;\varrho)d\gamma\right]\\&=\frac{\sqrt{T}}{T}u_j^{c\prime}\left[\frac{1}{T}\left(I_T-\frac{1}{T}\iota_T\iota_T'\right)C^h\iota_T-\left(I_T-\frac{1}{T}\iota_T\iota_T'\right)z(\varrho)\right].\end{aligned}$$

But we have seen before that ι_T is the eigenvector of C^c associated to its 0 eigenvalue, which means that $u_j^{c\prime}\iota_T=0$ for $j=1,\ldots,T-1$. As a result, the criterion function will be

$$\begin{aligned}&\left[\frac{\sqrt{T}}{T^2}\bar{U}^{c\prime}C^h\iota_T-\frac{\sqrt{T}}{T}\bar{U}^{c\prime}\Upsilon\varrho\right]'[\varsigma I_{T-1}+(\bar{\Lambda}^c)^2]^{-1}\\&\times\left[\frac{\sqrt{T}}{T^2}\bar{U}^{c\prime}C^h\iota_T-\frac{\sqrt{T}}{T}\bar{U}^{c\prime}\Upsilon\varrho\right],\end{aligned}$$

where the upper bar on $\bar{U}^c$ and $\bar{\Lambda}^c$ indicates that we have eliminated the elements associated to the 0 eigenvalue. As expected, this expression is quadratic in ϱ, so once again it is possible to find a closed-form analytical expression for the estimator.

For analogous reasons, we are also able to find closed-form expressions if we use ridge regularization instead, as long as we complement it with truncation of the eigenfunctions.

Appendix E. Computational aspects of natural cubic splines

If $\rho(\gamma;\varrho)$ is a (natural) cubic spline function defined over the interval $[\gamma_1,\gamma_{H+1}]$ that depends on $H+1$ parameters $\varrho=(\varrho_1,\ldots,\varrho_{H+1})'$, which are the knot values associated to the $H+1$ distinct knots $\gamma_1,\ldots,\gamma_{H+1}$, then

$$\rho(\gamma;\varrho)=\begin{cases}\rho_1(\gamma;\varrho) & \gamma\in[\gamma_1,\gamma_2]\\ \rho_2(\gamma;\varrho) & \gamma\in[\gamma_2,\gamma_3]\\ \vdots & \vdots\\ \rho_H(\gamma;\varrho) & \gamma\in[\gamma_H,\gamma_{H+1}]\end{cases}$$

with $\rho_i(\gamma;\varrho)$ $(i=1,\ldots,H)$ being the unique cubic polynomials in γ that satisfy the following conditions:

$$
\begin{pmatrix}
2(\gamma_3-\gamma_1) & \gamma_3-\gamma_2 & 0 & \cdots & & & & 0 \\
\gamma_3-\gamma_2 & 2(\gamma_4-\gamma_2) & \gamma_4-\gamma_3 & 0 & & & & \\
0 & \gamma_4-\gamma_3 & 2(\gamma_5-\gamma_3) & \gamma_5-\gamma_4 & 0 & & & \\
\vdots & \ddots & \ddots & \ddots & & & \ddots & \vdots \\
 & & & \ddots & \ddots & & \ddots & \ddots \\
 & & & \ddots & \gamma_{H-1}-\gamma_{H-2} & 2(\gamma_H-\gamma_{H-2}) & \gamma_H-\gamma_{H-1} & \\
0 & \cdots & & & 0 & \gamma_H-\gamma_{H-1} & 2(\gamma_{H+1}-\gamma_{H-1}) &
\end{pmatrix}
$$
$$
\times \begin{pmatrix} v_2(\varrho) \\ v_3(\varrho) \\ v_4(\varrho) \\ \vdots \\ v_{H-2}(\varrho) \\ v_{H-1}(\varrho) \\ v_H(\varrho) \end{pmatrix}
= 6 \begin{pmatrix} (\varrho_3-\varrho_3)/(\gamma_3-\gamma_2)-(\varrho_2-\varrho_1)/(\gamma_2-\gamma_1) \\ \vdots \\ (\varrho_{H+1}-\varrho_H)/(\gamma_{H+1}-\gamma_H)-(\varrho_H-\varrho_{H-1})/(\gamma_H-\gamma_{H-1}) \end{pmatrix}
$$

Box I.

1. $\rho_i(\gamma_{i+1};\varrho)=\varrho_{i+1}$ (interpolating property).
2. $\rho_{i-1}(\gamma_i;\varrho)=\rho_i(\gamma_i;\varrho)$ $(i=2,\ldots,H)$ (continuity).
3. $\partial\rho_{i-1}(\gamma_i;\varrho)/\partial\gamma=\partial\rho_i(\gamma_i;\varrho)/\partial\gamma$ $(i=2,\ldots,H)$ (continuity of the first derivative).
4. $\partial^2\rho_{i-1}(\gamma_i;\varrho)/(\partial\gamma)^2=\partial^2\rho_i(\gamma_i;\varrho)/(\partial\gamma)^2$ $(i=2,\ldots,H)$ (continuity of the second derivative).
5. $\partial^2\rho_1(\gamma_1;\varrho)/(\partial\gamma)^2=0$ and $\partial^2\rho_{H+1}(\gamma_{H+1};\varrho)/(\partial\gamma)^2=0$ ("natural" cubic spline).

The specific form of these cubic polynomials is

$$
\begin{aligned}
\rho_i(\gamma;\varrho) = {} & \frac{v_{i+1}(\varrho)(\gamma-\gamma_i)^3+v_i(\varrho)(\gamma_{i+1}-\gamma)^3}{6(\gamma_{i+1}-\gamma_i)} \\
& +\left[\frac{\varrho_{i+1}}{\gamma_{i+1}-\gamma_i}-\frac{\gamma_{i+1}-\gamma_i}{6}v_{i+1}(\varrho)\right](\gamma-\gamma_i) \\
& +\left[\frac{\varrho_i}{\gamma_{i+1}-\gamma_i}-\frac{\gamma_{i+1}-\gamma_i}{6}v_i(\varrho)\right](\gamma_{i+1}-\gamma),
\end{aligned}
$$

where the coefficients $v_i(\varrho)$ can be found by solving the linear system of $H-1$ equations:

$$
\begin{aligned}
&(\gamma_i-\gamma_{i-1})v_{i-1}(\varrho)+2(\gamma_{i+1}-\gamma_{i-1})v_i(\varrho)+(\gamma_{i+1}-\gamma_i)v_{i+1}(\varrho) \\
&\quad=6\left(\frac{\varrho_{i+1}-\varrho_i}{\gamma_{i+1}-\gamma_i}-\frac{\varrho_i-\varrho_{i-1}}{\gamma_i-\gamma_{i-1}}\right)
\end{aligned}
$$

for $i=2,\ldots,H$ in the $H-1$ unknowns $v_2(\varrho),\ldots,v_H(\varrho)$, with $v_1(\varrho)=0$ and $v_{H+1}(\varrho)=0$.

In matrix notation, we can write this system as given in Box I, which can be solved very efficiently because of its symmetric tridiagonal nature.

Importantly, in matrix notation, we can write this system as given in Box II so the independent term is linear in ϱ. This, coupled with the fact that the coefficient matrix is a function of the knots but not of the knot values, means that

$$v(\varrho)=A(\gamma)\varrho,$$

where $v(\varrho)=[v_1(\varrho),v_2(\varrho),\ldots,v_{H+1}(\varrho)]$ and $\gamma=(\gamma_1,\gamma_2,\ldots,\gamma_{H+1})$. As a result, $\rho_i(\gamma;\varrho)$ will be linear in ϱ too, and the same applies to $\rho(\gamma;\varrho)$, as long as we interpret this to mean that the coefficients of the linear combination will depend on γ and the interval to which it belongs.

Let us now consider the integrals required in the different inner product calculations for the case of the RA consumption CAPM model. Specifically, we need to compute

$$
\begin{aligned}
T\cdot c_{ts}^h &= (R_{m,t}R_{s,t}+R_{f,t}R_{f,s})\int_{\gamma_1}^{\gamma_{H+1}}k(x_t,\gamma)k(x_s,\gamma)d\gamma \\
&= (R_{m,t}R_{s,t}+R_{f,t}R_{f,s}) \\
&\quad\times\int_{\gamma_1}^{\gamma_{H+1}}\left[\left(\frac{\sum_{i=1}^N c_{it}}{\sum_{i=1}^N c_{it-3}}\right)\left(\frac{\sum_{i=1}^N c_{is}}{\sum_{i=1}^N c_{is-3}}\right)\right]^{-\gamma}d\gamma \\
&= (R_{m,t}R_{s,t}+R_{f,t}R_{f,s}) \\
&\quad\times\left[\ln\left(\frac{\sum_{i=1}^N c_{it}}{\sum_{i=1}^N c_{it-3}}\right)+\ln\left(\frac{\sum_{i=1}^N c_{is}}{\sum_{i=1}^N c_{is-3}}\right)\right]^{-1} \\
&\quad\times\left\{\left[\left(\frac{\sum_{i=1}^T c_{it}}{\sum_{i=1}^N c_{it-3}}\right)\left(\frac{\sum_{i=1}^N c_{is}}{\sum_{i=1}^N c_{is-3}}\right)\right]^{-\gamma_1}\right. \\
&\quad\left.-\left[\left(\frac{\sum_{i=1}^N c_{it}}{\sum_{i=1}^N c_{it-3}}\right)\left(\frac{\sum_{i=1}^N c_{is}}{\sum_{i=1}^N c_{is-3}}\right)\right]^{-\gamma_{H+1}}\right\},
\end{aligned}
$$

where we have used expression (F.4) in Appendix F.

In addition, we need to compute

$$
\begin{aligned}
T\cdot z_t(\varrho) &= (R_{m,t}+R_{f,t}) \\
&\quad\times\sum_{i=1}^H\left[\int_{\gamma_i}^{\gamma_{i+1}}\rho_i(\gamma;\varrho)\left(\frac{\sum_{i=1}^N c_{it}}{\sum_{i=1}^N c_{it-3}}\right)^{-\gamma}d\gamma\right].
\end{aligned}
$$

$$\begin{pmatrix} (\varrho_3-\varrho_2)/(\gamma_3-\gamma_2)-(\varrho_2-\varrho_1)/(\gamma_2-\gamma_1) \\ \vdots \\ (\varrho_{H+1}-\varrho_H)/(\gamma_{H+1}-\gamma_H)-(\varrho_H-\varrho_{H-1})/(\gamma_H-\gamma_{H-1}) \end{pmatrix}$$

$$= \begin{pmatrix} (\gamma_2-\gamma_1)^{-1} & -(\gamma_3-\gamma_2)^{-1}-(\gamma_2-\gamma_1)^{-1} & (\gamma_3-\gamma_2)^{-1} & 0 & \cdots & 0 \\ 0 & (\gamma_3-\gamma_2)^{-1} & -(\gamma_4-\gamma_3)^{-1}-(\gamma_3-\gamma_2)^{-1} & (\gamma_4-\gamma_3)^{-1} & 0 & \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots & \ddots \\ 0 & \cdots & & (\gamma_H-\gamma_{H-1})^{-1} & -(\gamma_{H+1}-\gamma_H)^{-1}-(\gamma_H-\gamma_{H-1})^{-1} & (\gamma_{H+1}-\gamma_H)^{-1} \end{pmatrix}$$

$$\times \begin{pmatrix} \varrho_1 \\ \varrho_2 \\ \varrho_3 \\ \vdots \\ \varrho_{H-1} \\ \varrho_H \\ \varrho_{H+1} \end{pmatrix}$$

Box II.

But

$$\int_{\gamma_i}^{\gamma_{i+1}} \rho_i(\gamma;\varrho)\left(\frac{\sum_{i=1}^{N} c_{it}}{\sum_{i=1}^{N} c_{it-3}}\right)^{-\gamma} d\gamma$$

$$= \int_{\gamma_i}^{\gamma_{i+1}} \left\{\left(\frac{\sum_{i=1}^{N} c_{it}}{\sum_{i=1}^{N} c_{it-3}}\right)^{-\gamma} \times \left[\frac{v_{i+1}(\varrho)(\gamma-\gamma_i)^3 + v_i(\varrho)(\gamma_{i+1}-\gamma)^3}{6(\gamma_{i+1}-\gamma_i)} + \left(\frac{\varrho_{i+1}}{\gamma_{i+1}-\gamma_i} - \frac{\gamma_{i+1}-\gamma_i}{6} v_{i+1}(\varrho)\right)(\gamma-\gamma_i) + \left(\frac{\varrho_i}{\gamma_{i+1}-\gamma_i} - \frac{\gamma_{i+1}-\gamma_i}{6} v_i(\varrho)\right)(\gamma_{i+1}-\gamma)\right]\right\} d\gamma,$$

which can also be easily obtained by using the expressions in Appendix F. Importantly, given that the resulting expression is linear in ϱ for each i, $z_t(\varrho)$ will be linear in ϱ too. In this sense, it is convenient to replace the $v_i(\varrho)$'s by $A(\gamma)\varrho$, so that we can obtain the matrix Υ. Although we have to resort to numerical quadrature to compute the integrals of the PIPO model, the linearity of $z_t(\varrho)$ in ϱ is preserved.

Finally, we need to compute

$$T \cdot d(\dot{\varrho}, \ddot{\varrho}) = 2\int_{\gamma_1}^{\gamma_{H+1}} \rho(\gamma;\dot{\varrho})\rho(\gamma;\ddot{\varrho})d\gamma = 2\sum_{i=1}^{H}\left[\int_{\gamma_i}^{\gamma_{i+1}} \rho_i(\gamma;\dot{\varrho})\rho_i(\gamma;\ddot{\varrho})d\gamma\right].$$

Once again, we can use the expressions in Appendix F to compute

$$\left[\int_{\gamma_i}^{\gamma_{i+1}} \rho_i(\gamma;\dot{\varrho})\rho_i(\gamma;\ddot{\varrho})d\gamma\right],$$

which will be a bilinear function of $\dot{\varrho}$ and $\ddot{\varrho}$. Like in case of $z_t(\varrho)$, if we write $\rho_i(\gamma;\varrho)$ as an explicit function of ϱ by replacing the $v_i(\varrho)$'s by $A(\gamma)\varrho$, we can also obtain the matrix Ω.

Appendix F. Some useful definite integrals

Let

$$p(x) = a(x-\ell)^3 + b(u-x)^3 + c(x-\ell) + d(u-x),$$
$$q(x) = e(x-\ell)^3 + f(u-x)^3 + g(x-\ell) + h(u-x).$$

Then

$$\begin{aligned} p(x)q(x) = {} & ae(x-\ell)^6 + (af+be)(u-x)^3(x-\ell)^3 \\ & + bf(u-x)^6 + (ag+ce)(x-\ell)^4 \\ & + (ah+de)(u-x)(x-\ell)^3 \\ & + (bg+cf)(x-\ell)(u-x)^3 \\ & + (bh+df)(u-x)^4 + cg(x-\ell)^2 \\ & + (ch+dg)(u-x)(x-\ell) + dh(u-x)^2. \end{aligned}$$

Given that

$$\int_\ell^u (ae(x-\ell)^6 + (af+be)(u-x)^3(x-\ell)^3 + bf(u-x)^6)dx = -\frac{1}{140}(\ell-u)^7(20ae+be+af+20bf),$$

$$\int_\ell^u ((ag+ce)(x-\ell)^4 + (ah+de)(u-x)(x-\ell)^3 + (bg+cf)(x-\ell)(u-x)^3 + (bh+df)(u-x)^4)dx = -\frac{1}{20}(\ell-u)^5 \times (4ce+de+4ag+ah+bg+cf+4bh+4df),$$

and

$$\int_\ell^u (cg(x-\ell)^2 + (ch+dg)(u-x)(x-\ell) + dh(u-x)^2)dx = -\frac{1}{6}(\ell-u)^3(2cg+ch+dg+2dh),$$

we can finally write

$$\begin{aligned} \int_\ell^u p(x)q(x)dx = {} & -\frac{1}{140}(\ell-u)^7(20ae+be+af+20bf) \\ & -\frac{1}{20}(\ell-u)^5(4ce+de+4ag \\ & +ah+bg+cf+4bh+4df) \\ & -\frac{1}{6}(\ell-u)^3(2cg+ch+dg+2dh). \end{aligned}$$

Similarly,

$$\int k^{-x}(x-\ell)^3dx$$
$$= -\frac{1}{k^x \ln^4 k}((x-\ell)^3 \times \ln^3 k + 3(x-\ell)^2 \ln^2 k + 6(x-\ell)\ln k + 6),$$

so

$$\int_\ell^u k^{-x}(x-\ell)^3dx = -\frac{1}{k^u \ln^4 k}((u-\ell)^3 \ln^3 k + 3(u-\ell)^2 \times \ln^2 k + 6(u-\ell)\ln k + 6) + \frac{6}{k^\ell \ln^4 k}.$$

Likewise,

$$\int k^{-x}(u-x)^3dx$$
$$= \frac{1}{k^x \ln^4 k}(-(u-x)^3 \times \ln^3 k + 3(u-x)^2 \ln^2 k - 6(u-x)\ln k + 6)$$

so

$$\int_\ell^u k^{-x}(u-x)^3dx$$
$$= \frac{6}{k^u \ln^4 k} + \frac{1}{k^\ell \ln^4 k}(-(u-\ell)^3 \times \ln^3 k + 3(u-\ell)^2 \ln^2 k - 6(u-\ell)\ln k + 6).$$

Analogously,

$$\int k^{-x}(x-\ell)dx = -\frac{1}{k^x \ln^2 k}((x-\ell)\ln k + 1),$$
$$\int k^{-x}(u-x)dx = \frac{1}{k^x \ln^2 k}((x-u)\ln k + 1),$$

so

$$\int_\ell^u k^{-x}(x-\ell)dx = -\frac{1}{k^u \ln^2 k}((u-\ell)\ln k + 1) + \frac{1}{k^\ell \ln^2 k},$$
$$\int_\ell^u k^{-x}(u-x)dx = \frac{1}{k^u \ln^2 k} - \frac{1}{k^\ell \ln^2 k}((\ell-u)\ln k + 1).$$

Finally, it is straightforward to show that

$$\int_\ell^u k^{-x}dx = \frac{1}{\ln k}(k^{-\ell} - k^{-u}) \tag{F.4}$$

because

$$\int k^{-x}dx = -\frac{k^{-x}}{\ln k}.$$

Unfortunately,

$$\int_\ell^u k^{-x}dx = \frac{1}{\ln k}\left(\frac{1}{k^\ell} - \frac{1}{k^u}\right)$$

becomes numerically unstable when k is close to 1, even though in the limit this definite integral is simply $(u-\ell)$. In our empirical application to the RA model, this will happen in those quarters in which consumption growth is positive or negative but very close to 0. To avoid this problem, define d such that

$$1-d = k,$$

and consider the Taylor expansion of $(1-d)^{-x}$, which is given by:

$$1 + xd + \frac{1}{2}x(x+1)d^2 + \frac{1}{3!}x(x+1)(x+2)d^3 + O\left(d^4\right).$$

On this basis, we can numerically approximate the required integral by integrating the Taylor expansion, which yields

$$\int_\ell^u k^{-x}dx = (u-\ell) + \frac{1}{2}(u^2-\ell^2)d + \left[\frac{1}{3}(u^3-\ell^3) + \frac{1}{2}(u^2-\ell^2)\right]\frac{1}{2}d^2 + \left[\frac{1}{4}(u^4-\ell^4) + (u^3-\ell^3) + (u^2-\ell^2)\right] \times \frac{1}{3!}d^3 + O\left(d^4\right).$$

We can use the same trick to compute

$$\int_\ell^u k^{-x}(x-\ell)^3dx.$$

Specifically, the relevant terms in the expansion of this integral will be:

$$\int_\ell^u (x-\ell)^3dx = \frac{1}{4}(\ell-u)^4,$$
$$\int_\ell^u x(x-\ell)^3dx = \frac{1}{20}(\ell-u)^4(\ell+4u),$$
$$\int_\ell^u x(x+1)(x-\ell)^3dx = \frac{1}{60}(\ell-u)^4\left(\ell^2+4\ell u+3\ell+10u^2+12u\right),$$
$$\int_\ell^u x(x+1)(x+2)(x-\ell)^3dx = \frac{1}{140}(\ell-u)^4\left(\ell^3+4\ell^2u+7\ell^2+10\ell u^2 + 28\ell u+14\ell+20u^3+70u^2+56u\right).$$

Similarly,

$$\int_\ell^u (u-x)^3dx = \frac{1}{4}(\ell-u)^4,$$
$$\int_\ell^u x(u-x)^3dx = \frac{1}{20}(\ell-u)^4(4\ell+u),$$
$$\int_\ell^u x(x+1)(u-x)^3dx = \frac{1}{60}(\ell-u)^4\left(10\ell^2+4\ell u+12\ell+u^2+3u\right),$$
$$\int_\ell^u x(x+1)(x+2)(u-x)^3dx = \frac{1}{140}(\ell-u)^4(20\ell^3+10\ell^2u+70\ell^2+4\ell u^2 + 28\ell u+56\ell+u^3+7u^2+14u).$$

We also need to compute integrals such as

$$\int_\ell^u k^{-x}(x-\ell)dx.$$

To do so, we can use

$$\int_\ell^u (x-\ell)dx = \frac{1}{2}(\ell-u)^2,$$
$$\int_\ell^u x(x-\ell)dx = \frac{1}{6}(\ell-u)^2(\ell+2u),$$
$$\int_\ell^u x(x+1)(x-\ell)dx$$

$$= \frac{1}{12}(\ell-u)^2\left(\ell^2+2\ell u+2\ell+3u^2+4u\right),$$

$$\int_\ell^u x(x+1)(x+2)(x-\ell)dx$$

$$= \frac{1}{60}(\ell-u)^2(3\ell^3+6\ell^2u+15\ell^2+9\ell u^2+30\ell u+20\ell+12u^3+45u^2+40u).$$

Finally, to compute

$$\int_\ell^u k^{-x}(u-x)dx$$

we can use

$$\int_\ell^u (u-x)dx = \frac{1}{2}(\ell-u)^2,$$

$$\int_\ell^u x(u-x)dx = \frac{1}{6}(\ell-u)^2(2\ell+u),$$

$$\int_\ell^u x(x+1)(u-x)dx$$

$$= \frac{1}{12}(\ell-u)^2\left(3\ell^2+2\ell u+4\ell+u^2+2u\right),$$

$$\int_\ell^u x(x+1)(x+2)(u-x)dx$$

$$= \frac{1}{60}(\ell-u)^2(12\ell^3+9\ell^2u+45\ell^2+6\ell u^2+30\ell u+40\ell+3u^3+15u^2+20u).$$

References

Alonso-Borrego, C., Arellano, M., 1999. Symmetrically normalized instrumental-variable estimation using panel data. Journal of Business and Economic Statistics 17, 36–49.

Anderson, T.W., Rubin, H., 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Annals of Mathematical Statistics 20, 46–63.

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

Arellano, M., 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., Sentana, E., 1999. Underidentification? Mimeo. CEMFI.

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

Burnside, A.C., 2007. Empirical asset pricing and statistical power in the presence of weak risk factors. NBER Working Paper 13357.

Carrasco, M., Chernov, M., Florens, J.P., Ghysels, E., 2007. Efficient estimation of general dynamic models with a continuum of moment conditions. Journal of Econometrics 140, 529–573.

Carrasco, M., Florens, J.P., 2000. Generalization of GMM to a continuum of moment conditions. Econometric Theory 17, 794–834.

Carrasco, M., Kotchoni, R., 2010. Efficient estimation using the characteristic function. Mimeo. University of Montreal.

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

Choi, I., Phillips, P.C.B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations. Journal of Econometrics 51, 113–150.

Clarida, R.R., Galí, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. Quarterly Journal of Economics 115, 147–180.

Cochrane, J.H., 2011. Determinacy and identification with taylor rules. Journal of Political Economy 119 (3), 565–615. http://dx.doi.org/10.1086/660817.

Cragg, J.G., Donald, S.G., 1993. Testing identifiability and specification in instrumental variable models. Econometric Theory 9, 222–240.

Cragg, J.G., Donald, S.G., 1997. Inferring the rank of a matrix. Journal of Econometrics 76, 223–250.

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

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

Galí, J., Gertler, M., López-Salido, J.D., 2001. European inflation dynamics. European Economic Review 45, 1237–1270.

Hansen, L.P., 1982a. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L.P., 1982b. Large sample properties of generalized method of moments estimators. Unpublished Proofs. University of Chicago. http://www.sciencedirect.com/science/article/pii/S0304407612001200?v=s5.

Hansen, L.P., 1985. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. Journal of Econometrics 30, 203–238.

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14, 262–280.

Hansen, L.P., Singleton, K.J., 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. Econometrica 50.

Hansen, L.P., Singleton, K.J., 1983. Stochastic consumption, risk aversion and the temporal behavior of asset returns. Journal of Political Economy 91.

Hansen, Lars Peter, Singleton, Kenneth J., 1996. Efficient estimation of linear asset-pricing models with moving average errors. Journal of Business & Economic Statistics 14, 53–68.

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

Hillier, G.H., 1990. On the normalization of structural equations: properties of direction estimators. Econometrica 58, 1181–1194.

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

Kim, H., Ogaki, M., 2009. Purchasing power parity and the Taylor rule. Working Paper No. 09-03. Ohio State University, Department of Economics.

Kleibergen, F., 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73, 1103–1123.

Kocherlakota, N., Pistaferri, L., 2009. Asset pricing implications of Pareto optimality with private information. Journal of Political Economy 117, 555–590.

Koerts, J., Abrahamse, A.P.J., 1969. On the Theory and Application of the General Linear Model. Rotterdam University Press, Rotterdam.

Koopmans, T.C., Hood, W.C., 1953. The estimation of simultaneous linear economic relationships. In: Hood, W.C., Koopmans, T.C. (Eds.), Studies in Econometric Method. In: Cowles Commission Monograph Number, vol. 14. Wiley, (Chapter 6).

Mavroeidis, S., 2005. Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve. Journal of Money, Credit and Banking 37, 421–448.

Mehra, R., Prescott, E.C., 1985. The equity premium: a puzzle. Journal of Monetary Economics 15, 145–161.

Nagel, S., Singleton, K.J., 2011. Estimation and evaluation of conditional asset pricing models. Journal of Finance 66 (3), 873–909.

Nason, J.M., Smith, G.W., 2008. Identifying the new Keynesian Phillips curve. Journal of Applied Econometrics 23, 525–551.

Phillips, P.C.B., 1989. Partially identified econometric models. Econometric Theory 5, 181–240.

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

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

Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. Econometrica 26, 393–415.

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.

Sargan, J.D., 1983a. Identification and lack of identification. Econometrica 51, 1605–1633.

Sargan, J.D., 1983b. Identification in models with autoregressive errors. In: Karlin, S., Amemiya, T., Goodman, L.A. (Eds.), Studies in Econometrics, Time Series, and Multivariate Statistics. Academic Press, New York.

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

Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20, 518–529.

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

Yildiz, N., 2012. Consistency of plug-in estimators of upper contour and level sets. Econometric Theory 28, 309–327. http://dx.doi.org/10.1017/S0266466611000144.