24/01/23

1st Part: Optimization Based Estimation and Hypothesis Testing

We will try to develop a somewhat general theory of Estimation and Statistical Testing procedures that emerge in the context of Econometrics via mathematical optimization. The criteria involved usually represent (part of) the structure of the statistical/econometric model at hand. The optimization in several cases (e.g. in slightly more complicated models than the linear), cannot be analytically performed, and we rely on numerical/computational methods. The properties of the resulting procedures are evaluated via asymptotic theory. Many of the standard procedures used in econometrics - eg. OLSE, GLSE, MLE, GMME are specific instances of such-like optimization based procedures.

We need first to fix a general framework in which to develop our theory:

General Framework

The general framework will consist of:

Dimensionality of the typical element of the sample

1. the sample: typically a random element $z_n \in \mathbb{R}^{k \times n}$

In our examples z_n will mostly be an n-sized collection of random matrices. E.g. in the usual linear model $z_n = (Y_n, X_n)$ with Y_n is a $n \times 1$ random vector (dependent variable), X_n is a $n \times p$ random

$k = p + 1$

matrix (regressors). Z_n can be perceived as a $n\times(p+L)$ random matrix. In the general form of the instrumental variables linear model $Z_n = (Y_n, X_n, W_n)^*$ with Y_n, X_n as before and W_n a random matrix $n\times q$ of instruments. □

[More examples will be given shortly] $*\ k = p+q+L$

I. the object of inference: The joint distribution of the sample Z_n is (at least partially) unknown. The object of the statistical inference is to use information from Z_n in order to retrieve (approximate) the unknown desired characteristics of the distribution $(:= D_0)$

Example: Conditional mean

the conditional (on the information set - σ algebra - generated by X_n) mean of Y_n, i.e. $\mathbb{E}(Y_n/\sigma(X_n))$ is partially unknown.

Example. Weak instruments: A random element R_n such that its deviation from Y_n is orthogonal to a "set of instruments" W_n, i.e. $\mathbb{E}(W_n(Y_n - R_n)) = 0_{q\times 1}$.

II. Exogenous (w.r.t. Z_n) structure employed for the object of inference: in our framework it is assumed that the object of inference D_0 (partially) depends on the unknown value of an Euclidean parameter

$$\Longrightarrow D_0 = D(\theta_0)$$

$\theta_0 \in \Theta \subseteq \mathbb{R}^p$ [$p \in \mathbb{N}^*$]. Knowledge of θ_0 would be considered as a sufficient reduction of the problem of knowing the desired characteristics of D_0. θ_0 is unknown; hence its recovery (or approximation) is the purpose of the statistical inference.

[We thus restrict ourselves into the framework of parametric description of the desired characteristics of the unknown D_0. Non parametric considerations - although very interesting - are out of the scope of our lectures]

i.e. $Y_n = X_n\theta_0 + \varepsilon_n$
$\mathbb{E}(\varepsilon_n / \sigma(X_n)) = 0_{n\times 1}$

Example Usual linear model: we assume that the aforementioned conditional mean, is linear w.r.t. the information i.e. is of the form $X_n\theta_0$, $\theta_0 \in \mathbb{R}^p$ ∎

Example Weak instruments linear model: we assume that

(*) $\mathbb{E}(W_n'(Y_n - X_n\theta_0)) = 0_{q\times 1}$, for $\theta_0 \in \mathbb{R}^p$. Notice that when $p=q$ and $W_n = X_n$, (*) also holds in the usual linear model. ∎

Example

a generalization of the usual linear model:

$$\mathbb{E}(Y_n / \sigma(X_n, W_n)) = g(\theta, X_n)$$

where $g : \Theta \times \mathbb{R}^{n\times p} \to \mathbb{R}^n$ is known:

e.g. $g(\theta, X_n) = (\exp(\theta' X_{(i)}))_{i=1,\dots,n}$

where $X_{(i)}$ is the i^{th} row of X_n.

Notice that in this case the conditional mean is not linear in X_n, but knowledge of the unknown θ_0 would be equivalent with knowledge of $\left(\exp(\theta_0' X_{(i)})\right)_{i=1,\dots,n} =$
$= \mathbb{E}\left(Y_n / \sigma(X_n)\right)$. $\blacksquare$

- Notice that in all the above cases, knowledge of θ_0, does not imply knowledge of the <u>entire</u> conditional (on the relevant information set) distribution of Y_n. ("only" the conditional mean is thus known.)

- Notice also in the above cases that θ_0 is independent of n, but $\mathbb{E}\left(Y_n / \sigma(X_n)\right)$ is not. Hence θ_0 essentially represents (some of) the properties of the conditional mean that do not depend on the information sets.

Since θ_0 is unknown, it will be "searched for" on a set of values for the parameter, that may contain θ_0. This will be denoted by Θ, referred to as the Parameter Space, and will be structured by any further exogenous information that may be available to the researcher beyond Z_n.

This defines a set of potential distributions for Z_n,

$$\Theta \ni \theta \longrightarrow D(\theta)$$

i.e. every possible value for the parameter θ defines a set of possible distributions of Z_n.

Example Usual linear Model

Suppose that no other exogenous information is known for θ_0, so Θ is maximal, i.e. $\Theta = \mathbb{R}^p$. If $\theta \in \mathbb{R}^p$, then $X_n\theta$ is a potential conditional mean for Y_n, and $D(\theta)$ is the set of conditional distributions with mean $X_n\theta$. Notice that in this case the correspondence $\theta \to D(\theta)$ is multivalued.

It could be the case that more information is known about $D(\theta_0)$, e.g. that $D(\theta_0) = N(X_n\theta_0, I_{n\times n})$ with $I_{n\times n}$ the relevant identity. In such a case "only" the conditional mean is unknown - since θ_0 is unknown - and incorporating this information we obtain that the correspondence

$$\theta \longrightarrow D(\theta) = N(X_n\theta, I_{n\times n}) \quad \text{is}$$

actually a function (to each possible value of θ a unique Gaussian distribution is associated). ■

Definition Returning to the general case the correspondence

$$\Theta \ni \theta \longrightarrow D(\theta)$$

defines a (parametric) statistical Model for the approximation of θ_0; it is the set of distributions

$$\{ D(\theta), \theta \in \Theta \} \quad (*)$$

- when the correspondence $\theta \to D(\theta)$ is actually multi-valued, the model is termed semi-parametric (θ does not uniquely specify the distribution; essentially an "infinite dimensional part" is left unparametrized.
- when the correspondence $\theta \to D(\theta)$ is actually a function, the model is termed (fully) parametric.

Example

- the $\{\mathbb{E}(Y_n/\sigma(X_n)) = X_n\theta, \theta \in \mathbb{R}^p\}$ (*)

is a semi-parametric linear model.

- the $\{Y_n/\sigma(X_n) \sim N(X_n\theta, I_{n\times n}),$

$\theta \in \mathbb{R}^p\}$

is a (Gaussian) parametric linear model. ■

Definition The model $\{D(\theta), \theta \in \Theta\}$ is termed econometric when its specification is partially base on exogenous information about the distributions involved derived from the economic theory and/or other empirical properties of economic phenomena.

Example In the usual linear model Y_n is a time series of observed stock logarithmic excess returns, and $X_n = W_n$, $p=1$ and is a vector of market logarithmic returns, then the financial economics CAPM predicts that $\mathbb{E}(Y_n/\sigma(X_n)) = X_n\theta_0$ for some unknown $\theta_0 \in \mathbb{R}$, hence the econometric semiparametric linear model (*) with $p=1$. ■ 1

Example

(*) Suppose that the firm $j \in \{1,2\}$ decides to enter the market m, $(Z_{j,m}=1)$ or not $(Z_{j,m}=1)$ $m \in \{1,2,\dots,M\}$, based on the profit function

$$\Pi_{j,m} = \left[\varepsilon_{j,m} - \theta_{0j} Z_{-j,m}\right] 1_{Z_{j,m}=1}, \text{ with } \varepsilon_{j,m} \sim \text{Unif}[0,1]$$

$Z_{-j,m}$ is the other firms decision to enter market M, $\theta_0 = (\theta_{0,1}, \theta_{0,2}) \in \Theta = (0,1)\times(0,1)$. $\varepsilon_{j,m}$ is the stochastic utility of j from entering in M. We can prove that the

Stochastic game has the following outcomes:

i. $(Z_{1,m}, Z_{2,m}) = (1,1)$, if $\varepsilon_{j,m} \geq \theta_{0j}$ $\forall j=1,2$

ii. $(Z_{1,m}, Z_{2,m}) = (1,0)$ if $\varepsilon_{1,m} \geq \theta_{01}$, $\varepsilon_{2m} < \theta_{02}$

iii. $(Z_{1,m}, Z_{2,m}) = (0,1)$ if $\varepsilon_{1,m} < \theta_{01}$, $\varepsilon_{2m} \geq \theta_{02}$

iv. $(Z_{1,m}, Z_{2,m}) = \left\{\begin{matrix}(0,1)\\(1,0)\end{matrix}\right\}$ if $\varepsilon_{j,m} < \theta_{0,j}$ $\forall j=1,2$

↳ Multiple Nash equilibria

(4)

Without any further assumptions we have that (easy calculations based on independence)

$$\mathbb{P}(\{i\}) = \mathbb{P}(\varepsilon_{1,m} \geq \theta_{01,m})\,\mathbb{P}(\varepsilon_{2,m} \geq \theta_{02,m})$$
$$= (1-\theta_{01,m})(1-\theta_{02,m})$$

$$\mathbb{P}(\{ii\}) \leq \mathbb{P}\left((Z_{1,m}, Z_{2,m}) = (1,0)\right) \leq \mathbb{P}(Z_{2,m}=0)$$

$$\Leftrightarrow (1-\theta_{0,1})\theta_{0,2} \leq \mathbb{P}\left((Z_{1,m}, Z_{2,m})=(1,0)\right) \leq \theta_{0,2}$$

which can be equivalently be rewritten as the following system of moment inequalities:

$$\left.\begin{array}{l}\mathbb{E}\left(z_{1,m}z_{2,m}-(1-\theta_{0,1})(1-\theta_{0,2})\right)=0\\ \mathbb{E}\left(z_{1,m}(1-z_{2,m})-(1-\theta_{0,1})\theta_{0,2}\right)\geq 0\\ \mathbb{E}\left(\theta_{0,2}-z_{1,m}(1-z_{2,m})\right)\geq 0\end{array}\right\}$$

we will not further discuss this example for a while!

Given a sample of $(z_{1,m}, z_{2,m})_{m=1,\dots,M}$ of observed decisions for the two firms we obtain the econometric parametric model of moment inequalities

$$\left\{\left.\begin{array}{l}\mathbb{E}\left(z_{1,m}z_{2,m}-(1-\theta_1)(1-\theta_2)\right)=0\\ \mathbb{E}\left(z_{1,m}(1-z_{2,m})-(1-\theta_1)\theta_2\right)\geq 0\\ \mathbb{E}\left(\theta_2-z_{1,m}(1-z_{2,m})\right)\geq 0\end{array}\right\},\ \theta=\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}\in\Theta=(0,1)\times(0,1)\right\}$$

Example GARCH(1,1)-Model

(*)

The empirical characteristics of financial logarithmic excess returns imply certain properties of conditional hetero-skedasticity. Such properties can be partly modelled by by stochastic processes that are defined as solutions to Stochastic Recurrence Equations like:

- given z_t iid with $\mathbb{E}(z_0)=0$, $\mathbb{E}(z_0^2)=1$

(*)
- $y_t = z_t\sigma_t$, $t\in\mathbb{Z}$
- $\sigma_t^2 = w_0 + (a_0 z_{t-1}^2 + b_0)\sigma_{t-1}^2$, $t\in\mathbb{Z}$

with $w_0>0$, $a_0, b_0\geq 0$ unknown.

It turns out that when $\mathbb{E}\left(\ln\left(\alpha_0 Z_0^2 + b_0\right)\right) < 0$

there exists a unique stationary and ergodic solution to $(*)$ specified as:

$$y_t = Z_t \left(\omega_0 \left(1 + \sum_{j=0}^{\infty} \prod_{i=0}^{j} \left(\alpha_0 Z_{t-i}^2 + \beta_0\right)\right)\right)^{1/2}, \quad t \in \mathbb{Z}$$

which is termed as a GARCH process. Even if the above hold, $\theta_0 := (\omega_0, \alpha_0, b_0)$ is unknown and the relevant semiparametric model that is thus structured, for a given time series sample of observable returns $(y_t)_{t=1,\dots,n}$

$$\begin{cases} y_t = z_t \sqrt{h_t} & , \ t = 1, \dots, n \\ h_t = \omega + \alpha y_{t-1}^2 + \beta h_{t-1} & \text{[* corrected typo]} \end{cases}$$

$$\left\{ \theta = (\omega, \alpha, \beta) \in \Theta, \ \Theta = \{(\omega, \alpha, \beta) \in \mathbb{R}^3 : \omega > 0, \ \alpha, \beta \geq 0, \ \mathbb{E} \ln(\alpha z_0^2 + \beta) < 0\} \right\}$$

$\blacksquare$

Remark:

Notice that both the previous examples reveal that the representations of an econometric model can be quite diverse: moment conditions, recurrence equations, etc. In all cases though everything is reducible to structure about the particular characteristics of interest for the joint distribution of the sample. $\blacksquare$

Definition. The statistical model is well-specified iff θ_0 lies in it.

- In all other cases the model is misspecified. Misspecification may occur due to that the model is not correctly parameterized, e.g. some regressor is omitted in the linear model, or the conditional mean is not linear, or due to that $\theta_0 \notin \Theta$, e.g. $\mathbb{E}\left(\mathbb{R}(\alpha_0 z_0^2 + b_0)\right) > 0$ in the previous example.

Assumption. We will be working with correctly specified models (Specification analysis is also part of statistical inference; we will not touch this!)

Inference

1. **Estimation:** Given the model an estimator of θ_0 is any measurable function $\hat{\theta}_n : \mathbb{R}^{kn} \to \Theta$

↳ this is needed in order for $\hat{\theta}_n$ to have a well-defined distribution. Its properties will depend on that.

For fixed n the properties of the estimators are generally unknown. Those become more discernible when $n \to +\infty$ due to the limit theory part of Probability theory. Thus:

Definition $\hat{\theta}_n$ is weakly consistent iff

$$\forall \varepsilon > 0, \lim_{n \to +\infty} \mathbb{P}(\|\hat{\theta}_n - \theta_0\| > \varepsilon) = 0$$

↳ Euclidean distance between $\hat{\theta}_n$ and θ_0

↳ Consistency derivations can be facilitated by LLNs. [Minimal Property]

Definition. if $\hat{\theta}_n$ is weakly consistent and (r_n) is a real sequence such that $r_n \to +\infty$, then $\hat{\theta}_n$ has rate r_n, iff $\forall \varepsilon > 0, \exists M > 0$:

$$\lim_{n \to +\infty} \mathbb{P}(r_n \|\hat{\theta}_n - \theta_0\| > M) \leq \varepsilon.$$

↳ when it exists r_n is termed rate of convergence [Desired as fast as possible]

Definition. If $\hat{\theta}_n$ is weakly consistent with rate r_n, then it is termed asymptotically Gaussian iff $\exists Z \sim N(0, V_{p \times p})$ such that $r_n(\hat{\theta}_n - \theta_0) \rightsquigarrow Z$

↳ Convergence in distribution

↳ the extraction of rates and asymptotic distributions can be facilitated by CLTs.

V is the asymptotic variance of $\hat{\theta}_n$. Desired as "small" as possible.

B. Inference: Suppose that the analyst wishes to empirically validate the hypothesis that $\theta_0 \in \Theta^* \subseteq \Theta$ against $\theta_0 \in \Theta_* \subseteq \Theta$ with $\Theta^* \cap \Theta_* = \phi$. It thus specifies the hypothesis structure

$$H_0 : \theta_0 \in \Theta^* \qquad (**)$$
$$H_1 : \theta_0 \in \Theta_*.$$

A statistical test of $(**)$ given a significance level $\alpha \in (0,1)$ is a measurable function

$$t_n(\alpha) : \mathbb{R}^{kn} \longrightarrow \{\text{cannot reject } H_0, \text{reject } H_0\}$$

Definition: The test is termed asymptotically conservative iff $\lim_{n\to+\infty} \mathbb{P}(\text{reject } H_0 \text{ by } t_n \mid H_0 \text{ holds}) \leq \alpha$. (exact when $= \alpha$)

The test is termed consistent iff

$$\lim_{n\to+\infty} \mathbb{P}(\text{reject } H_0 \text{ by } t_n \mid H_1 \text{ holds}) = 1.$$

[Minimal Properties]

Optimization Based Estimators and Testing Procedures

In several cases already mentioned before, the structure of the statistical model implies that θ_0 can be recovered via some sort of variation principle; i.e. as an optimizer of (a usually intractable) real function.

Example Consider the usual linear model enhanced with a restriction of the existence of the second conditional moments, and homoskedasticity, i.e.

$$\left\{ \begin{array}{l} \mathbb{E}(Y_n / \sigma(X_n)) = X_n\theta, \ \theta \in \Theta \\ Var(Y_n / \sigma(X_n)) = I_{n\times n} \quad (*) \end{array} \right\}$$

In such a case by construction (why) θ_0 satisfies

$$\theta_0 = \underset{\theta \in \Theta}{\operatorname{argmin}} \ \mathbb{E}\left(\left((Y_n - X_n\theta)'(Y_n - X_n\theta)\right) / \sigma(X_n)\right)$$

If the criterion $\theta \to \mathbb{E}\left(((Y_n - X_n\theta)'(Y_n - X_n\theta)) / \sigma(X_n)\right)$ where actually known then it could be the case that θ_0 could be recovered by solving the optimization problem. However the criterion is unknown since it depends on θ_0 via $\mathbb{E}(\cdot / \sigma(X))$.

(*) without loss of generality

i.e. the structure of the statistical/econometric model is such that, $\exists\, C_{D_0} : \Theta \to \mathbb{R}$

$$: \quad \theta_0 = \underset{\theta \in \Theta}{\operatorname{argmin}}\, C_{D_0}(\theta).$$

$\hookrightarrow$ it may also depend on the sample

However C_{D_0} is intractable due to its dependence on D_0. In several cases C_{D_0} is also approximable by its empirical analogue:

Example (continued) the empirical analogue of $\mathbb{E}\left(\left(Y_n - X_n\theta\right)'\left(Y_n - X_n\theta\right)/\sigma(X_n)\right)$ is the empirical mean $\frac{1}{n}\left(Y_n - X_n\theta\right)'\left(Y_n - X_n\theta\right)$.

Hence there exists a $C_n(\theta) : \mathbb{R}^{k n} \times \Theta \to \mathbb{R}$ that is tractable (via the sample Z_n) and somehow approximates C_{D_0}.

Thus we immediately obtain the following definition:

Definition Given the aforementioned structure an Optimization based estimator (OE) is defined by

$$\theta_n \in \underset{\theta \in \Theta}{\operatorname{argmin}}\, C_n(\theta).$$

Example. In the (weak) instrumental variables linear model the empirical analogue of the moment conditions that define the structure of the model is $\frac{1}{n}W_n(Y_n - X_n\theta)$.

$\mathbb{E}(W_n(Y_n - X_n\theta)) : \Theta \to \mathbb{R}^q$ when well defined.

Let V_q be a strictly positive definite $q \times q$ matrix i.e. $\forall z \in \mathbb{R}^q$ $z'V_q z = 0 \iff z = 0_{q\times 1}$. As a matter of fact, the function $z \to (z'V_q z)^{1/2}$ is a norm on $\mathbb{R}^q$ that is uniquely minimized at $z = 0_{q\times 1}$.

Hence, given V_q, in this instance*

$$C(\theta) := \mathbb{E}(W_n(Y_n - X_n\theta))' V_q \mathbb{E}(W_n(Y_n - X_n\theta))$$

has the property that $\theta_0 \in \operatorname{argmin}_{\theta\in\Theta} C(\theta)$. Given the remarks above about the empirical analogue an OE estimator of θ_0 can be defined by

$$(\alpha) \quad \theta_n \in \operatorname{argmin}_{\theta\in\Theta} \left(\frac{1}{n}W_n(Y_n - X_n\theta)\right)' V_q \left(\frac{1}{n}W_n(Y_n - X_n\theta)\right)$$

(α) is termed as the IV estimator of θ_0 (IVE).

1. How is V_q chosen?
2. Are other norms in $\mathbb{R}^q$ usable?

$\boxtimes$

* Optimization-wise without loss of generality we can work with the squared norm - why?

Example. In the aforementioned GARCH(1,1) model it can be proven that

(A) $$\theta_0 \in \operatorname*{argmin}_{\theta \in \Theta} \underbrace{\mathbb{E}\left(\ln h_t(\theta) + \frac{z_t^2 \sigma_t^2}{h_t(\theta)}\right)}_{(B)}$$

with $h_t(\theta)$ defined by the stochastic recurrence equation

$$h_t(\theta) = \omega + \alpha y_{t-1}^2 + b h_{t-1}(\theta), \quad t \in \mathbb{Z}$$

It can also be proven that for arbitrary $h_0 > 0$, the empirical analogue

$$C_n(\theta) := \frac{1}{n} \sum_{t=1}^{n} \left(\ln h_t^*(\theta) + y_t^2 / h_t^*(\theta)\right)$$

with $$h_0^* = h_0, \quad h_t^*(\theta) = \omega + \alpha y_{t-1}^2 + b h_{t-1}^*(\theta),$$
$$t = 1, \dots, n$$

provides an "asymptotically plausible" approximation of (B), hence an OE estimator can be defined by

$$\theta_n \in \operatorname*{argmin}_{\theta \in \Theta} C_n(\theta).$$

1. θ_n is the Gaussian Quasi Maximum Likelihood Estimator (QMLE) of θ_0.
2. Why (A) holds? 3. In what sense is the aforementioned approximation plausible? ☒

Returning to the general case:

Question: How can we be sure that $\hat{\theta}_n$ exists?

— Meaning of existence: a. $\underset{\theta\in\Theta}{\operatorname{argmin}}\, G_n(\theta) \neq \emptyset$ with probability 1

(actually it would be even more helpful if the argmin is a singleton)

b. $\hat{\theta}_n$ should be measurable (i.e. have a well defined probability distribution so that its properties are definable, at least as $n\to\infty$)

Answer: a. can be ensured by several conditions:

e.g. if $G_n:\Theta\to\mathbb{R}$ is continuous with probability 1 and Θ is compact (closed and bounded) then $\underset{\theta\in\Theta}{\operatorname{argmin}}\, G_n(\theta)\neq\emptyset$, by a Weierstrass Theorem type argument. (not necessarily singleton though — then a selection must be used)

e.g. if $G_n:\Theta\to\mathbb{R}$ is strictly convex, and Θ is a closed convex subset of $\mathbb{R}^q$ with probability

1, then $\arg\min_{\theta \in \Theta} G_n(\theta)$ singleton. E.g. this is ensured in the usual linear model if rank $X_n = q$, and $\Theta = \mathbb{R}^p$. Then $\hat{\theta}_n = (X_n' X_n)^{-1} X_n' Y_n$ (why?)

b. Under slightly stronger conditions (e.g. joint continuity of G_n w.r.t. θ, Z_n), and due to the fact that Θ is Euclidean it can be proven (using a result called Measurable Projections Theorem) that $\hat{\theta}_n$ is indeed a random vector (thus has a well defined distribution).

Question: Is $\hat{\theta}_n$ generally analytically traceable, i.e. can the defining optimization problem be analytically solved?

Answer. Generally it isn't. The OLSE estimator ③ is a simple special case of solvability (but this can become also non traceable for $\Theta \neq \mathbb{R}^q$ - can you derive the OLSE for $\Theta = \mathbb{Z}^q$ - the estimator components can only assume integral values). Consider the Gaussian QMLE in the GARCH(1,1) example. You can easily see that it does not have a closed form; Thus:

1. The OE estimators do not generally have closed forms due to the complexity of the optimization problem at hand. [Finite sample properties are difficult to establish!]

2. Their derivation is thus generally based on numerical/computational procedures (this implies that their definition should also involve, and their properties depend on, the algorithm used)

3. Their properties must be thus derivable from the properties of the optimization problem.

We will focus our analysis in the limit theory suchlike estimators and corresponding testing procedures. Our derivations will involve high level assumptions and general properties on the issue of approximation of optimization problems.

Before doing so we need the notion of identification (otherwise analysis will at least become quite complicated; we will try to obtain a flavor of the partial identification framework via the Nash equilibria example later on.)

Definition For the statistical model at hand, θ_0 is identified by the criterion $C(\theta)$ iff $\theta_0 = \underset{\theta \in \Theta}{\arg\min}\, C(\theta)$.

Remark This essentially means that θ_0 is uniquely recoverable via minimization of $C(\theta)$. This fails if $\underset{\Theta}{\operatorname{argmin}}\, C(\theta)$ is non singleton.

When identification fails, the part of the statistical model represented by C, contains not enough information about θ_0. Exogenous further information is thus required.

Identification holds when - for example - $C(\theta)$ is strictly convex and additionally Θ is convex.

Example: In the usual linear model with homoskedasticity we have that:

$$\mathbb{E}\left[(Y_n - X_n\theta)'(Y_n - X_n\theta)\big|_{\sigma(X_n)}\right] =$$

$$\mathbb{E}\left(\left(X_n(\theta_0-\theta)+\varepsilon_n\right)'\left(X_n(\theta_0-\theta)+\varepsilon_n\right)\big|\sigma(X_n)\right) =$$

$$= (\theta_0-\theta)'\,\mathbb{E}\left(X_nX_n'/\sigma(X_n)\right)(\theta_0-\theta) + \mathbb{E}\left(\varepsilon_n'\varepsilon_n/\sigma(X_n)\right)$$

$$+\ 2(\theta_0-\theta)'\,\mathbb{E}\left(X_n'\varepsilon/\sigma(X_n)\right)$$

$$= (\theta_0-\theta)'X_nX_n'(\theta_0-\theta) + \eta\,,$$ and this is uniquely minimized at θ_0 iff $\operatorname{rank}(X_n) = p$. ☒

Remark We must not confuse the issue of uniqueness of $\arg\min_{\theta\in\Theta} C(\theta)$, with the issue of uniqueness of $\hat{\theta}_n$.

The former may hold but the latter may not (as mentioned before some selection procedure may be needed for $\hat{\theta}_n$). The usual linear model and the OLSE may confuse us due to the conditional (on $\sigma(X_n)$) analysis. Under unconditional analysis the identification condition becomes $\operatorname{rank} \mathbb{E}(X_n) = p$, a sufficient condition of which is $\operatorname{rank}(X_n) = p$

What is important in our case is an asymptotic version of the notion of identification:

Definition Suppose that as $n \to +\infty$ C_n converges (in some manner that we will discuss later) to a deterministic function $C^*: \Theta \to \mathbb{R}$.

independent of the sample.

Then θ_0 is asymptotically identifiable by $C_n(\theta)$ iff $\theta_0 = \arg\min_{\theta\in\Theta} C^*(\theta)$.

In many cases $C = C^*$ and the two notions coincide. (we will examine several examples)

Example In the usual linear model this holds when $\frac{X_n X_n'}{n}$ converges in probability to a matrix of rank q. This is stronger than $\text{rank}\, X_n = q$. ∎

Appendix

* this colouring denotes typos corrections (24/01/23)

* This colouring denotes endnotes numberings in the main text.

Endnotes

1. Notice that in this special case the regression matrix may also contain other regressors, e.g. a constant vector of 1, etc. This augmented structure is useful in testing the validity of the economic relation implied by the CAPM; e.g. if the coefficient of the constant regressor is "systematically" statistically significant, then this constitutes empirical evidence against the "universal" validity of the CAPM. ∎

→

(2) As commented in the follow up notes of consistency a more general framework of existence, is that C_n is bounded from below (in θ), with probability 1, if a non necessarily zero "optimization error" is allowed in the definition:

for $u_n \geq 0$ with probability 1, then $\hat{\theta}_n$ is defined by the relation:

$$C_n(\hat{\theta}_n) \leq \inf_{\theta \in \Theta} C_n(\theta) + u_n \quad (*)$$

When $u_n = 0$ (and if $\arg\min_{\theta \in \Theta} C_n(\theta) \neq \emptyset$) we recover the original definition. If $\arg\min_{\theta \in \Theta} C_n(\theta)$ with positive probability then for any $u_n > 0$ with probability 1, $(*)$ provides a well defined estimator - why?

$\blacksquare$

(3) Actually, and since $\operatorname{rank} X_n = p$ (with probability 1) and thus C_n is strictly convex in Θ (with probability 1) the derivation of the OLSE for a general Θ can be proven to emerge via $\hat{\theta}_n \in \arg\min_{\theta \in \Theta} \|(x_n' x_n)^{-1} x_n' y_n - \theta\|$, and this has a unique solution if Θ is closed and convex.

$\blacksquare$

(4) Even though we have not managed to examine this strategic interaction, we have specified a game with "stochastic" utility. The outcomes are stochastic Nash equilibria. $\blacksquare$