Inverse Problems in Econometrics: Examples and Specific Theoretical Problems

Jean-Pierre Florens

[1] *Toulouse School of Economics*

Mois de la statistique - CIRM - Marseille

Février 2016

Contents

$\boxed{\text{Structural modelling}}$ statistics plus economic theory.

- Statistical part: data generated by a probability measure characterized by a function F

  e.g. $X \sim P$ *cdf* F

  $(x_1, ...x_n)$ iid sample of X.

- Economic model: definition of the objects of interest (parameters, functions...) denoted in general by φ.

- Relation between statistics and economics:

  an implicit relation:

$$A(\varphi, F) = 0$$

"Inverse problem".

From the data F is estimated by $\hat{F}$ and φ is estimated by solving $A(\varphi, \hat{F}) = 0$

(generalized method of moments).

$\exists\ h(\varphi, X)$ such that

$$A(\varphi, F) = E(h(\varphi, X))$$

Linear wrt F but not wrt φ.

Conditional method of moments:
$\exists\ h(\varphi, X)$ and Z function of X such that

$$A(\varphi, F) = E(h(\varphi, X)|Z)$$

Some econometric terms

$$\varphi \in \mathcal{E}, F \in \mathcal{F} \quad A : \mathcal{E} \times \mathcal{F} \to \mathcal{H}$$

suitable functional spaces.

- Well specified model:

$$\exists(\varphi_*, F_*) \text{ such that } A(\varphi_*, F_*) = 0 \quad X \sim F_*$$

$$\varphi_* \in \mathcal{E} \ F_* \in \mathcal{F}.$$

- Identified model (locally identified).
  if $A(\varphi, F_*) = 0 \Rightarrow \varphi = \varphi_* \quad \forall F_*$
  (if $A(\varphi, F_*) = 0$ and φ in a neighborhood of $\varphi_* \Rightarrow \varphi = \varphi_*$)
  else $\{\varphi / A(\varphi, F_*) = 0\}$ identified set.
- Non overidentified model.
  $\forall F \in \mathcal{F} \ \exists \varphi \in \mathcal{E} / A(\varphi, F) = 0$ (in particular for an estimator of F).
  else: overidentification.

A model is well-posed if it is well-specified, identified and non overidentified and if the unique φ corresponding to F is a continuous function of F.

- Exogeneity: $X = (Y, Z)$

  Z is exogenous if the solution of $A(\varphi, F) = 0$ does not depend on the marginal distribution of Z (up to null sets). All the information on φ is captured by the conditional probability of Y given Z.

  Contrary: Endogenous.

  Fundamental concept because econometrics is not in general experimental and because data are generated by equilibrium.

$$\begin{aligned}&A(\varphi, F) = r_F - K_F\varphi\\&K_F : \mathcal{E} \to \mathcal{F} \text{ linear operator}\\&r_F \in \mathcal{F}\\&\varphi \in \mathcal{E}\end{aligned}$$

A simple case: linear functional regression.

$Z, \varphi \in$ Hilbert space
$Y \in \mathbb{R}$ $U \in \mathbb{R}$

$$Y = \langle Z, \varphi\rangle + U \quad U \text{ random noise} \quad E(ZU) = 0$$

$$\Rightarrow \quad \underbrace{VarZ}_{K_F}(\varphi) = \underbrace{Cov(Z, Y)}_{r_F}$$

Toulouse School of Economics

- Usually we have data $(x_1, ..., x_n)$ and functional objects $r(), \varphi()$ depending on the distribution of the data and associated by a relation

  $$\boxed{r = T(\varphi)}$$

  $\hat{r}(.)$ is an estimation of r and

  $$\boxed{\hat{r} = T(\varphi) + U}$$

  where U is a noise with known properties
  (ex: $E(U) = 0 \; V(U) = \Sigma$) Statistical Inverse Problem
- In most of the case T is unknown and should be estimated using the same data sample $(x_1, ..., x_n)$.
  Ex: $T =$ Covariance Operator

  $$T(\varphi) = E(W\langle Z, \varphi\rangle)$$

  T : conditional expectation operator

  $$T(\varphi) = E(\varphi(Z)|W)$$

- The functional parameters of interest are <u>simple</u> (very smooth and with simple shape).
  Prior knowledge on φ (increasing, convex) $\Rightarrow$ Bayesian approach.
- But T has also usually an important smoothing power.
- The primary estimation of r or T are usually <u>non parametric estimation</u>

  (e.g. $r = E(Y|Z)\ T(\varphi) = E(\varphi(Z)|W)\ X = (Y, Z, W)$)

  $\Rightarrow$ Two levels of selection of the regularisation parameters (estimation of r and T, inversion of $\hat{T}$).
- The operator T may be not injective (or its estimator)
  $\rightarrow$ Identification problem.

- Three sources of error on the estimation of φ.

$$\left\{\begin{array}{l} -\text{ Error } \hat{r} - r \\ -\text{ Error } \hat{T} - T \\ -\text{ Regularisation bias} \end{array}\right.$$

- Importance of tests
  - Parametric model/non parametric
  - Tests between solutions of different inverse problems.

$$\left.\begin{array}{l} \hat{r}_1 = T_1(\varphi_1) + U_1 \\ \hat{r}_2 = T_2(\varphi_2) + U_2 \end{array}\right\} \text{ test } \varphi_2 = A(\varphi_1)$$

    Ex: exogeneity tests.
  - Tests of shape constraints.

- Successive inverse problems
  Ex: $r = T(\varphi)$ $\quad r =$ Distribution function

  r may be estimated directly $(\frac{1}{n}\sum_i \mathbb{1}(y_i \leq t))$ or
  r may result from a deconvolution.
  $r =$ distribution function of Y^* but Y^* is not observable and we observe $Y = Y^* + \eta$

  Ex: $\varphi(s, t)$ solutions of an inverse problem but interest to $m(s)$ such that

  $$m'(s) = \varphi(s, m(s))$$

  Theory of consumer surplus.
- Linear model with a partially unknown operator $r = K^\sigma \varphi$. Joint estimaton of φ and σ (example: convolution model with normal error and unknown variance).

III-A

- Usual regression model: $(Y,Z) \in \mathbb{R} \times \mathbb{R}^p$

$$\begin{aligned} Y &= E(Y|Z) + U \\ &= \varphi(Z) + U \quad E(U|Z) = 0 \end{aligned}$$

- Instrumental variables model: $(Y,Z,W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$

$$Y = \varphi(Z) + U \quad E(U|W) = 0$$

Linear integral equation of type I.

$$\int \varphi(z) f(z|w) dz = \int y f(y|w) dy$$

$$K\varphi = r$$

Very important in econometric: Endogeneity/Selection bias/treatment models.

III-B

- treatment model (counterfactual model)

$$Y = \varphi(\zeta) + U$$

ζ non random treatment level
U random component

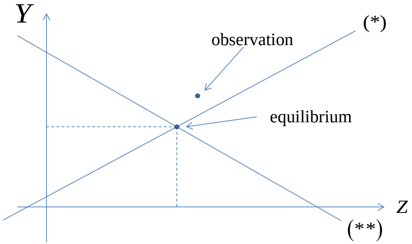
Assignment mechanism: choice of Z non independent of U. (not a randomized experiment)

$$\Rightarrow \quad Y = \varphi(Z) + U \quad \text{but} \quad E(U|Z) \neq 0$$

Identification by observation of W (Instrumental variable) explaining Z but mean independent of U ($E(U|W) = 0$).

- endogeneity: systems of equations (linear for simplicity)

$$\begin{cases} Y = aZ + bW_1 + U_1 & (*) \\ \\ Z = cY + dW_2 + U_2 & (**) \end{cases} \qquad E\left(\begin{array}{c} U_1 \\ U_2 \end{array} \middle| W_1, W_2 \right) = 0$$



The two equations are not regression models

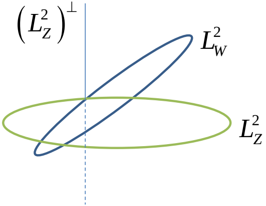
$$(E(U_1|Z, W_1) \neq 0).$$

III-C Identification

- Correct specification: $r \in \mathcal{R}ange(K)$.
- Identification: K one to one.
  Dependence condition between W and Z.
  Spaces; $L^2_Z, L^2_W \ldots$ wrt the true distribution.
  L^2_W does not contain a function orthogonal to L^2_W.

$$E\left(\varphi(Z)|W\right) = 0$$

$$\Rightarrow \varphi = 0 \quad \text{(a.s)}$$



In most of the case K is compact (implied by an Hilbert Schmidt condition) :

$$\int \left(\frac{f(z,w)}{f(z)f(w)} \right)^2 f(z)f(w)dzdw < \infty$$

$$K^* \psi(w) = E(\psi(W)|Z)$$

Singular value decomposition of K: $(\lambda_j, \varphi_j(z), \psi_j(w))$

K one to one $\Leftrightarrow \lambda_j \neq 0 \; \forall j$.
K one to one: "Strong Identification"
"Completness"

(remember that in a statistical model a statistic $\varphi(x)$ is complete if
$E(\varphi(x)|\theta) = 0 \Rightarrow \varphi = 0 \quad \varphi \in L^2$.
L^p complete: $\varphi \in L^p$).

Non testable assumption.
Important econometric literature about conditions implying this property.

Ex: $\left| \begin{array}{l} Y = \varphi(Z) + U \\ Z = W + V \quad V \perp\!\!\!\perp W \quad \text{Characteristic function of } V \neq 0 \end{array} \right.$
Non identified model: section IV.

III-D Regularity

- Usual hypothesis: Hölder condition:

$$\varphi \in \mathcal{R}(K^*K)^{\frac{\beta}{2}}$$

  More generally $\varphi \in \mathcal{R}(g(K^*K))$

- Hilbert scale assumption
  L operator defining an Hilbert scale

$$\varphi \in \mathcal{D}(L^b)$$

$$K \sim L^{-a} \quad \exists \underline{c}, \bar{c} \quad \underline{c}\|L^{-a}\varphi\| \leq \|K\varphi\| \leq \bar{c}\|L^{-a}\varphi\|$$

III-E Estimation

(y_i, z_i, w_i) iid sample

- Tikhonov regularization/kernel estimation

$$\min\langle r - K\varphi, r - K\varphi\rangle + \alpha\langle\varphi,\varphi\rangle \Rightarrow \varphi_\alpha = (\alpha I + K^*K)^{-1}K^*r$$

$$\begin{aligned}
\alpha\varphi + \hat{K}^*\hat{K}\varphi &= \hat{K}^*\hat{r} \\
\hat{r} = \hat{E}(Y|W) &= \int y\hat{f}(y|w)dy \\
\hat{K} = \hat{E}(\varphi(Z)|W) &= \int \varphi(z)\hat{f}(z|w)dz \\
\hat{K}^* = \hat{E}(\psi(W)|Z) &= \int \psi(W)\hat{f}(w|z)
\end{aligned}$$

$\hat{f}$ estimation by kernel.

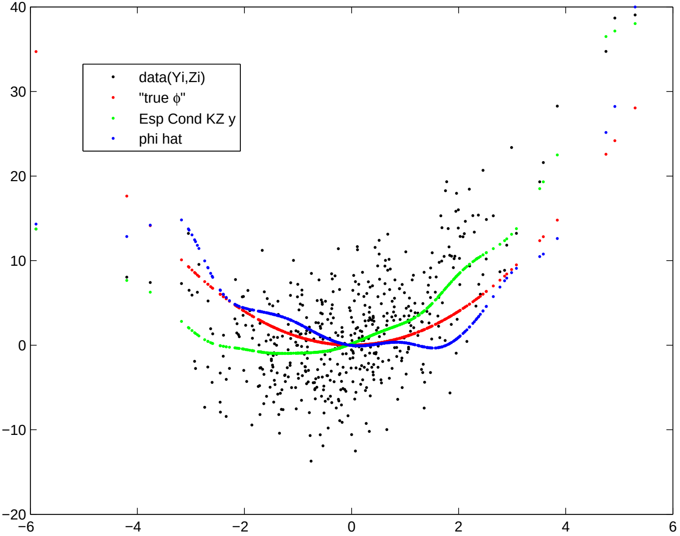
Under some approximation this estimation reduces to a matrix computation.

- Alternative: Sieve estimation $\hat{f}$ and φ approximated on a basis of functions.

Example:

$$\begin{aligned} Y &= Z^2 + U \quad U = \rho V + \varepsilon \\ Z &= W + V \end{aligned}$$

Toulouse School of Economics


data(Yi,Zi)
"true φ"
Esp Cond KZ y
phi hat
40
30
20
10
0
−10
−20
−6
−4
−2
0
2
4
6

- Ex: Instrumental variables

$$
\begin{aligned}
E(Y|W) &= E(\varphi(Z)|W) \\
r &= K\varphi
\end{aligned}
$$

  The same data are used for the estimation of r and K, the conditional expectation operator.

- Tikhonov regularisation:
  $\|\hat{r} - \hat{K}\varphi\|^2 \sim O_p(\delta) \quad \varphi \in \mathcal{R}(K^*K)^{\frac{\beta}{2}}$

$$\hat{\varphi}_\alpha = (\alpha I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\hat{r}$$

$$
\begin{aligned}
\hat{\varphi}_\alpha &= (\alpha I + \hat{K}^*\hat{K})^{-1}(\hat{K}^*\hat{r} - \hat{K}^*\hat{K}\varphi) \quad \mathrm{I} \\
&+ \left[(\alpha I + \hat{K}^*\hat{K})^{-1}\hat{K}^*\hat{K} - (\alpha I + K^*K)^{-1}K^*K)\right]\varphi \quad \mathrm{II} \\
&+ (\alpha I + K^*K)^{-1}K^*K\varphi - \varphi \quad \mathrm{III}
\end{aligned}
$$

I: variance $(\frac{\delta}{\alpha})$ III bias (α^β): I and III are usual.
II: due to the estimation of K

$$\begin{aligned} \mathrm{II} &= -\alpha(\alpha I + \hat{K}^*\hat{K})^{-1}(\hat{K}^*\hat{K} - K'K)(\alpha I + K^*K)^{-1}\varphi \\ &= -\alpha(\alpha I + \hat{K}^*\hat{K})^{-1}\hat{K}^*(\hat{K} - K)(\alpha I + K^*K)^{-1}\varphi \\ &\quad -\alpha(\alpha I + \hat{K}^*\hat{K})^{-1}(\hat{K}^* - \hat{K}^*)K(\alpha I + K^*K)^{-1}\varphi \end{aligned}$$

$$-\alpha(\alpha I + K^*K)^{-1}\varphi = \varphi - (\alpha I + K^*K)^{-1}K^*K\varphi$$

Two components:

- convergence rate of $\hat{K} - K$ or $\hat{K}^* - K^*$
- regularisation bias $\alpha(\alpha I + K^* K)^{-1}\varphi$

If the model is identified (K one ton one) this regularisation bias goes to 0 at the speed α^β and this contribute to control the second term. In general case (identified + good assumptions in the estimation of K and K^*) the second term is negligable and the rate is

$$O_p\left(\frac{\delta}{\alpha} + \alpha^\beta\right)$$

If α optimal rate $\delta^{\frac{\beta}{\beta+1}}$ In the IV case: $\delta = n^{-\frac{2s}{2s+q}}$
s regularity of $E(Y|W)$
q: clim W

$$\Rightarrow \qquad \text{rate of } \|\hat{\varphi}_\alpha - \varphi\|^2 - O_p\left(n^{-\frac{2s}{2s+q} \times \frac{\beta}{\beta+1}}\right).$$

Under more assumptions $\beta = \frac{b}{a}$ $\quad b$ smoothness of φ $\quad a=$ smoothness of $f(Z|W)$

$$s = a + b \quad E(Y|W) = E(\varphi(Z)|W)$$

$$\Rightarrow \text{ rate } = n^{-\frac{2b}{a+b+1}} \; (q = 1)$$

Selection of the regularisation parameter
Empirical rules

$$\min \frac{1}{\alpha}\|\hat{r} - \hat{K}\hat{\varphi}_\alpha\|^2 \text{ or} \min \|\hat{\varphi}^\alpha\|^2\|\hat{r} - \hat{K}\hat{\varphi}_\alpha\|^2$$

Extension to IV models of the cross validation approach.

Many extensions:

- other regularisation (iterative)
- penalization by the norm of derivatives

Asymptotic normality: α fixed (bias) functional result.

$$\alpha \to 0 : \frac{\sqrt{n}\langle \hat{\varphi}^\alpha - \varphi, \psi\rangle}{\sigma\|\alpha(\alpha I + K^*K)^{-1}\psi\|} \to N(0,1)$$

- Non identified model

$$\mathcal{N}(K) \neq \{0\} \quad \varphi = \varphi_0 + \varphi_1 \quad \begin{array}{l} \varphi_0 \in \mathcal{N}(K) \\ \varphi_1 \perp \mathcal{N}(K) \end{array}$$

$\hat{\varphi}_\alpha \to \varphi_1$ under usual assumptions if K is given at a speed depending on the regularity of φ_1.
If K is estimated the regularisation bias $\alpha(\alpha I + K^*K)^{-1}\varphi$ does not go to 0 and the elimination of II require more assumptions on $\|\hat{K} - K\|$ or $\|\hat{K}^* - K^*\|$.

- Reduction of the curse of dimensionality
  - Additive model

$$Y = \varphi_1(Z_1) + \varphi_2(Z_2) + U \quad E(U|W) = 0$$

  - Partially linear model

$$Y = \varphi(Z) + X'\beta + U \quad E(U|W) = 0$$

Main question: rate of convergence of $\hat{\beta}$. Under which condition $\hat{\beta}$ converges at $\sqrt{n}$ speed?

- Transformation models:
    - $\varphi(Y) = X'\beta + U \quad E(U|X,W) = 0$
      normalisation on β.
      ex: $Y \in [0,1] \quad Y$ probability
      $Y = F(X'\beta + U) \quad \varphi = F^{-1}$.
    - More generally:

$$\varphi(Y) = \psi(Z) + X'\beta + U \quad E(U|X,W) = 0$$

Y, Z "endogenous".
Application: Two sided market
Two equations linking two proportions.

$$\varphi \in L^2_Y \ \psi \in L^2_Z \ E(\psi) = 0$$
$$K(\varphi, \psi) = E(\varphi - \psi|W, X)$$

- Panel data models. $\begin{cases} t \text{ time} \\ i \text{ individual} \end{cases}$

$$Y_{ti} = \varphi(Z_{ti}) + \eta_i + U_{ti}$$

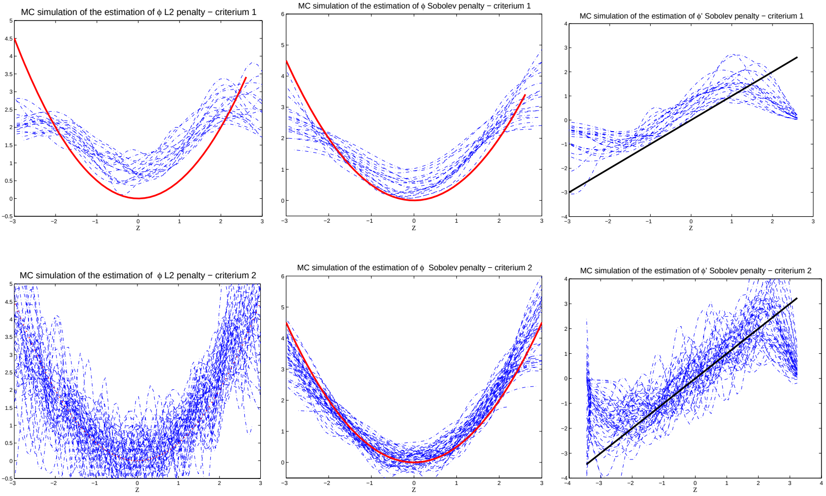
$$\eta_i \text{ heterogeneity effect } .E(U_{ti}(|W) = 0$$

$$Y_{ti} - Y_{t_1 i} = \varphi(Z_{ti}) - \varphi(Z_{t_1 i}) + U_{ti} - U_{t-1i}$$

$$\varphi \in \mathcal{E} \quad E(\varphi(Z_t) - \varphi(Z_{t-1})|W = E(Y_{ti} - Y_{t-1i}|W).$$

$$K\varphi = E(\varphi(Z_t) - \varphi(Z_{t-1})|W).$$

φ is identified up to an additive constant. Estimation of the derivative of φ using a Tikhonov method with a penalty of the L^2 norm of the derivative.


MC simulation of the estimation of ϕ L2 penalty − criterium 1
MC simulation of the estimation of ϕ Sobolev penalty − criterium 1
MC simulation of the estimation of ϕ' Sobolev penalty − criterium 1
MC simulation of the estimation of ϕ L2 penalty − criterium 2
MC simulation of the estimation of ϕ Sobolev penalty − criterium 2
MC simulation of the estimation of ϕ' Sobolev penalty − criterium 2
Z

- Discrete endogenous variable models (classification problem)

$$Y^* = \varphi(Z) + U \quad E(U|W) = 0 \quad (Y, Z, W) \text{ random vector}$$
$$Y^* = E(Y^*|W) - \varepsilon = E(\varphi(Z)|W) - \varepsilon$$
$$Y^* \text{ unobservable but } Y = \mathbb{1}(Y^* \geq 0)$$

Object of interest: φ (example of Z treatment).

Assumption: ε is independent of W with a known distribution characterised by the cdf G.

$\Rightarrow$ We may identified $p(W) = P(Y = 1|W)$

$$\Rightarrow \quad p(W) = G(E(\varphi(Z)|W)) \quad \underbrace{G^{-1}(p(W))}_{r} = \underbrace{E(\varphi(Z)|W)}_{K\varphi}$$

- Quantile models: $(Y, Z) \in \mathbb{R} \times \mathbb{R}^p$ random element.

  Object of interest: Conditional distribution of $Y|Z$ describes by its quantile function

$$
\begin{array}{l}
Y = \varphi(Z, U) \\
\varphi(Z, .) \uparrow \quad U \sim U[0, 1] \quad Z \perp\!\!\!\perp U \\
U = F(Y|Z)
\end{array}
$$

- Quantile Instrumental variables models.

$$\begin{aligned}
&(Y, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q \\
&Y = \varphi(Z, U) \\
&\varphi(Z, .) \uparrow, U \sim U[0,1] \quad U \perp\!\!\!\perp W
\end{aligned}$$

$\Rightarrow \varphi$ characterized by a non linear integral equation:

$$F(y, z|w) = \frac{\partial}{\partial z} P(Y \leq y, Z \leq z | W = w)$$

$$\int F(\varphi(z, u), z|w) dz = 1 \quad T(\varphi) = 0$$

F may be estimated non parametrically

- Examples of interest: computation of the (robust) frontier.

  Z level of production of a firm
  Y cost
  U inefficiency component
  $\varphi(z, o)$ minimum cost for producting z: efficiency frontier.
  $\varphi(z, \alpha)$ α "small": Robust frontier

  Quantile I.V. estimation of the cost frontier under endogeneity of the production level.

- other example:

$$
\begin{aligned}
&Y = \text{Duration} \\
&U \sim Exp(1) \quad U \perp\!\!\!\perp W \\
&\varphi = \text{inverse of integrated hazard}
\end{aligned}
$$

  Duration model with endogenous co-factors (not time dependent).

- Identification (unicity of φ)

  Local identification (unicity in a neighborhood of the true value). Dependence condition between Z and W given U (conditional completness).

  $$E\left(a(Z,U)|W,U\right)) = 0 \Rightarrow a = 0 \quad \text{(a.s)}$$

  Global condition: Same type of condition for a family of perturbations of the true data generating process.

Estimation: Landweber recursive algorithm.

$$\varphi_{K+1} = \varphi_K + \widehat{T'^{*}_{\varphi_K}}(\hat{T}(\varphi_K))$$

$\rightarrow$ choice of $K : \min K \|T(\hat{\varphi}_K)\|^2$

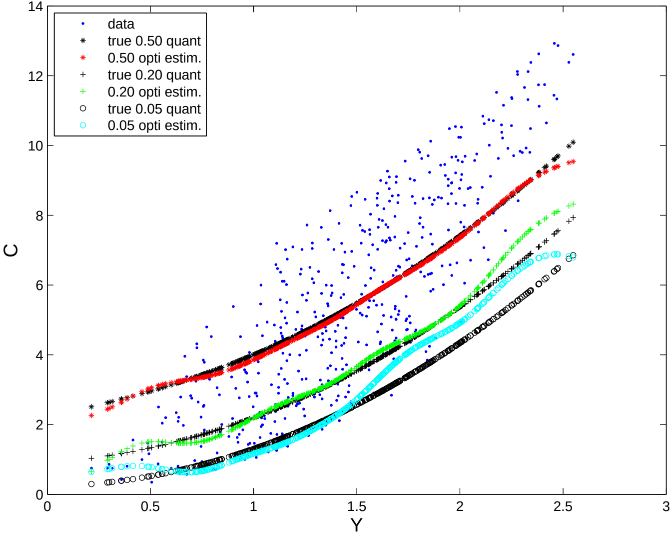
- Particular case: Separability assumption.

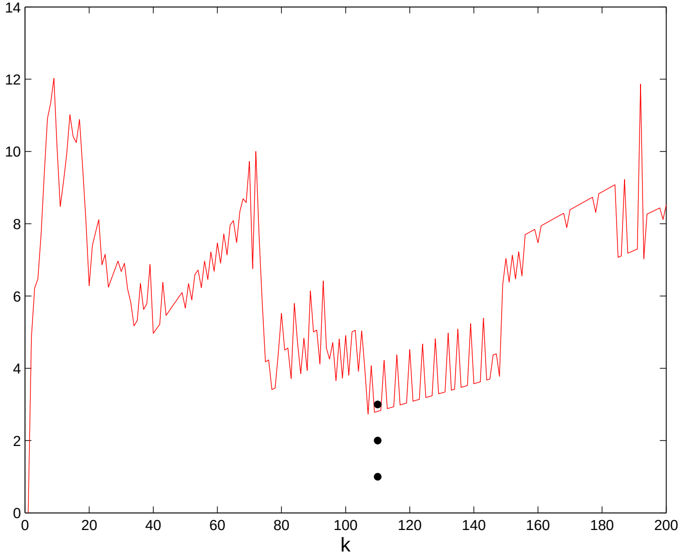
$$Y = \varphi_1(Z) + \varphi_2(U) + U$$

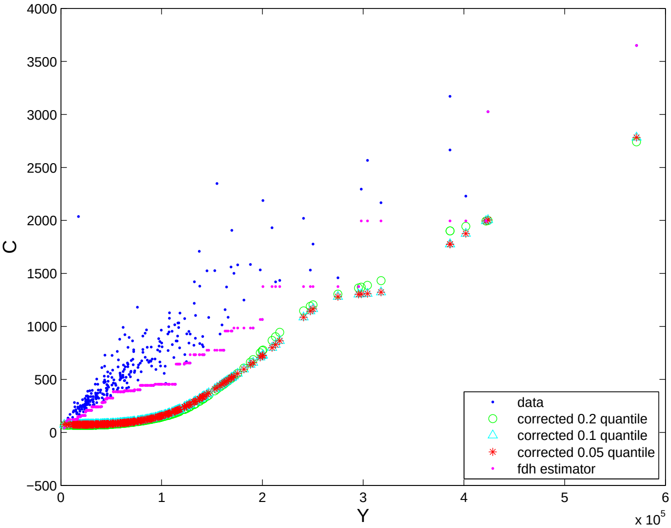
still with $U \perp\!\!\!\perp W$

Interest of $U \perp\!\!\!\perp W$ relatively to $E(\varphi_2(U)|W) = 0$ Identification and estimation with "few" instruments.

ex: $Z \in \mathbb{R} \quad W \in \{0, 1\}$.
Continuous endogenous variables and discrete instruments.


data
true 0.50 quant
0.50 opti estim.
true 0.20 quant
0.20 opti estim.
true 0.05 quant
0.05 opti estim.
C
Y
0
2
4
6
8
10
12
14
0
0.5
1
1.5
2
2.5
3


0
2
4
6
8
10
12
14
0
20
40
60
80
100
120
140
160
180
200
k


4000
3500
3000
2500
2000
1500
1000
500
0
−500
C
0
1
2
3
4
5
6
Y
x 10^5
data
corrected 0.2 quantile
corrected 0.1 quantile
corrected 0.05 quantile
fdh estimator

- additively separable cases:
  Y_t stochastic process.
  Two filtrations $\mathcal{Z}_t$ and $\mathcal{W}_t$
  IV decomposition: $Y_t = \wedge_t + U_t$
  such that $\wedge_t$ is $\mathcal{F}_t$ adapted

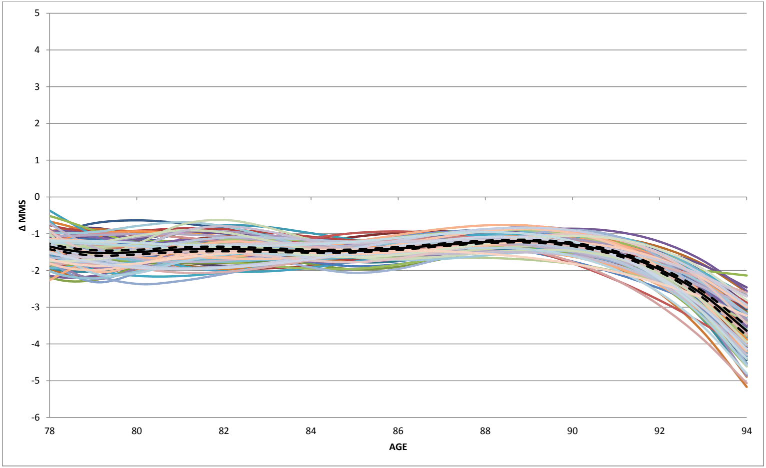
$$E(U_t - U_s|\mathcal{W}_s) = 0 \quad 0 \leq s \leq t$$

  Then: If $dY_t = h_t\, dt + dM_t$ decomposition wrt $\mathcal{W}_t$

$$\Rightarrow \wedge_t = \int_0^t \lambda_s ds \text{ where } \lambda_t \text{ is the solution of}$$
$$h_t = E(\lambda_t|\mathcal{W}_t)$$

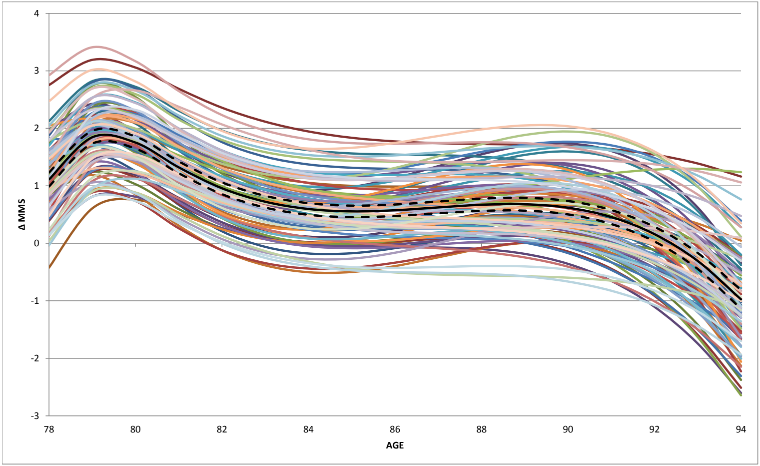
  Application:

  macroeconomic time series models
  diffusion process with endogenous explanatory variable in the drift
  empirical application to Alzheimer decease.



ATE Distribution Estimates: Exogeneity Assumption

Toulouse School of Economics



ATE Distribution Estimates: Endogeneity Assumption

- Non separable dynamic models
  Objective:
  - duration models with endogenous co-factors, possibly time dependent
  - diffusion depending on endogenous variables in the volatility

Y_t process. two filtrations $\mathcal{Z}_t$ and $\mathcal{W}_t$

Model:
$\exists\ \Phi_t$ increasing sequence of stopping times relatively to $\mathcal{Z}_t$

$$Y_{\Phi_t} = U_t \quad (U_s)_s \perp\!\!\!\perp (\mathcal{W}_t)_t$$

Then:

$$\begin{aligned} \mathcal{X}_t &= \mathcal{Z}_t \vee \mathcal{Y}_t \\ dY_t &= k_t dt + dE_t \quad wrt \mathcal{X}_t \\ U_t &= H_t + M_t \end{aligned}$$

$\Rightarrow$

$$E_{\mathcal{Z}} \left[\int_0^{\Phi_t} E(k_s | \mathcal{Y}_s \vee \mathcal{W}_s) ds \right] = H_t$$

Sequence of non linear integral equations.

- Regression case: $Y \in \mathcal{F}$ $Z \in \mathcal{E}$ Hilbert spaces

$$Y = \Pi Z + U$$

π: linear operator U random noise in $\mathcal{F}$

$$Cov(Z, U) = 0$$

iid observations $(y_i, z_i)_{i=1,\ldots,n}$
ex:

  - $\mathcal{F} = \mathbb{R}$ $Y =$ demand to a service in a geographical zone (Hospital, Post office, Bank...)
    Z: density of population in the zone.
  - Reduced form of a social network model
    $u \in \{1, 2, \ldots\}$ population of individuals
    $\begin{cases} Z(u) \\ Y(u) \end{cases}$ variables

$$Y(u) = f\left((Y(v))_{v \neq u}, (Z(v))_v\right) + \varepsilon(u)$$

    f restricted by a network structure (possibly unknown).
    Structural model.

Resolution of this model:

$$Y = g(Z) + U \; g \text{ linear}$$

- Time dependent variables. In a geographical zone:

$$\begin{aligned} Y(t) &= \text{electricity consumption at time } t \\ Z(s) &= \text{temperature at time } s \\ Y(t) &= \int Z(s)\pi(t,s)ds + U(t) \end{aligned}$$

  Π possibly constrained (triangular, convolution).
- Estimation under exogeneity assumption.

$$min \sum_{i=1}^{n} \underbrace{\|y_i - \Pi z_i\|^2}_{\text{norm in } \mathcal{F}} + \alpha \underbrace{\|\Pi\|^2}_{\substack{\text{Hilbert Schmidt norm} \\ \|\Pi\|^2 = tr\Pi\Pi^*}}$$

Easy computation:

$$\hat{\Pi}_\alpha^* \psi = \frac{1}{n} \sum_{i=1}^{n} \Big(\langle y_i, \psi \rangle - \langle z_i, \hat{\Pi}_\alpha^* \psi \rangle \Big)$$

$\langle z_i, \hat{\Pi}_\alpha^* \psi \rangle$ solution of a linear systems.

$$(\alpha I + M)v = Mw \quad M = \left(\frac{\langle z_i, z_j \rangle}{n} \right)_{ij}$$

$$v = \Big(\langle z_i, \hat{\Pi}_\alpha^* \psi \rangle \Big)_i \qquad w = (\langle y_i, \psi \rangle)_i$$

- Rate of convergence, asymptotic normality for fixed α
  Source condition: $\exists R / \Pi^* = r(V_Z) R$
  r analytical function

  Ex: Hölder condition $r(V_Z) = V_Z^\beta \ \beta > 1 \quad \Pi^* = V_z^\beta R$
  $\rightarrow$ rate $\|\hat{\Pi}_\alpha^* - \Pi\|_{HS}^2 = O_p \left(\frac{1}{n\alpha} + \alpha^{\beta \wedge 2} \right)$
  Asymptotic normality for fixed α and tests.
  Data driven selection of α.

Toulouse School of Economics

- Discretized variables.

  z_i, y_i not observed but z_i^m, y_i^m are observed such that $\|z_i^m - z_i\| \sim O(f(m))$ $\|y_i^m - y_i\| = O(f(m))$

  same asymptotic if $\frac{f(m)}{\alpha n} = O\left(\alpha^{\beta \wedge 2}\right)$

- Endogeneity of Z.

  Instruments W $Cov(U, W) = 0$

  $\Rightarrow Cov(Y, W) = \Pi Cov(Z, W)$

  $\Rightarrow \hat{\Pi}_\alpha^* = (\alpha I + \hat{C}_{ZW} \hat{C}_{WZ})^{-1} \hat{C}_{ZW} \hat{C}_{WY}$

  Same type of asymptotic results.

Definitions

- Game models very important in economic literature.
  - Auctions and procurements
  - Oligopoly competition
  - Contract theory (labor contract, principal agent models...)
  - Many applications in industrial economy

  Other applications: psychology, political sciences...
- Base line model for games with incomplete information.
  One game. N player $j = 1, .., N$
  Each player has a private signal: ξ_j

$$\xi_j \text{ iid } \xi_j \sim F \quad (forexample)$$

each player knows ξ_j and F and plays

$$x_j = \sigma(\xi_j, F)$$

σ: strategy (e.g. Nash equilibrium).
The statistician observes $x_1, ...x_n$ (or the winner only), knows σ and want estimate F.

- Example: First price auction:
  ξ_j private value of the object x_j price proposed by the player
  gain of the player $\begin{cases} 0 \text{ if } x_j \text{ is not the highest price} \\ \xi_j - x_j \text{ if } \xi_j \text{ is the highest price.} \end{cases}$
  $\Rightarrow$ Nash equilibrium

$$x_j = \xi_j - \frac{\int_0^{\xi_j} F^{N-1}(u)du}{F^{N-1}(\xi_j)} \quad \text{if } \ \xi_j \in [0,1]$$

- many extensions of this model.
  - ξ_j non iid
  - explanatory variables

$$\xi_j | Z \sim F^z \text{ conditional distribution}$$

    σ depends on W
  - σ not completely given.

- Data: $(l, j) \begin{cases} l \text{ game} & l = 1, ..., L \\ j \text{ player} & j = 1, ..., N_l \end{cases}$

  Parameters: F^Z, θ

$$x_{jl} = \sigma(\xi_{j_l}, F^{z_l}, w_{z_l}, \theta)$$

  x_{jl}, z_l and w_{jl} observed

- Baseline model: $X = \sigma(\xi, F)\ \xi \sim F$

$$\sigma(., F) \text{ increasing } \xi \in [\underline{\xi}, \overline{\xi}]$$

$$G(x) = P(X \leq x) = P(\xi \leq \sigma_F^{-1}(x)) = F \circ \sigma_F^{-1}(x)$$

$$\boxed{G(x) = F \circ \sigma_F^{-1}(x)}$$

  Inverse problem (non linear in general).

$$\boxed{G^{-1}(u) = \sigma_F \circ F^{-1}(u)}$$

  quantile form.

- other examples:
    - Third price auction:

$$X = \xi + \frac{1}{\lambda} ln \left\{ 1 + \frac{\lambda}{N-2} \frac{F(\xi)}{F(\xi)} \right\}$$

    - Contract models (simplest case)

$$X = \xi + \frac{F(\xi)}{f(\xi)} \quad f: \text{ density of } \xi$$

- Analysis in terms of distribution function.

$$C = \varphi \circ \sigma_F^{-1} = T(F)$$

Non linear inverse problem. T given.
G estimated by the empirical distribution function (possibly smoothed).
Numerical regularized solution of $\hat{C} = T(F)$ (iterative method).
Local analysis: Frechet derivative of T:

$$T'_F(\tilde{F})(\xi) = \alpha(\xi)\tilde{F}(\xi) + \beta(\xi)\int_0^\xi F^{N-2}(u)\tilde{F}(u)du$$

$$\alpha(\xi) = \frac{N\int_0^\xi F^{N-1}(u)du}{F^w(\xi)} \quad \beta(\xi) = \frac{-(N-1)}{F^{N-1}(\xi)}$$

- Analysis in terms of quantile functions.
  Interest: linear inverse problem:

$$G^{-1}(u) = \frac{N}{\alpha^N} \int_0^\alpha u^{N-1} F^{-1}(u) du$$
$$H(u) = \frac{N}{\alpha^N} \int_0^\alpha u^{N-1} \psi(u) du$$

H and ψ quantile functions of the bids and of the prices.

$$\varphi = F^{-1} \quad r = G^{-1} \frac{\alpha^N}{N}$$

$r = K\varphi$

Several solutions:

- estimation smooth of r

$$
\begin{aligned}
\Rightarrow \quad & r'(\alpha) = \alpha^{N-1} F^{-1}(\alpha) \\
\Rightarrow \quad & \hat{F}^{-1}(\alpha) = \tfrac{1}{\alpha^{N-1}} \left(\hat{G}^{-1}(\alpha) \tfrac{\alpha^N}{N} \right)'
\end{aligned}
$$

  rate of estimation of $\hat{F}^{-1}(\alpha)$= rate of estimation of $G^{-1}(\alpha)'$.

- G^{-1} estimated by the empirical quantile function and regularised inversion. (under boundary restriction and shape constraint).

Toulouse School of Economics

- The hazard rate games models.

$$X = \sigma(\xi, F) = a(\xi, \frac{F}{f}(\xi)) \quad f = F'$$

$$\begin{aligned} \Rightarrow G^{-1}(\alpha) &= a(F^{-1}(\alpha), \alpha F^{-1'}(\alpha)) \quad \text{a given} \\ r &= a(\varphi, \alpha\varphi') \end{aligned}$$

Differential equation. Well-posed inverse problem.

Example: Third price auction. Solution :

$$\varphi(\alpha) = \frac{1}{\lambda} \ln \frac{1}{\alpha^N} \int_0^\alpha N u^{N-1} e^{\lambda G^{-1}(u)} du$$

Asymptotic theory: Application of the behavior of functions of order statistics.

General results: $F^{-1}, F, F^{-1'}$ and f converges at the $\sqrt{n}$ rate to Gaussian processes.

X-A GMM

- One of the more popular approaches in econometrics

$$X \sim F \quad \text{iid sample}$$

$$\exists h / E^F(h(X,\theta)) = 0 \quad dim\theta \text{ finite}$$

Over identified (F constrained by this equation) if $dimh > \theta$

Extensions to non iid sampling or to conditional moments.

Different approaches in the literature.

  - $\min \| \frac{1}{n} \Sigma h(x_i, \theta) \|_V^2 \left.\begin{matrix} \text{GMM} \\ \text{CUE} \end{matrix}\right\}$ V may depend on θ
  - resolution of $F \min d(F - \hat{F}_n) / E^F(h) = 0 \}$GEL

Reformulation of the problem.
Parameter:

(θ, f) f density of F wrt π

$f \in L^2_\Pi$ such that $\int h(x,\theta)f(x)d\pi = 0$

$$\hat{r}_n(t) = \frac{1}{n}\sum_{i=1}^{n} k(x_i, t) \quad k : X \times T \to \mathbb{R} \quad (T, \mathcal{T}, P)$$

$$\hat{r}(t) = \int k(x,t)f(x)\pi(dx) + U(t)$$

$$E(U) = 0 \quad Var(U) = \frac{1}{n}\Sigma$$

$$\begin{cases} \hat{r} = Kf + U \\ f \in \mathcal{N}(R_\theta) \; R_\theta(f) = \begin{pmatrix} \displaystyle\int f(x)\pi(dx) = 1 \\ \int h(x,\theta)f(x)\pi(dx) = 0 \end{pmatrix} \end{cases}$$

possible extension to more general problems.

X-B

- Bayesian analysis
  - linear models without constraints.

    Different notations:
    - Sampling model

      $$y^\delta | x \sim \mathcal{N}(Kx, \delta\Sigma)$$

      $$x \in \mathcal{X} \quad y \in y \quad K : \mathcal{X} \to Y \quad \Sigma \text{ trace class operator}$$
    - Prior probability

      $$x|\alpha \sim N(x_0, \frac{\delta}{\alpha}\Omega_o)$$

      (interesting case: $L = \Omega_o^{-\frac{1}{2s}}$ defining an Hilbert scale and $K \sim L^{-a}$).

Toulouse School of Economics

Hypothesis: $\mathcal{R}(K\Omega_o^{\frac{1}{2}}) \subset \mathcal{D}(\Sigma^{-1})$

$$\Rightarrow x|y^\delta, \alpha \sim N\left(A(y^\delta - Kx_0) + x_0, \frac{\delta}{\alpha}(\Omega_0 - AK\Omega_o)\right)$$

$A = \Omega_o^{\frac{1}{2}}(\alpha I + B^* B)^{-1}(\Sigma^{-\frac{1}{2}} B)^*$ continuous
$B = \Sigma^{-\frac{1}{2}} K \Omega_o^{\frac{1}{2}}$

Frequentist analysis of the posterior mean and of the posterior distribution.

Adaptive selection of α by an empirical bayes method.

$$\left.\begin{array}{l} \mu|\alpha \\ y^\delta|\mu, \alpha \end{array}\right\} \rightarrow y^\delta|\alpha$$

$y^\delta|\alpha$: dominated model
$\hat{\alpha}$: max integrated likelihood or positive mode.

X-C Bayesian GMM

$$\begin{aligned} \theta \in \Theta &\subset \mathbb{R}^k \\ f|\theta &\sim N(f_{o\theta}, \Omega_{o\theta}) \\ \hat{r}|f,\theta &\sim N(Kf, \frac{1}{n}\Sigma) \quad \Sigma \text{ estimated.} \end{aligned}$$

Prior such that

$$Prob\left(\int fd\pi = 1 \text{ and } \int h(x,\theta)f(x)\pi(x) = 0\right) = 1$$

$$\Omega_{o\theta}^{\frac{1}{2}}1 = 0 \quad \Omega_{o\theta}^{\frac{1}{2}}h = 0$$

$\rightarrow$ Posterior analysis:

Under some regularity assumption

- $\mu(\theta\ r_n) \propto \mu(\theta) exp - \frac{1}{2}\sum_{j=0}^{\infty}\frac{\langle\sqrt{n}(r_n - Kf_{o\theta}), \Sigma^{-\frac{1}{2}}\psi_\theta\rangle^2}{1+n\lambda_j}$

  $(l_{j\theta}, \psi_{j\theta}, p_{i\theta})$ SVD of $\Sigma^{-\frac{1}{2}}K\Omega_{o\theta}^{\frac{1}{2}}$

  λ_j eigen values of $\Omega_{o\theta}$ (independent of θ in some cases).

$$\text{Several properties} \begin{cases} \text{diffuse prior} \\ \text{frequentist asymptotic properties} \end{cases}$$

- $f|\theta, r_n \sim N(\hat{f}_{o\theta}, \hat{\Omega}_\theta)$

$$\hat{f}_\theta = f_{o\theta} + A(r_n - K f_{o\theta}) \quad \hat{\Omega}_\theta = \Omega_o - AK\Omega_{o\theta}$$

$$A = \Omega_{o\theta}^{\frac{1}{2}} \left(\frac{1}{n} I + \Omega_{o\theta}^{\frac{1}{2}} K^* \Sigma^{-1} K \Omega_{o\theta}^{\frac{1}{2}} \right)^{-1} \left(\Sigma^{-1} K \Omega_{o\theta} \right)^*$$

f generated by the posterior satisfies the moment conditions.

The posterior of f given θ revises the prior on f except in the directions of the moments constraints.

- Presentation biased in direction of my own works and of co-authors and PhD students.
- Many others works:
- Radom coefficient models

$$Y = X'\theta \quad \theta \text{ random} \qquad X \perp\!\!\!\perp \theta$$

  Object of interest: distribution of θ
  Lead to
  - deconvolution on the sphere
  - Random transform
- Functional GMM

$$E(h(\varphi, X)|Z) = 0 \quad \varphi \text{ function}$$

- Type II integral equations:

$$\varphi(z) - E(\varphi(Z_1)|Z_2 = z) = E(Y|Z_z)$$

- Data driven selection of the regularisation parameter and oracle inegalities.
- More complex functional equations

$$\begin{aligned} Y_t &= \varphi(Z_t) + U_t + E \\ E(U_t - U_s|\mathcal{W}_s) &= 0 \end{aligned}$$

Z_t diffusion conditional to the filtration generated by Y, Z and W.

$$dZ_t = \mu_t dt + \sigma_t dB_t$$

$$E\left(\frac{dY_t}{dt}|\mathcal{W}_t\right) = E\left(\frac{\partial\varphi}{\partial z}\mu_t + \frac{1}{2}\frac{\partial^2\varphi}{\partial z^2}\sigma_t^2|\mathcal{W}_t\right)$$

- Development of softwares with a choice of approaches (choice of the regularisation and of this estimation)
- Models on networks or on Riemmanian manifolds (use of the Laplacian).