Estimation des paramètres à partir des moments conditionnels en présence de censure

Valentin Patilea

INSA-IRMAR Rennes & CREST-ENSAI
Work in progress!
Work in progress!

Joint research project with
- Pascal Lavergne (SFU)
- Olivier Lopez (CREST-ENSAI et IRMAR)
A general statistical problem

The observations: independent copies of

\[ Z = (Y', X')' \in \mathbb{R}^{d+q}, \quad d \geq 1, \quad q \geq .1 \]
A general statistical problem

- The observations: independent copies of

\[ Z = (Y', X')' \in \mathbb{R}^{d+q}, \quad d \geq 1, \quad q \geq 0.1 \]

- Let \( g(z, \theta) = (g^{(1)}(z, \theta), \ldots, g^{(r)}(z, \theta))' - r - \text{vector valued function} \)
A general statistical problem

- The observations: independent copies of

$$Z = (Y', X')' \in \mathbb{R}^{d+q}, \quad d \geq 1, \quad q \geq .1$$

- Let $$g(z, \theta) = (g^{(1)}(z, \theta), \ldots, g^{(r)}(z, \theta))' - r-\text{vector valued function}$$

- The unknown parameter $$\theta \in \Theta \subset \mathbb{R}^p, \quad p \geq 1$$
A general statistical problem

The observations: independent copies of

\[ Z = (Y', X')' \in \mathbb{R}^{d+q}, \quad d \geq 1, \; q \geq 1 \]

Let \( g(z, \theta) = (g^{(1)}(z, \theta), \ldots, g^{(r)}(z, \theta))' - r-\text{vector valued function} \)

The unknown parameter \( \theta \in \Theta \subset \mathbb{R}^p, \; p \geq 1 \)

Assumption: for a unique \( \theta_0 \in \Theta \)

\[ \mathbb{E} [g(Z, \theta_0)|X] = 0 \quad \text{a.s.} \] (1)
A general statistical problem

The observations: independent copies of

\[ Z = (Y', X')' \in \mathbb{R}^{d+q}, \quad d \geq 1, \quad q \geq .1 \]

Let \( g(z, \theta) = (g^{(1)}(z, \theta), \ldots, g^{(r)}(z, \theta))' - r - \) vector valued function

The unknown parameter \( \theta \in \Theta \subset \mathbb{R}^p, \ p \geq 1 \)

Assumption: for a unique \( \theta_0 \in \Theta \)

\[ \mathbb{E}[g(Z, \theta_0)|X] = 0 \quad \text{a.s.} \quad (1) \]

The problem: estimate \( \theta_0 \) identified by the conditional moment restrictions (1)
A classical approach: GMM (or GEE)

- Reformulate a set of unconditional moment restrictions from (1)

Consider a $m \times r$ matrix $A(X, \theta)$ of "instruments" and $\rho(Z, \theta) = A(X, \theta)g(Z, \theta)$.

Clearly, $E[\rho(Z, \theta_0)] = E\left\{ A(X, \theta_0)E[g(Z, \theta_0) | X] \right\} = 0$

Under "suitable" conditions ($m \geq r$, ...), one has the identification $E[\rho(Z, \theta)] = 0 \Rightarrow \theta = \theta_0$ (2)

The associated GMM estimator $\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \rho(Z_i, \theta)' \hat{\Omega} \sum_{j=1}^n \rho(Z_j, \theta)$

where $\hat{\Omega}$ is a $m \times m$ positive semi-definite (random) matrix.
A classical approach: GMM (or GEE)

- Reformulate a set of **unconditional** moment restrictions from (1)

- Consider a \( m \times r \) matrix \( A(X, \theta) \) of “instruments” and

\[
\rho(Z, \theta) = A(X, \theta)g(Z, \theta)
\]
A classical approach: GMM (or GEE)

- Reformulate a set of unconditional moment restrictions from (1)
- Consider a \( m \times r \) matrix \( A(X, \theta) \) of "instruments" and
  \[ \rho(Z, \theta) = A(X, \theta)g(Z, \theta) \]
- Clearly,
  \[ \mathbb{E} [\rho(Z, \theta_0)] = \mathbb{E} \{ A(X, \theta_0)\mathbb{E} [g(Z, \theta_0)|X]\} = 0 \]
A classical approach: GMM (or GEE)

- Reformulate a set of unconditional moment restrictions from (1)

- Consider a $m \times r$ matrix $A(X, \theta)$ of "instruments" and
  \[
  \rho(Z, \theta) = A(X, \theta)g(Z, \theta)
  \]

- Clearly,
  \[
  \mathbb{E} [\rho(Z, \theta_0)] = \mathbb{E} \{A(X, \theta_0)\mathbb{E} [g(Z, \theta_0)|X]\} = 0
  \]

- Under "suitable" conditions ($m \geq r$, ...), one has the identification
  \[
  \mathbb{E} [\rho(Z, \theta)] = 0 \quad \Rightarrow \quad \theta = \theta_0 \quad (2)
  \]
A classical approach: GMM (or GEE)

- Reformulate a set of **unconditional** moment restrictions from (1)

- Consider a $m \times r$ matrix $A(X, \theta)$ of “instruments” and

  $$\rho(Z, \theta) = A(X, \theta)g(Z, \theta)$$

- Clearly,

  $$\mathbb{E} [\rho(Z, \theta_0)] = \mathbb{E} \{A(X, \theta_0)\mathbb{E} [g(Z, \theta_0)|X]\} = 0$$

- Under “suitable” conditions ($m \geq r$, ...), one has the identification

  $$\mathbb{E} [\rho(Z, \theta)] = 0 \Rightarrow \theta = \theta_0$$  \hspace{1cm} (2)

- The associated GMM estimator

  $$\hat{\theta} = \arg \min_{\theta} \left[ \sum_{i=1}^{n} \rho(Z_i, \theta) \right]' \hat{\Omega} \left[ \sum_{j=1}^{n} \rho(Z_j, \theta) \right] = \arg \min_{\theta} \left\| \sum_{i=1}^{n} \rho(Z_i, \theta) \right\|_{\hat{\Omega}}^2$$

  where $\hat{\Omega}$ is a $m \times m$ positive semi-definite (random) matrix.
The classical “challenges”
The classical “challenges”

- Find a set of instruments that guarantee the identification condition
The classical “challenges”

- Find a set of instruments that guarantee the identification condition
- Find the (asymptotically) ‘optimal’ matrix $\hat{\Omega}$ for given instruments – easy

In general, a nonparametric estimate of the optimal instruments is required.
The classical “challenges”

- Find a set of instruments that guarantee the identification condition
- Find the (asymptotically) ‘optimal’ matrix $\hat{\Omega}$ for given instruments – easy
- Find the (asymptotically) ‘optimal’ set of instruments – combine with the optimal $\hat{\Omega}$ to get an (asymptotically) efficient GMM estimator

$A(X, \theta_0) = E\left[\partial g(Z, \theta_0) / \partial \theta | X\right] E\left[g(Z, \theta_0) g(Z, \theta_0)^T | X\right]^{-1}$

In general, a nonparametric estimate of the optimal instruments is required.
The classical “challenges”

- Find a set of instruments that guarantee the identification condition
- Find the (asymptotically) ‘optimal’ matrix $\hat{\Omega}$ for given instruments – easy
- Find the (asymptotically) ‘optimal’ set of instruments – combine with the optimal $\hat{\Omega}$ to get an (asymptotically) efficient GMM estimator

The (asymptotically) ‘optimal’ instruments

$$A(X, \theta_0) = \mathbb{E} \left[ \frac{\partial g(Z, \theta_0)}{\partial \theta'} | X \right]' \mathbb{E} \left[ g(Z, \theta_0) g(Z, \theta_0)' | X \right]^{-1}$$

In general, a nonparametric estimate of the optimal instruments is required
Estimating the optimal instruments is not very common in applications.
Estimating the optimal instruments is not very common in applications

Quite often, with real data, one uses some “convenient” instruments
• Estimating the optimal instruments is not very common in applications

• Quite often, with real data, one uses some “convenient” instruments

• In simple linear regression \( g(Z, \theta) = Y - \theta^{(1)} - \theta^{(2)} X \) and the optimal instruments are \( A(X, \theta) = (1, X)' \)
Estimating the optimal instruments is not very common in applications.

Quite often, with real data, one uses some “convenient” instruments.

In simple linear regression \( g(Z, \theta) = Y - \theta^{(1)} - \theta^{(2)} X \) and the optimal instruments are \( A(X, \theta) = (1, X)' \).

There is a big temptation to use such simple instruments with nonlinear regression!

In general, this produces inconsistent GMM estimators.
Estimating the optimal instruments is not very common in applications

Quite often, with real data, one uses some “convenient” instruments

In simple linear regression \( g(Z, \theta) = Y - \theta^{(1)} - \theta^{(2)} X \) and the optimal instruments are \( A(X, \theta) = (1, X)' \)

There is a big temptation to use such simple instruments with nonlinear regression!
In general, this produces inconsistent GMM estimators

Inconsistency can appear even with optimal instruments!
Suppose $\mathbb{E}[Y \mid X] = \theta_0^2 X + \theta_0 X^2$ with $\theta_0 = 5/4$ and $V(Y \mid X) \equiv \sigma^2$. The optimal instrument is $A(X, \theta_0) = 2\theta_0 X + X^2$. If $X$ is $N(-1, 1)$ distributed, the unconditional moment restriction $\mathbb{E}[(Y - \theta_0^2 X - \theta_0 X^2)A(X, \theta_0)] = 0$ admits the solutions $\theta = \pm 5/4$. Since $\theta_0$ is unknown in practice, one may try to directly solve $\mathbb{E}[(Y - \theta_0^2 X - \theta_0 X^2)A(X, \theta_0)] = 0$. The solutions are $\theta = \pm 5/4$ and $\theta = -3$ when $X$ is $N(1, 1)$ distributed.
Suppose $\mathbb{E}[Y \mid X] = \theta_0^2 X + \theta_0 X^2$ with $\theta_0 = \frac{5}{4}$ and $V(Y \mid X) \equiv \sigma^2$.

The optimal instrument is $A(X, \theta_0) = 2\theta_0 X + X^2$. 

Since $\theta_0$ is unknown in practice, one may try to directly solve $\mathbb{E}[(Y - \theta_0^2 X - \theta_0 X^2) A(X, \theta_0)] = 0$. The solutions are $\theta = \pm \frac{5}{4}$ and $\theta = -\frac{3}{4}$ when $X$ is $N(1, 1)$ distributed.
Inconsistent GMM optimal instruments – example

Suppose $\mathbb{E}[Y \mid X] = \theta_0^2 X + \theta_0 X^2$ with $\theta_0 = 5/4$ and $V(Y \mid X) \equiv \sigma^2$

The optimal instrument is $A(X, \theta_0) = 2\theta_0 X + X^2$

If $X$ is $N(-1, 1)$ distributed, the unconditional moment restriction

$$\mathbb{E}[(Y - \theta^2 X - \theta X^2)A(X, \theta_0)] = 0$$

admits the solutions $\theta = \pm 5/4$. 
Suppose $\mathbb{E}[Y \mid X] = \theta_0^2 X + \theta_0 X^2$ with $\theta_0 = 5/4$ and $V(Y \mid X) \equiv \sigma^2$

The optimal instrument is $A(X, \theta_0) = 2\theta_0 X + X^2$

If $X$ is $N(-1, 1)$ distributed, the unconditional moment restriction

$$\mathbb{E}[(Y - \theta^2 X - \theta X^2)A(X, \theta_0)] = 0$$

admits the solutions $\theta = \pm 5/4$.

Since $\theta_0$ is unknown in practice, one may try to directly solve

$$\mathbb{E}[(Y - \theta^2 X - \theta X^2)A(X, \theta)] = 0.$$  

The solutions are $\theta = \pm 5/4$ and $\theta = -3$ when $X$ is $N(1, 1)$ distributed.
Alternative approach – Empirical likelihood


**Advantages:**
- No need to estimate the optimal instruments
- The global identification problem can be avoided
- The estimator $\hat{\theta}$ is asymptotically efficient
Alternative approach – Empirical likelihood


**Advantages:**
- No need to estimate the optimal instruments
- The global identification problem can be avoided
- The estimator $\hat{\theta}$ is asymptotically efficient

**Drawbacks:**
- The bandwidth should vanish at some "good rate" – no hint how to choose it in practice.
- The calculation of the EL-based estimators is not simple
Alternative approach – Empirical likelihood


Advantages:
- No need to estimate the optimal instruments
- The global identification problem can be avoided
- The estimator $\hat{\theta}$ is asymptotically efficient

Drawbacks:
- The bandwidth should vanish at some “good rate” – no hint how to choose it in practice.
- The calculation of the EL-based estimators is not simple
Suppose that the moment restriction identifying the parameter $\theta$ is

$$E[g(Y, X, \theta_0) | X] = 0 \text{ a.s.}$$

with $Y \in \mathbb{R}$.
Suppose that the moment restriction identifying the parameter $\theta$ is

$$\mathbb{E}[g(Y, X, \theta_0) | X] = 0 \text{ a.s.}$$

with $Y \in \mathbb{R}$.

**BUT** instead of observing $Y_1, \ldots, Y_n$, we observe independent copies of

$$T = Y \wedge C \quad \text{and} \quad \delta = 1 \{Y \leq C\},$$

where $C$ is a random censoring.
Suppose that the moment restriction identifying the parameter $\theta$ is

$$\mathbb{E} [g(Y, X, \theta_0) | X] = 0 \quad \text{a.s.}$$

with $Y \in \mathbb{R}$.

**BUT** instead of observing $Y_1, \ldots, Y_n$, we observe independent copies of

$$T = Y \wedge C \quad \text{and} \quad \delta = 1\{Y \leq C\},$$

where $C$ is a random censoring.

A new function $\phi(T, \delta, X, \theta)$ should be defined such that

$$\mathbb{E} [\phi(T, \delta, X, \theta) | X] = 0 \quad \text{a.s.} \iff \theta = \theta_0.$$
Conditional moments with random censored data

- Suppose that the moment restriction identifying the parameter \( \theta \) is
  \[
  \mathbb{E}[g(Y, X, \theta_0) | X] = 0 \quad \text{a.s.}
  \]
  with \( Y \in \mathbb{R} \).

- **BUT** instead of observing \( Y_1, \ldots, Y_n \), we observe independent copies of
  \[
  T = Y \wedge C \quad \text{and} \quad \delta = 1\{Y \leq C\},
  \]
  where \( C \) is a random censoring.

- A new function \( \phi(T, \delta, X, \theta) \) should be defined such that
  \[
  \mathbb{E}[\phi(T, \delta, X, \theta) | X] = 0 \quad \text{a.s.} \iff \theta = \theta_0.
  \]

- **First, some identification assumptions are needed!**
Suppose that

Y and C are independent
Suppose that
- $Y$ and $C$ are independent
- $\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y)$.


Identification assumptions

Suppose that

- $Y$ and $C$ are independent
- $\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y)$.


Let $\phi(T, \delta, X, \theta) = \delta[1 - G(T-)]^{-1}g(Y, X, \theta)$, where $G(t) = \mathbb{P}(C \leq t)$. 

V. Patilea (INSA-IRMAR)
Identification assumptions

Suppose that

- $Y$ and $C$ are independent
- $\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y)$.


Let $\phi(T, \delta, X, \theta) = \delta [1 - G(T -)]^{-1} g(Y, X, \theta)$, where $G(t) = \mathbb{P}(C \leq t)$.

Indeed,

$$
\mathbb{E}[\phi(T, \delta, X, \theta) \mid X] = \mathbb{E}[\delta [1 - G(Y -)]^{-1} g(Y, X, \theta) \mid X]
$$
Identification assumptions

- Suppose that
  - $Y$ and $C$ are independent
  - $\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y)$.


- Let $\phi(T, \delta, X, \theta) = \delta[1 - G(T-)]^{-1}g(Y, X, \theta)$, where $G(t) = \mathbb{P}(C \leq t)$.

- Indeed,

$$
\mathbb{E}[\phi(T, \delta, X, \theta) \mid X] = \mathbb{E}[\delta[1 - G(Y-)]^{-1}g(Y, X, \theta) \mid X] \\
= \mathbb{E}[[1 - G(Y-)]^{-1}\mathbb{E}\{\delta \mid Y, X\}g(Y, X, \theta) \mid X]
$$
Identification assumptions

- Suppose that
  - $Y$ and $C$ are independent
  - $\mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y)$.
  

- Let $\phi(T, \delta, X, \theta) = \delta[1 - G(T -)]^{-1}g(Y, X, \theta)$, where $G(t) = \mathbb{P}(C \leq t)$.

- Indeed,

\[
\mathbb{E}[\phi(T, \delta, X, \theta) \mid X] = \mathbb{E}[\delta[1 - G(Y -)]^{-1}g(Y, X, \theta) \mid X] \\
= \mathbb{E}[[1 - G(Y -)]^{-1}\mathbb{E}\{\delta \mid Y, X\}g(Y, X, \theta) \mid X] \\
= \mathbb{E}[[1 - G(Y -)]^{-1}\mathbb{P}(Y \leq C \mid Y)g(Y, X, \theta) \mid X]
\]
Identification assumptions

- Suppose that
  - \( Y \) and \( C \) are independent
  - \( \mathbb{P}(Y \leq C \mid X, Y) = \mathbb{P}(Y \leq C \mid Y) \).
  

- Let \( \phi(T, \delta, X, \theta) = \delta[1 - G(T-)]^{-1}g(Y, X, \theta) \), where
  \( G(t) = \mathbb{P}(C \leq t) \).

- Indeed,

\[
\mathbb{E}[\phi(T, \delta, X, \theta) \mid X] = \mathbb{E}[\delta[1 - G(Y-)]^{-1}g(Y, X, \theta) \mid X] \\
= \mathbb{E}[[1 - G(Y-)]^{-1}\mathbb{E}\{\delta \mid Y, X\}g(Y, X, \theta) \mid X] \\
= \mathbb{E}[[1 - G(Y-)]^{-1}\mathbb{P}(Y \leq C \mid Y)g(Y, X, \theta) \mid X] \\
= \mathbb{E}[g(Y, X, \theta) \mid X].
\]
When a more general relationship between the censoring time and the covariates is required:

\[ Y \text{ and } C \text{ are independent conditionally on } X \]
Alternative identification assumptions

- When a more general relationship between the censoring time and the covariates is required:
  
  $Y$ and $C$ are independent conditionally on $X$

- In this case

  $$\phi(T, \delta, X, \theta) = \frac{\delta}{1 - G(T - | X)} g(Y, X, \theta),$$

  where $G(t | X) = \mathbb{P}(C \leq t | X)$. 

Alternative identification assumptions

When a more general relationship between the censoring time and the covariates is required:

$Y$ and $C$ are independent conditionally on $X$

In this case

$$
\phi(T, \delta, X, \theta) = \frac{\delta}{1 - G(T - | X)} g(Y, X, \theta),
$$

where $G(t | X) = \mathbb{P}(C \leq t | X)$.

Indeed,

$$
\mathbb{E}[\phi(T, \delta, X, \theta) | X] = \mathbb{E}[g(Y, X, \theta) | X].
$$
Alternative identification assumptions

- When a more general relationship between the censoring time and the covariates is required:

  \[ Y \text{ and } C \text{ are independent conditionally on } X \]

- In this case

  \[
  \phi(T, \delta, X, \theta) = \frac{\delta}{1 - G(T - \mid X)} g(Y, X, \theta),
  \]

  where \( G(t \mid X) = \mathbb{P}(C \leq t \mid X) \).

- Indeed,

  \[
  \mathbb{E}[\phi(T, \delta, X, \theta) \mid X] = \mathbb{E}[g(Y, X, \theta) \mid X].
  \]

- This framework is more complicated when \( G(T - \mid X) \) has to be estimated nonparametrically!
Hereafter, consider Stute’s identification assumptions.
Hereafter, consider Stute’s identification assumptions.

Consider, for the moment, that $G(t) = \mathbb{P}(C \leq t)$ is given.
Hereafter, consider Stute’s identification assumptions.

Consider, for the moment, that $G(t) = P(C \leq t)$ is given.

Assume that the conditioning variables $X$ admit a density $f(\cdot)$. 
Hereafter, consider Stute’s identification assumptions.

Consider, for the moment, that \( G(t) = \mathbb{P}(C \leq t) \) is given.

Assume that the conditioning variables \( X \) admit a density \( f(\cdot) \).

Extension to vectors \( X \) with continuous and discrete components is possible.
Consider the criterion

\[ M(\theta) = \mathbb{E} [\phi'(T, \delta, X, \theta) \mathbb{E} [\phi(T, \delta, X, \theta) | X] f(X)] \]

\[ = \mathbb{E} [\mathbb{E} [\phi'(T, \delta, X, \theta) | X] \mathbb{E} [\phi(T, \delta, X, \theta) | X] f(X)] \geq 0 \]
Consider the criterion

\[ M(\theta) = \mathbb{E} \left[ \phi'(T, \delta, X, \theta) \mathbb{E} \left[ \phi(T, \delta, X, \theta) | X \right] f(X) \right] \]
\[ = \mathbb{E} \left[ \mathbb{E} \left[ \phi'(T, \delta, X, \theta) | X \right] \mathbb{E} \left[ \phi(T, \delta, X, \theta) | X \right] f(X) \right] \geq 0 \]

We have

\[ M(\theta) = 0 \iff \theta = \theta_0 \]

that is the initial \textit{conditional} moment restriction is equivalent to the \textit{unconditional} moment condition defined by \( M(\cdot) \).
Consider the criterion

\[ M(\theta) = \mathbb{E} \left[ \phi'(T, \delta, X, \theta) \mathbb{E} [\phi(T, \delta, X, \theta) | X] f(X) \right] \]

\[ = \mathbb{E} \left[ \mathbb{E} [\phi'(T, \delta, X, \theta) | X] \mathbb{E} [\phi(T, \delta, X, \theta) | X] f(X) \right] \geq 0 \]

We have

\[ M(\theta) = 0 \iff \theta = \theta_0 \]

that is the initial *conditional* moment restriction is equivalent to the *unconditional* moment condition defined by \( M(\cdot) \).

More generally, for some positive definite matrix-valued map \( P(\cdot) \),

\[ M(\theta) = \mathbb{E} \left[ \{ P^{-1/2}(X)\phi(T, \delta, X, \theta)\} \mathbb{E} \left[ P^{-1/2}(X)\phi(T, \delta, X, \theta) | X \right] f(X) \right] \]
The idea

- Build a sample counterpart of $M(\theta)$ and minimize it w.r.t. $\theta$
The idea

- Build a sample counterpart of $M(\theta)$ and minimize it w.r.t. $\theta$
- Let $M_{n,h}(\theta)$ be an estimate of $M(\theta)$ defined as

$$\frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} \phi'_i(\theta) P_n^{-1/2}(X_i) P_n^{-1/2}(X_j) \phi_j(\theta) K_{ij}$$

where $\phi_i(\theta) = \phi(T_i, \delta_i, X_i, \theta)$

$$K_{ij} = K_{ij}(h) = \frac{1}{hq} K \left( \frac{X_i - X_j}{h} \right)$$

and $P_n(\cdot)$ is some positive-definite sample counterpart of $P(\cdot)$. 
The idea

- Build a sample counterpart of $M(\theta)$ and minimize it w.r.t. $\theta$
- Let $M_n,h(\theta)$ be an estimate of $M(\theta)$ defined as

$$
\frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} \phi_i'(\theta) P_n^{-1/2}(X_i) P_n^{-1/2}(X_j) \phi_j(\theta) K_{ij}
$$

where $\phi_i(\theta) = \phi(T_i, \delta_i, X_i, \theta)$

$$
K_{ij} = K_{ij}(h) = \frac{1}{h^q} K \left( \frac{X_i - X_j}{h} \right)
$$

and $P_n(\cdot)$ is some positive-definite sample counterpart of $P(\cdot)$.

- The smooth GMM estimator

$$
\tilde{\theta}_{n,h} = \arg \min_{\theta \in \Theta} M_n(\theta)
$$
Asymptotics – consistency (the case $G(\cdot)$ known)

- We show consistency *uniformly* in the bandwidth, *not necessarily vanishing*. 
We show consistency *uniformly* in the bandwidth, *not necessarily vanishing*.

No instruments are needed!
We show consistency \textit{uniformly} in the bandwidth, \textit{not necessarily vanishing}.

No instruments are needed! Why this should work?
Asymptotics – consistency (the case $G(\cdot)$ known)

- We show consistency *uniformly* in the bandwidth, *not necessarily vanishing*.
- No instruments are needed! Why this should work?
- Simplify $P_n(X) = I_d$. If $\mathcal{F}[l](\cdot)$ is the FT of a function $l(\cdot)$,

$$
\mathbb{E} M_{n,h} (\theta) = \frac{1}{2} \mathbb{E} \left[ \phi'_1(\theta) \phi_2(\theta) h^{-q} K \left( (X_1 - X_2)/h \right) \right]
$$
We show consistency uniformly in the bandwidth, not necessarily vanishing.

No instruments are needed! Why this should work?

Simplify $P_n(X) = ld$. If $\mathcal{F}[l](\cdot)$ is the FT of a function $l(\cdot)$,

$$
\mathbb{E} M_{n,h}(\theta)
= \frac{1}{2} \mathbb{E} \left[ \phi'_1(\theta) \phi_2(\theta) h^{-q} K \left( (X_1 - X_2)/h \right) \right]
= \frac{(2\pi)^{-q/2}}{2} \mathbb{E} \left[ \phi'_1(\theta) \phi_2(\theta) \int_{\mathbb{R}^q} \exp \left( -it(X_1 - X_2) \right) \mathcal{F}[K](ht) \, dt \right]
$$
Asymptotics – consistency (the case $G(\cdot)$ known)

We show consistency *uniformly* in the bandwidth, *not necessarily vanishing*.

No instruments are needed! Why this should work?

Simplify $P_n(X) = l_d$. If $\mathcal{F}[l](\cdot)$ is the FT of a function $l(\cdot)$,

\[
\mathbb{E}M_{n,h}(\theta) = \frac{1}{2} \mathbb{E} \left[ \phi_1'(\theta) \phi_2(\theta) h^{-q} K ((X_1 - X_2)/h) \right] = \frac{(2\pi)^{-q/2}}{2} \mathbb{E} \left[ \phi_1'(\theta) \phi_2(\theta) \int_{\mathbb{R}^q} \exp \left( -it(X_1 - X_2) \right) \mathcal{F}[K](ht) \, dt \right] = \frac{(2\pi)^{q/2}}{2} \sum_{k=1}^{r} \left\{ \int_{\mathbb{R}^q} \left| \mathcal{F} \left[ \mathbb{E}[\phi_1^{(k)}(\theta)|X = \cdot] f(\cdot) \right] (t) \right|^2 \mathcal{F}[K](ht) \, dt \right\}
\]
We show consistency \textit{uniformly} in the bandwidth, \textit{not necessarily vanishing}.

No instruments are needed! Why this should work?

Simplify $P_n(X) = Id$. If $\mathcal{F}[l](\cdot)$ is the FT of a function $l(\cdot)$,

$$
\mathbb{E}M_{n,h}(\theta) = \frac{1}{2} \mathbb{E} \left[ \phi'_1(\theta) \phi_2(\theta) h^{-q} K \left( \frac{(X_1 - X_2)}{h} \right) \right] \\
= \frac{(2\pi)^{-q/2}}{2} \mathbb{E} \left[ \phi'_1(\theta) \phi_2(\theta) \int_{\mathbb{R}^q} \exp \left( -it(X_1 - X_2) \right) \mathcal{F}[K](ht) \, dt \right] \\
= \frac{(2\pi)^{q/2}}{2} \sum_{k=1}^{r} \left\{ \int_{\mathbb{R}^q} \left| \mathcal{F} \left[ \mathbb{E}[\phi_1^{(k)}(\theta) \mid X = \cdot] f(\cdot) \right] (t) \right|^2 \mathcal{F}[K](ht) \, dt \right\}
$$

If $\mathcal{F}[K](\cdot)$ is strictly positive on $\mathbb{R}^q$, $\mathbb{E}M_{n,h}(\theta) \geq 0$ and

$$
\mathbb{E}M_{n,h}(\theta) = 0 \quad \text{iff} \quad \theta = \theta_0.
$$
The strictly positive $\mathcal{F}[K](\cdot)$ – fulfilled by products of the triangular, normal, Laplace, Cauchy, ... densities
The strictly positive $F[K](\cdot)$ – fulfilled by products of the triangular, normal, Laplace, Cauchy, ... densities

higher-order kernels are allowed
The strictly positive $F[K](\cdot)$ – fulfilled by products of the triangular, normal, Laplace, Cauchy, ... densities

higher-order kernels are allowed

**Theorem**

*For an i.i.d. sample and under ‘suitable’ assumptions,*

$$\tilde{\theta}_{n,h} - \theta_0 = o_P(1)$$

uniformly in $h \in \{1 \geq h > 0 : nh^{2q} \geq \ln n\}$. 
The assumptions for consistency ($G(\cdot)$ known)

- The parameter space $\Theta$ is compact

The families $G_k = \{ \phi_{(k)}(\cdot, \cdot, \cdot, \theta) : \theta \in \Theta \}$, $1 \leq k \leq r$, are VC-classes for an envelope $G$ with $\mathbb{E}G^2 < \infty$. 

V. Patilea (INSA-IRMAR)

MAS-SMAI 2008, Rennes

August, 2008 17 / 24
The assumptions for consistency ($G(\cdot)$ known)

- The parameter space $\Theta$ is compact
- $\theta_0 \in \Theta$ is unique satisfying the conditional moment restriction

\[ K(x) = \tilde{K}(x_1) \ldots \tilde{K}(x_q) \]

$\tilde{K}(\cdot)$ is a symmetric, squared-integrable, bounded function of bounded variation with strictly positive Fourier transform. The integral of $\tilde{K}(\cdot)$ equals one.

$\forall n$, $P_n(\cdot)$ is a $r \times r$ symmetric positive definite non-random matrix-function; there is a symmetric positive definite matrix function $P(\cdot)$ such that $\forall u$, $P_n(u) - P(u) = o(1)$.

Moreover, $0 < c_1 \leq \inf_u \lambda_{\min}(u) \leq \sup_u \lambda_{\max}(u) \leq c_2 < \infty$, where $\lambda_{\min}(u)$ and $\lambda_{\max}(u)$ denote the smallest and largest eigenvalue of $P(u)$ or $P_n(u)$.

The function $\sup_\theta \|E[\phi(T, \delta, X, \theta)]|_X=x\|_f(x)$ is in $L_1 \cap L_2$. For all $x$, the map $\theta \mapsto E[\phi(T, \delta, X, \theta)|X=x]$ is continuous.

The families $G_k = \{ \phi(k)(\cdot, \cdot, \cdot, \theta) : \theta \in \Theta \}$, $1 \leq k \leq r$, are VC-classes for an envelope $G$ with $E_G^2 < \infty$.  

V. Patilea (INSA-IRMAR)  
MAS-SMAI 2008, Rennes  
August, 2008  
17 / 24
The assumptions for consistency ($G(\cdot)$ known)

- The parameter space $\Theta$ is compact
- $\theta_0 \in \Theta$ is unique satisfying the conditional moment restriction
- $K(x) = \tilde{K}(x^{(1)}) \ldots \tilde{K}(x^{(q)})$ with $\tilde{K}(\cdot)$ a symmetric, squared-integrable, bounded function of bounded variation with strictly positive Fourier transform. The integral of $\tilde{K}(\cdot)$ equals one.
The assumptions for consistency \((G(\cdot) \text{ known})\)

- The parameter space \(\Theta\) is compact
- \(\theta_0 \in \Theta\) is unique satisfying the conditional moment restriction
- \(K(x) = \tilde{K}(x^{(1)}) \ldots \tilde{K}(x^{(q)})\) with \(\tilde{K}(\cdot)\) a symmetric, squared-integrable, bounded function of bounded variation with strictly positive Fourier transform. The integral of \(\tilde{K}(\cdot)\) equals one.
- \(\forall n, \ P_n(\cdot)\) is a \(r \times r\) symmetric positive definite non-random matrix-function; there is a symmetric positive definite matrix function \(P(\cdot)\) such that \(\forall u, \ P_n(u) - P(u) = o(1)\). Moreover,

\[
0 < c_1 \leq \inf_u \lambda_{\text{min}}(u) \leq \sup_u \lambda_{\text{max}}(u) \leq c_2 < \infty,
\]

where \(\lambda_{\text{min}}(u)\) [\(\lambda_{\text{max}}(u)\)] denote the smallest [largest] eigenvalue of \(P(u)\) or \(P_n(u)\).
The assumptions for consistency ($G(\cdot)$ known)

- The parameter space $\Theta$ is compact.
- $\theta_0 \in \Theta$ is unique satisfying the conditional moment restriction.
- $K(x) = \tilde{K}(x^{(1)}) \ldots \tilde{K}(x^{(q)})$ with $\tilde{K}(\cdot)$ a symmetric, squared-integrable, bounded function of bounded variation with strictly positive Fourier transform. The integral of $\tilde{K}(\cdot)$ equals one.
- $\forall n$, $P_n(\cdot)$ is a $r \times r$ symmetric positive definite non-random matrix-function; there is a symmetric positive definite matrix function $P(\cdot)$ such that $\forall u$, $P_n(u) - P(u) = o(1)$. Moreover,

$$0 < c_1 \leq \inf_u \lambda_{\text{min}}(u) \leq \sup_u \lambda_{\text{max}}(u) \leq c_2 < \infty,$$

where $\lambda_{\text{min}}(u)$ [$\lambda_{\text{max}}(u)$] denote the smallest [largest] eigenvalue of $P(u)$ or $P_n(u)$.

- The function $\sup_{\theta} \| \mathbb{E}[\phi(T, \delta, X, \theta) \mid X = x] \| f(x)$ is in $L^1 \cap L^2$. For all $x$, the map $\theta \mapsto \mathbb{E}[\phi(T, \delta, X, \theta) \mid X = x]$ is continuous.
The assumptions for consistency (\(G(\cdot)\) known)

- The parameter space \(\Theta\) is compact
- \(\theta_0 \in \Theta\) is unique satisfying the conditional moment restriction
- \(K(x) = \tilde{K}(x^{(1)}) \ldots \tilde{K}(x^{(q)})\) with \(\tilde{K}(\cdot)\) a symmetric, squared-integrable, bounded function of bounded variation with strictly positive Fourier transform. The integral of \(\tilde{K}(\cdot)\) equals one.
- \(\forall n, P_n(\cdot)\) is a \(r \times r\) symmetric positive definite non-random matrix-function; there is a symmetric positive definite matrix function \(P(\cdot)\) such that \(\forall u, P_n(u) - P(u) = o(1)\). Moreover,
  \[
  0 < c_1 \leq \inf_u \lambda_{\text{min}}(u) \leq \sup_u \lambda_{\text{max}}(u) \leq c_2 < \infty,
  \]
  where \(\lambda_{\text{min}}(u) [\lambda_{\text{max}}(u)]\) denote the smallest [largest] eigenvalue of \(P(u)\) or \(P_n(u)\).
- The function \(\sup_{\theta} \| \mathbb{E}[\phi(T, \delta, X, \theta) | X = x] \| f(x)\) is in \(L^1 \cap L^2\). For all \(x\), the map \(\theta \mapsto \mathbb{E}[\phi(T, \delta, X, \theta) | X = x]\) is continuous.
- The families \(G_k = \{\phi^{(k)}(\cdot, \cdot, \cdot, \theta) : \theta \in \Theta\}, 1 \leq k \leq r\), are VC-classes for an envelope \(G\) with \(\mathbb{E} G^2 < \infty\).
The general case – unknown $G(\cdot)$

- $G(\cdot)$ can be estimated nonparametrically by the Kaplan-Meier estimator.
The general case – **unknown** \( G(\cdot) \)

- \( G(\cdot) \) can be estimated nonparametrically by the Kaplan-Meier estimator.
- Define

\[
\hat{\phi}_{in}(\theta) = \frac{\delta_i}{1 - \hat{G}(T_i)} g(T_i, X_i, \theta)
\]
The general case – unknown $G(\cdot)$

- $G(\cdot)$ can be estimated nonparametrically by the Kaplan-Meier estimator.

Define

$$\hat{\phi}_{in}(\theta) = \frac{\delta_i}{1 - \hat{G}(T_i -)} g(T_i, X_i, \theta)$$

- Let $\hat{M}_{n,h}(\theta)$ be equal to

$$\frac{1}{2n(n - 1)} \sum_{1 \leq i \neq j \leq n} \hat{\phi}_{in}(\theta) P_n^{-1/2}(X_i) P_n^{-1/2}(X_j) \hat{\phi}_{jn}(\theta) K_{ij}$$
The general case – unknown $G(\cdot)$

- $G(\cdot)$ can be estimated nonparametrically by the Kaplan-Meier estimator.

Define

$$\hat{\phi}_{in}(\theta) = \frac{\delta_i}{1 - \hat{G}(T_i, X_i, \theta)} g(T_i, X_i, \theta)$$

- Let $\hat{M}_{n,h}(\theta)$ be equal to

$$\frac{1}{2n(n-1)} \sum_{1 \leq i \neq j \leq n} \hat{\phi}_{in}(\theta) P_n^{-1/2}(X_i) P_n^{-1/2}(X_j) \hat{\phi}_{jn}(\theta) K_{ij}$$

- The smooth GMM estimator becomes

$$\hat{\theta}_{n,h} = \arg \min_{\theta \in \Theta} \hat{M}_{n,h}(\theta)$$
Since
\[
\left| \frac{\delta_i}{1 - G(T_i^-)} - \frac{\delta_i}{1 - \hat{G}(T_i^-)} \right| \leq \sup_{1 \leq i \leq n} \left| \hat{G}(T_i^-) - G(T_i^-) \right|
\]
\[
\times \frac{\delta_i}{[1 - G(T_i^-)]^2} \frac{1 - G(T_i^-)}{1 - \hat{G}(T_i^-)}
\]
\[
= o_p(1) \frac{\delta_i}{[1 - G(T_i^-)]^2}
\]
Since
\[
\left| \frac{\delta_i}{1 - G(T_i -)} - \frac{\delta_i}{1 - \hat{G}(T_i -)} \right| \leq \sup_{1 \leq i \leq n} |\hat{G}(T_i -) - G(T_i -)| \times \frac{\delta_i}{[1 - G(T_i -)]^2} \frac{1 - G(T_i -)}{1 - \hat{G}(T_i -)} = o_p(1) \frac{\delta_i}{[1 - G(T_i -)]^2}
\]
under some additional integrability assumptions,
\[
\sup_{\theta \in \Theta} |\hat{M}_{n,h}(\theta) - M_{n,h}(\theta)| = o_p(1)
\]
uniformly in \( h \in \{1 \geq h > 0 : nh^{2q} \geq \ln n\} \).

The uniform-in bandwidth consistency of \( \hat{\theta}_{n,h} \) follows.
The general case with alternative identification assumptions

- \( G(\cdot \mid X) \) can be estimated nonparametrically by the conditional Kaplan-Meier estimator (Beran, 1981), but the properties of \( \hat{G}(\cdot \mid X) \) are more complicated.
The general case with alternative identification assumptions

- $G(\cdot \mid X)$ can be estimated nonparametrically by the conditional Kaplan-Meier estimator (Beran, 1981), but the properties of $\hat{G}(\cdot \mid X)$ are more complicated.

- Future research ...
Asymptotic normality – the case $G(\cdot)$ known

There exists zero-mean stochastic processes $A_{n,h}$, $h \in [0, 1]$, such that

$$\sqrt{n} (\tilde{\theta}_{n,h} - \theta_0) - A_{n,h} = o_p(1)$$

uniformly in $h \in \mathcal{H}_n = \{ 1 \geq h > 0 : Cn h^{4q} \geq n^\varepsilon \}$, for some constant $C > 0$ and some (arbitrarily) small $\varepsilon > 0$. 
Asymptotic normality – the case $G(\cdot)$ known

There exists zero-mean stochastic processes $A_{n,h}, h \in [0, 1]$, such that

$$\sqrt{n}(\tilde{\theta}_{n,h} - \theta_0) - A_{n,h} = o_p(1)$$

uniformly in $h \in \mathcal{H}_n = \{1 \geq h > 0 : Cnh^{4q} \geq n^\varepsilon\}$, for some constant $C > 0$ and some (arbitrarily) small $\varepsilon > 0$.

In particular, this asymptotic equivalence holds for fixed $h$. 
Asymptotic normality – the case $G(\cdot)$ known

There exists zero-mean stochastic processes $A_{n,h}$, $h \in [0, 1]$, such that

$$\sqrt{n}(\tilde{\theta}_{n,h} - \theta_0) - A_{n,h} = o_p(1)$$

uniformly in $h \in \mathcal{H}_n = \{1 \geq h > 0 : Cnh^{4q} \geq n^\varepsilon\}$, for some constant $C > 0$ and some (arbitrarily) small $\varepsilon > 0$.

In particular, this asymptotic equivalence holds for fixed $h$.

The asymptotic law of $\sqrt{n}(\tilde{\theta}_{n,h} - \theta_0)$ is obtained from the gaussian limit of $A_{n,h}$, $h \in [0, 1]$.
When the bandwidth is taken in a subset of $\mathcal{H}_n$ that decreases to zero, the asymptotic variance of $A_{n,h}$ is equal to the semiparametric efficiency bound for the moment condition

$$\mathbb{E}[\phi(T, \delta, X, \theta_0) \mid X] = 0 \text{ a.s.},$$

provided the limit $P(\cdot)$ of the weight matrices $P_n$ is suitably chosen.
Asymptotic normality – \( G(\cdot) \) known (cont’d)

- When the bandwidth is taken in a subset of \( \mathcal{H}_n \) that decreases to zero, the asymptotic variance of \( A_{n,h} \) is equal to the semiparametric efficiency bound for the moment condition

\[
\mathbb{E}[\phi(T, \delta, X, \theta_0) \mid X] = 0 \text{ a.s.},
\]

provided the limit \( P(\cdot) \) of the weight matrices \( P_n \) is suitably chosen.

- The ‘optimal’ choice for \( P(\cdot) \) is

\[
\text{Var} \left[ \phi(T, \delta, X, \theta_0) \mid X \right] f(X)
\]

which, in general, has to be estimated nonparametrically (e.g., kernel smoothing).
When the bandwidth is taken in a subset of $\mathcal{H}_n$ that decreases to zero, the asymptotic variance of $A_{n,h}$ is equal to the semiparametric efficiency bound for the moment condition

$$\mathbb{E}[\phi(T, \delta, X, \theta_0) | X] = 0 \ a.s.,$$

provided the limit $P(\cdot)$ of the weight matrices $P_n$ is suitably chosen.

The ‘optimal’ choice for $P(\cdot)$ is

$$\text{Var} [\phi(T, \delta, X, \theta_0) | X] f(X)$$

which, in general, has to be estimated nonparametrically (e.g., kernel smoothing).

The ‘optimal’ choice for $P(\cdot)$ does not involve $\mathbb{E}[\nabla_\theta \phi(T, \delta, X, \theta) | X]$ which is usually more difficult to estimate nonparametrically.
A new, simple, easy to implement estimation with conditional moment restrictions in the presence of random censoring is proposed.
A new, simple, easy to implement estimation with conditional moment restrictions in the presence of random censoring is proposed.

Test statistics for testing restriction on the parameters could be defined.
A new, simple, easy to implement estimation with conditional moment restrictions in the presence of random censoring is proposed.

Test statistics for testing restriction on the parameters could be defined.

It is based on kernel smoothing but it yields $\sqrt{n}$-consistent estimators even for fixed bandwidths.
A new, simple, easy to implement estimation with conditional moment restrictions in the presence of random censoring is proposed.

Test statistics for testing restriction on the parameters could be defined.

It is based on kernel smoothing but it yields $\sqrt{n}$--consistent estimators even for fixed bandwidths.

The results hold *uniformly* in the bandwidth – data-driven bandwidth selection is allowed.
A new, simple, easy to implement estimation with conditional moment restrictions in the presence of random censoring is proposed.

Test statistics for testing restriction on the parameters could be defined.

It is based on kernel smoothing but it yields $\sqrt{n}$–consistent estimators even for fixed bandwidths.

The results hold *uniformly* in the bandwidth – data-driven bandwidth selection is allowed.

It allows for non differentiable functions $\theta \mapsto g(z, \theta)$ in particular, it applies to quantile restrictions.
Using I.I.D. representations of Kaplan-Meier integrals and suitable bounds for

\[ \sup_{1 \leq i \leq n} \left| \frac{\delta_i}{1 - G(T_i)} - \frac{\delta_i}{1 - \hat{G}(T_i)} \right| \]

deduce the $\sqrt{n}$–convergence and the asymptotic normality for $\hat{\theta}_{n,h}$
Asymptotic normality – $G(\cdot)$ unknown

- Using I.I.D. representations of Kaplan-Meier integrals and suitable bounds for
  
  $$\sup_{1 \leq i \leq n} \left| \frac{\delta_i}{1 - G(T_i^-)} - \frac{\delta_i}{1 - \hat{G}(T_i^-)} \right|$$

  deduce the $\sqrt{n}$–convergence and the asymptotic normality for $\hat{\theta}_{n,h}$

- The variance of the feasible estimator $\hat{\theta}_{n,h}$ is modified due to the contribution of KM estimation