Semi-Nonparametric Modeling and Estimation

Herman J. Bierens

Pennsylvania State University
1 What are semi-nonparametric models?

Semi-nonparametric (SNP) models are models for which the functional form is only partly parametrized and where the non-specified part is an unknown function.

Examples:

- **SNP Discrete choice models:**
  \[
  \Pr[Y = 1|X] = F(\theta'X),
  \]
  where \( F(x) \) is an unknown distribution function, \( Y \in \{0, 1\} \) is the dependent variable and \( X \) is a vector of co-variates.
• SNP index regression models:

\[ Y = g(\theta'X) + U, \ E[U|X] = 0, \]

where \( g(x) \) is an unknown monotonic increasing response function.

If \( Y \) has support \( \mathbb{R} \) then without loss of generality we may assume that

\[ g(x) = \ln \left( \frac{F(x)}{1 - F(x)} \right), \]

where \( F(x) \) is an unknown distribution function.
SNP mixed proportional hazard models:
The conditional survival function of a duration $Y$ takes the form

$$
\Pr [Y > t|X] = E \left[ \exp \left( -V. \exp(\theta'X) \int_0^t \lambda(\tau|\alpha) d\tau \right) \right| X]
= \int_0^\infty \exp \left( -v. \exp(\theta'X) \int_0^t \lambda(\tau|\alpha) d\tau \right) dG(v),
$$

where $X$ is a vector of covariates, $\exp(\theta'X)$ is the systematic hazard, $\lambda(t|\alpha)$ is the baseline hazard, for example the Weibull hazard

$$
\lambda(t|\alpha) = \alpha.t^{\alpha-1}, \; \alpha > 0,
$$

and $V$ represent unobserved heterogeneity with unknown distribution function $G(v)$. 
SNP first-price auction models without binding reservation price.

The equilibrium bid function takes the form

\[ B = V - \int_0^V \frac{F(x|X)^{I-1} dx}{F(V|X)^{I-1}} \]

where \( B \) is a bid, \( X \) is a vector of auction-specific covariates, \( I \) is the number of bidders, possibly dependent on \( X \), and \( V \) is a random drawing from the conditional value distribution \( F(v|X) \).

Suppose that

\[ \ln V = \theta' X + U \]

where \( U \) is independent of \( X \). Then

\[ F(v|X) = G(v \cdot \exp(-\theta' X)) \]

where \( G \) is the unknown distribution function of \( \exp(U) \).
In all these examples the unknown function involved takes the form of a distribution function.

Therefore, the following questions arise

- (a) How to model these distribution functions or their corresponding density functions in a flexible way;
- (b) How to estimate them consistently.

The answers are:

- (a) Use Hilbert space theory.
- (b) Use sieve estimation.

I will discuss the latter first.
2 Sieve estimation

2.1 General conditions

Consider a model that involves an unknown Euclidean parameter vector $\theta_0 \in \mathbb{R}^p$ and an unknown absolutely continuous distribution function $F_0$ with density $f_0$ and support

$$X = \{x \in \mathbb{R} : f_0(x) > 0\}.$$

The parameter vector $\theta_0$ is contained in a compact set $\Theta \subset \mathbb{R}^p$, and the distribution function $F_0$ is contained in a compact metric space $\mathcal{F}$ of distribution functions, endowed with the "sup" metric

$$\sup_{x \in X} |F_1(x) - F_2(x)|.$$
Suppose that the pair \((\theta_0, F_0)\) is identified by the minimum of a continuous real function \(Q(\theta, F)\) on \(\Theta \times \mathcal{F}\):

\[
(\theta_0, F_0) = \arg \min_{(\theta, F) \in \Theta \times \mathcal{F}} Q(\theta, F)
\]

is unique,

\[
\overline{Q}(\theta, F) \text{ is continuous on } \Theta \times \mathcal{F}
\]

Let \(\hat{Q}_N(\theta, F)\) be the sample counterpart of \(Q(\theta, F)\), where \(N\) is the sample size, satisfying

\[
p \lim_{N \to \infty} \sup_{(\theta, F) \in \Theta \times \mathcal{F}} \left| \hat{Q}_N(\theta, F) - Q(\theta, F) \right| = 0
\]

Denote

\[
\left(\hat{\theta}_N, \hat{F}_N\right) = \arg \min_{(\theta, F) \in \Theta \times \mathcal{F}} \hat{Q}_N(\theta, F)
\]

Then similar to standard M-estimation theory,

\[
p \lim_{N \to \infty} \hat{\theta}_N = \theta_0, \quad p \lim_{N \to \infty} \sup_{x \in X} \left| \hat{F}_N(x) - F_0(x) \right| = 0.
\]

However, the problem is that the computation of \(\hat{F}_N\) is not feasible!
2.2 Sieve estimators

Suppose that it is possible to construct an increasing sequence \( \{F_n\}_{n=1}^{\infty} \) of (sieve) subspaces of \( \mathcal{F} \) such that for each \( n \geq 1 \) the computation of
\[
\left( \hat{\theta}_{n,N}, \hat{F}_{n,N} \right) = \arg \min_{(\theta,F) \in \Theta \times F_n} Q_N(\theta, F)
\]
is feasible.

If \( \{F_n\}_{n=1}^{\infty} \) is dense in \( \mathcal{F} \), i.e.,
\[
\mathcal{F} = \overline{\bigcup_{n=1}^{\infty} F_n},
\]
where the bar denoted the closure, then
\[
p \lim_{N \to \infty} \hat{\theta}_{n,N,N} = \theta_0, \ p \lim_{N \to \infty} \sup_{x \in \mathcal{X}} \left| \hat{F}_{n,N,N}(x) - F_0(x) \right| = 0.
\]
for any subsequence \( n_N \to \infty \) as \( N \to \infty \).

Note that the condition \( \mathcal{F} = \overline{\bigcup_{n=1}^{\infty} F_n} \) is equivalent to:
\[
\forall F \in \mathcal{F} \& \forall n \exists F_n \in F_n : \lim_{n \to \infty} \sup_{x \in \mathcal{X}} \left| F_n(x) - F(x) \right| = 0.
\]
2.3 Asymptotic normality?

Under what conditions is \( \hat{\theta}_{n,N} \) asymptotically normally distributed, i.e.,

\[
\sqrt{N} \left( \hat{\theta}_{n,N} - \theta_0 \right) \overset{d}{\to} N_p [0, \Sigma]
\]

There is some literature on this problem. See


However, in practice these conditions are difficult to verify, and they often assume implicitly that asymptotic normality holds.
Therefore, suppose that

\[ F_0 \in \bigcup_{n=1}^{\infty} \mathcal{F}_n \]

Then there exists a smallest \( n \) such that

\[ F_0 \in \mathcal{F}_n \]

This smallest \( n \) can be estimated consistently via an information criterion.

Given a consistent estimator of \( n \), the model becomes completely parametric, so that asymptotic normality can be derived in the standard way.
Questions:

• How to construct a compact metric space $\mathcal{F}$ of distribution functions;
• How to construct feasible sieve spaces $\mathcal{F}_n$.

The answers are again: Use Hilbert space theory.
3 Hilbert spaces of functions

Let $w(x)$ be a density on $\mathbb{R}$.

Consider the space $L^2(w)$ of Borel measurable real function $f(x)$ on $\mathbb{R}$ satisfying

$$\int f(x)^2 w(x) \, dx < \infty.$$ 

Endow the space $L^2(w)$ with the inner product

$$\langle f, g \rangle = \int f(x) g(x) w(x) \, dx$$

and associated norm

$$\|f\| = \sqrt{\langle f, f \rangle} = \sqrt{\int f(x)^2 w(x) \, dx}$$

and metric

$$\|f - g\| = \sqrt{\int (f(x) - g(x))^2 w(x) \, dx}$$
Then $L^2(w)$ is a Hilbert space:

A Hilbert space $\mathcal{H}$ is a vector space endowed with an inner product and associated norm and metric such that every Cauchy sequence in $\mathcal{H}$ has a limit in $\mathcal{H}$.

In particular, for any sequence $f_n \in L^2(w)$ satisfying

$$\lim_{\min(m,k) \to \infty} \|f_m - f_k\| = 0$$

(which makes $f_n$ a Cauchy sequence) there exists an $f \in L^2(w)$ such that

$$\lim_{n \to \infty} \|f_n - f\| = 0.$$
3.1 Complete orthonormal sequences

Let \( \{ \varphi_j(x) \}_{j=0}^{\infty} \) be an orthonormal sequence in \( L^2(w) \):

\[
\langle \varphi_i, \varphi_j \rangle = \int \varphi_i(x) \varphi_j(x) w(x) dx = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}
\]

Without loss of generality we may assume that

\( \varphi_0(x) \equiv 1 \).

Let \( f_n(x) \) be the projection of \( f(x) \) on \( \{ \varphi_j(x) \}_{j=0}^{n} \):

\[
f_n(x) = \sum_{j=0}^{n} \gamma_j \varphi_j(x), \text{ where } ||f - f_n||^2 \text{ is minimal.}
\]

Then

\[
\gamma_j = \langle f, \varphi_j \rangle = \int f(x) \varphi_j(x) w(x) dx, \quad \sum_{j=0}^{\infty} \gamma_j^2 < \infty.
\]

The \( \gamma_j \)'s involved are called the Fourier coefficients of \( f(x) \).
An orthonormal sequence \( \{\varphi_j(x)\}_{j=0}^{\infty} \) in \( L^2(w) \) is called complete if for arbitrary \( f \in L^2(w) \),

\[
\lim_{n \to \infty} \| f - f_n \| = 0
\]

where

\[
f_n(x) = \sum_{j=0}^{n} \gamma_j \varphi_j(x) \quad \text{with} \quad \gamma_j = \langle f, \varphi_j \rangle
\]

Then the equality

\[
f(x) = \sum_{j=0}^{\infty} \gamma_j \varphi_j(x)
\]

holds for all \( x \) in a set \( B \) satisfying

\[
\int_B w(x)dx = 1
\]
3.2 Examples of complete orthonormal sequences

3.2.1 Hermite polynomials

In the case
\[ w(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi} \]
the Hermite polynomials form a complete orthonormal sequence in the corresponding Hilbert space \( L^2(w) \).

Hermite polynomials \( \varphi_k(x) \) on \( \mathbb{R} \) can be generated recursively by
\[
\sqrt{k + 1}\varphi_{k+1}(x) - x \cdot \varphi_k(x) + \sqrt{k}\varphi_{k-1}(x) = 0, \quad k \geq 1,
\]
starting from
\[
\varphi_0(x) = 1, \quad \varphi_1(x) = x.
\]
3.2.2 Legendre polynomials

In the case that $w(u)$ is the uniform density on $[0, 1]$, 
\[ w(u) = I (0 \leq u \leq 1) \]
the Legendre polynomials form a complete orthonormal sequence in the corresponding Hilbert space $L^2(w)$.

Legendre polynomials $\varphi_k(u)$ on $[0, 1]$ can be generated recursively by

\[
\begin{align*}
\frac{(k + 1) / 2}{\sqrt{2k + 3\sqrt{2k + 1}}} \varphi_{k+1}(u) + (0.5 - u) \cdot \varphi_k(u) \\
+ \frac{k/2}{\sqrt{2k + 1}\sqrt{2k - 1}} \varphi_{k-1}(u) = 0, \ k \geq 1,
\end{align*}
\]
starting from
\[ \varphi_0(u) = 1, \varphi_1(u) = \sqrt{3} (2u - 1). \]
3.2.3 Cosine and Fourier series

Other complete orthonormal sequences in the case that \( w(u) \) is the uniform density on \([0, 1]\),

\[
w(u) = I (0 \leq u \leq 1),
\]

are the cosine sequence

\[
\varphi_k(u) = \begin{cases} 
1 & \text{for } k = 0 \\
\sqrt{2} \cos(k\pi u) & \text{for } k \geq 1
\end{cases}
\]

and the Fourier series

\[
\varphi_k(u) = \begin{cases} 
1 & \text{for } k = 0 \\
\sqrt{2} \sin(2k\pi u) & \text{for odd } k \geq 1 \\
\sqrt{2} \cos(2k\pi u) & \text{for even } k \geq 2
\end{cases}
\]
3.2.4 Chebyshev polynomials

Consider the weight function

\[ w(u) = \frac{1}{\pi \sqrt{u(1-u)}}.I(0 \leq u \leq 1) \]

which is the density of the distribution function

\[ W(u) = 1 - \pi^{-1} \arccos(2u - 1). \]

The Chebyshev polynomials

\[ \varphi_k(u) = \begin{cases} 
1 & \text{for } k = 0 \\
\sqrt{2}. \cos(k. \arccos(2u - 1)) & \text{for } k \geq 1 
\end{cases} \]

form a complete orthonormal sequence in the corresponding Hilbert space \( L^2(w) \).
4 Densities and distribution functions

4.1 Series representation of densities

Given a density $w(x)$ with support $X \subset \mathbb{R}$ and corresponding complete orthonormal sequence $\varphi_j(x)$, every density function $f(x)$ with support contained in $X$ can be written as

$$f(x) = w(x) \left( \sum_{j=0}^{\infty} \gamma_j \varphi_j(x) \right)^2 \text{ a.e. on } X$$

where

$$\sum_{j=0}^{\infty} \gamma_j^2 = 1$$

However, there are uncountable many sequences $\gamma_j$ for which this is true!
In particular, let for an arbitrary \( \tau \in \mathbb{X} \)
\[
g(x|\tau) = (I(x \leq \tau) - I(x > \tau)) \sqrt{f(x)/w(x)}
\]
Then
\[
f(x) = w(x)g(x|\tau)^2
\]
and
\[
g(.|\tau) \in L^2(w).
\]
Consequently
\[
g(x|\tau) = \sum_{j=0}^{\infty} \gamma_j \phi_j(x) \text{ a.e. on } \mathbb{X}
\]
where
\[
\gamma_j = \int g(x|\tau)\phi_j(x)w(x)dx
\]
\[
= \int_{-\infty}^{\tau} \phi_j(x)\sqrt{f(x).w(x)}dx - \int_{\tau}^{\infty} \phi_j(x)\sqrt{f(x).w(x)}dx
\]
Since \( \tau \) is arbitrary, there are uncountable many of such se-
quences \( \gamma_j \).
Because we can always choose
\[ \gamma_0 \in \left( 0, \int \sqrt{f(x)w(x)} \, dx \right) \]
the condition
\[ \sum_{j=0}^{\infty} \gamma_j^2 = 1 \]
can be implemented by reparametrizing the \( \gamma_j \)'s as
\[
\begin{align*}
\gamma_0 &= \frac{1}{\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2}} \\
\gamma_j &= \frac{\delta_j}{\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2}}, \quad j \geq 1,
\end{align*}
\]
where
\[ \sum_{m=1}^{\infty} \delta_m^2 < \infty. \]
**Theorem:** Given a density \( w(x) \) with support \( \mathbb{X} \subset \mathbb{R} \), and corresponding complete orthonormal sequence \( \varphi_j(x) \), for every density function \( f(x) \) with support contained in \( \mathbb{X} \) there exist uncountable many sequences \( \{ \delta_j \}_{j=1}^{\infty} \) satisfying

\[
\sum_{m=1}^{\infty} \delta_m^2 < \infty
\]

such that

\[
f(x) = w(x) \frac{\left(1 + \sum_{j=1}^{\infty} \delta_j \varphi_j(x)\right)^2}{1 + \sum_{m=1}^{\infty} \delta_m^2} \text{ a.e. on } \mathbb{X}.
\]

Moreover, let

\[
f_n(x) = w(x) \frac{\left(1 + \sum_{j=1}^{n} \delta_j \varphi_j(x)\right)^2}{1 + \sum_{m=1}^{n} \delta_m^2}
\]

Then

\[
\lim_{n \to \infty} \int |f_n(x) - f(x)| \, dx.
\]

use this approach to generalize the standard normal density to

\[ f_n(x) = \frac{\exp \left( -\frac{x^2}{2} \right)}{\sqrt{2\pi}} \times \frac{(1 + \sum_{k=1}^{n} \delta_k \varphi_k(x))^2}{1 + \sum_{m=1}^{n} \delta_m^2} \]

where the \( \varphi_k(x) \)'s are Hermite polynomials:

\[
\varphi_0(x) = 1, \quad \varphi_1(x) = x,
\]

\[
\sqrt{k + 1} \varphi_{k+1}(x) - \frac{1}{\sqrt{k + 1}} x \cdot \varphi_k(x) + \sqrt{k} \varphi_{k-1}(x) = 0, \quad k \geq 1,
\]

They call these densities semi-nonparametric (SNP) densities.
The integrals

\[
\int_{-\infty}^{x} \varphi_k(z) \frac{\exp \left( -\frac{z^2}{2} \right)}{\sqrt{2\pi}} dz,
\]

\[
\int_{-\infty}^{x} \varphi_k(z) \varphi_m(z) \frac{\exp \left( -\frac{z^2}{2} \right)}{\sqrt{2\pi}} dz
\]

can be computed recursively as well, so that the SNP distribution function

\[
F_n(x) = \int_{-\infty}^{x} f_n(z) dz
\]

can be computed straightforwardly.
4.2 Density and distribution functions on the unit interval

Let \( G(x) \) be an a priori chosen distribution function with density \( g(x) \) and support
\[
\mathbb{X} = \{ x \in \mathbb{R} : g(x) > 0 \},
\]
Every absolutely continuous distribution function \( F(x) \) with support \( \mathbb{X} \) can be written as
\[
F(x) = H(G(x))
\]
where \( H(u) \) is an absolutely continuous distribution function on \([0, 1]\), namely
\[
H(u) = F \left( G^{-1}(u) \right)
\]
The role of the a priori chosen distribution function $G(x)$ is two-fold:

- $G(x)$ determines the support of $F(x)$.
- $G(x)$ is an initial guess of $F(x)$: If $F(x) = G(x)$, then $H(u) = u$.

The density $f(x)$ of

$$F(x) = H(G(x))$$

can be written as

$$f(x) = h(G(x))g(x)$$

where $h(u)$ is the density of $H(u)$.

Therefore, in modeling general density and distribution functions semi-nonparametrically, it suffices to model the density $h(u)$ and its c.d.f. $H(u)$ semi-nonparametrically.
Theorem: For every density function \( h(u) \) on \([0, 1]\) there exist uncountable many sequences \( \{\delta_j\}_{j=1}^{\infty} \) satisfying
\[
\sum_{m=1}^{\infty} \delta_m^2 < \infty
\]
such that,
\[
h(u) = \frac{(1 + \sum_{k=1}^{\infty} \delta_k \varphi_k(u))^2}{1 + \sum_{m=1}^{\infty} \delta_m^2} \text{ a.e. on } [0, 1]
\]
where the \( \varphi_k(x) \)'s are the Legendre polynomials or the cosine sequence
\[
\varphi_k(x) = \sqrt{2} \cos(k\pi u),
\]
for example. The corresponding SNP densities
\[
h_n(u) = \frac{(1 + \sum_{k=1}^{n} \delta_k \varphi_k(u))^2}{1 + \sum_{m=1}^{n} \delta_m^2}
\]
satisfy
\[
\lim_{n \to \infty} \int_0^1 |h_n(u) - h(u)| \, du.
\]
The advantage of the cosine sequence

\[ \varphi_k(u) = \sqrt{2} \cos(k \pi u), \]

is that then the SNP distribution function

\[ H_n(u) = \int_0^u h_n(z) \, dz \]

has a closed form:

\[
H_n(u) = u \\
\frac{1}{1 + \sum_{m=1}^{n} \delta_m^2} \left[ 2\sqrt{2} \sum_{k=1}^{n} \delta_k \frac{\sin(k \pi u)}{k \pi} + \sum_{m=1}^{n} \delta_m^2 \frac{\sin(2m \pi u)}{2m \pi} + 2 \sum_{k=2}^{n} \sum_{m=1}^{k-1} \delta_k \delta_m \frac{\sin((k + m) \pi u)}{(k + m) \pi} + 2 \sum_{k=2}^{n} \sum_{m=1}^{k-1} \delta_k \delta_m \frac{\sin((k - m) \pi u)}{(k - m) \pi} \right]
\]

has shown that in the case of Legendre polynomials $\varphi_k(u)$ the SNP c.d.f. $H_n(u)$ has a closed form expression too, as

$$H_n(u|\delta) = \frac{(1, \delta') L_{n+1} \Pi_{n+1}(u) L'_{n+1}(\frac{1}{\delta})}{1 + \delta' \delta}$$

where $\delta = (\delta_1, ..., \delta_n)'$,

$$\Pi_{n+1}(u) = \left( \frac{u^{i+j+1}}{i+j+1}; \ i, j = 0, 1, ..., n \right)$$

and $L_{n+1}$ is the $(n+1) \times (n+1)$ lower-triangular matrix with rows formed by the coefficients of the polynomials $\varphi_k(u)$ for $k = 0, 1, ..., n$.

However, the matrix $L_{n+1}$ can only be computed with sufficient accuracy for $n \leq 15$. 
4.3 Compact metric spaces of density and distribution functions

Theorem: Let $\mathcal{D}$ be the space of densities on $[0, 1]$ of the type

$$h(u) = \frac{(1 + \sum_{k=1}^{\infty} \delta_k \varphi_k(u))^2}{1 + \sum_{m=1}^{\infty} \delta_m^2},$$

subject to the restrictions

$$|\delta_k| \leq \overline{\delta}_k$$

for some a priori chosen positive sequence $\overline{\delta}_k > 0$ satisfying

$$\sum_{k=1}^{\infty} \overline{\delta}_k^2 < \infty.$$

For example, let

$$\overline{\delta}_k = c \left(1 + \sqrt{k \ln(k)}\right)^{-1}$$

for some large $c > 0$. 
If we endow $\mathcal{D}$ with the metric

$$\int_0^1 |h_1(u) - h_2(u)| \, du.$$  

then $\mathcal{D}$ is compact.

Let $\mathcal{D}_n$ be the space of SNP densities on $[0, 1]$ of the type

$$h_n(u) = \frac{(1 + \sum_{k=1}^n \delta_k \varphi_k(u))^2}{1 + \sum_{m=1}^n \delta_m^2},$$

subject to the same restrictions on the $\delta_k$'s as before, and endowed with the same metric as $\mathcal{D}$.

Then the sequence $\mathcal{D}_n$ is dense in $\mathcal{D}$:

$$\mathcal{D} = \bigcup_{n=1}^\infty \mathcal{D}_n.$$
Corollary 1: The space

\[ C = \left\{ H(u) = \int_0^u h(v)dv, \; h \in D \right\} \]

of distribution functions on \([0, 1]\) endowed with the "sup" metric

\[ \sup_{0 \leq u \leq 1} |H_1(u) - H_2(u)| \]

is compact.

The spaces

\[ C_n = \left\{ H_n(u) = \int_0^u h_n(v)dv, \; h_n \in D_n \right\} \]

endowed with the sup metric are dense in \( C \):

\[ C = \overline{\bigcup_{n=1}^{\infty} C_n} .\]
Corollary 2: Let $G(x)$ be an a priori chosen absolutely continuous distribution function with support $\mathbb{X} \subset \mathbb{R}$, and let

$$\mathcal{F} = \{ F(x) = H(G(x)) : H \in \mathcal{C} \}$$

$$\mathcal{F}_n = \{ F(x) = H_n(G(x)) : H_n \in \mathcal{C}_n \}$$

Endow these spaces with the sup metric

$$\sup_{x \in \mathbb{X}} |F_1(x) - F_2(x)|.$$

Then $\mathcal{F}$ is compact and $\{ \mathcal{F}_n \}_{n=1}^{\infty}$ is dense in $\mathcal{F}$:

$$\mathcal{F} = \bigcup_{n=1}^{\infty} \mathcal{F}_n.$$
5 The mixed proportional hazard model with fixed right censoring

Let $T$ be a duration. The mixed proportional hazard (MPH) model assumes that, conditional on an observable vector $X$ of covariates and an variable $V > 0$ representing unobserved heterogeneity, the conditional hazard takes the form

$$\lim_{\varepsilon \downarrow 0} \frac{\Pr [T \in [t, t + \varepsilon) | T > t, X, V]}{\varepsilon} = V.\psi_0 (X) . \lambda_0 (t)$$

where

- $\psi_0 (X) > 0$ is the systematic hazard
- $\lambda_0 (t) \geq 0$ is the baseline hazard
- $V$ and $X$ are independent
Then
\[
\Pr [T > t | X] = \int_0^\infty \exp (-v \psi_0 (X) \Lambda_0 (t)) dG_0(v)
\]
where
- \(G_0(v) = \Pr [V \leq v]\)
- \(\Lambda_0 (t) = \int_0^t \lambda_0 (\tau) d\tau\) is the integrated baseline hazard.


have shown that \(G_0, \lambda_0\) and \(\psi_0\) are non-parametrically identified if
- \(E[V] = 1\)
- \(\text{Var}(X)\) is non-singular
- \(\Lambda_0 (c) = 1\) for some fixed \(c > 0\)
Let

- \( H_0(u) = \int_0^\infty u^v \, dG_0(v), \) which is a distribution function on \([0,1]\)
- \( h_0(1) = 1, \) where \( h_0(u) = \int_0^\infty u v^{v-1} \, dG_0(v), \) because then \( E[V] = 1. \)
- \( \psi_0(X) = \exp(\beta_0 X) \)

Then

\[
\Pr [T > t \mid X] = \int_0^\infty \exp \left( -v \psi_0(X) \Lambda_0(t) \right) dG_0(v) \\
= H_0(\exp \left( -\exp(\beta_0 X) \Lambda_0(t) \right))
\]
5.1 Right-censoring

Usually the duration $T$ is only observed up to an upper bound $\bar{T}$, which may vary with $X$. Here I will assume that $\bar{T}$ is a fixed constant:

$$\bar{T} = \bar{t}$$

Denote

- $Y = \min \left( \frac{T}{\bar{t}}, 1 \right)$
- $F_0(u) = \Lambda_0\left( u \bar{t} \right) / \Lambda_0\left( \bar{t} \right)$, $u \in [0, 1]$,
- $\alpha_0 = \ln \left( \Lambda_0\left( \bar{t} \right) \right)$
- $\theta_0 = (\alpha_0, \beta_0')$
- $\mu(\theta_0, X) = \exp(\alpha_0 + \beta_0'X)$
Then, with \( \mu(\theta_0, X) = \exp(\alpha_0 + \beta_0'X) \),
\[
\Pr[Y = 1|X] = H_0(\exp(-\mu(\theta_0, X)))
\]
whereas for \( \tau \in [0, 1) \), with \( F_0(\tau) = \Lambda_0(\tau \bar{t}) / \Lambda_0(\bar{t}) \),
\[
\Pr[Y \leq \tau|X, Y < 1]
= \frac{1 - H_0(\exp(-\mu(\theta_0, X).F_0(\tau)))}{1 - H_0(\exp(-\mu(\theta_0, X)))}.
\]
Therefore,
\[
\Pr[Y \leq \tau|X] = \Psi(\tau|X, \theta_0, F_0, H_0)
\]
where
\[
\Psi(\tau|X, \theta, F, H)
= \begin{cases} 
0 & \text{if } \tau < 0, \\
1 - H(\exp(-\mu(\theta, X)F(\tau))) & \text{if } 0 \leq \tau < 1, \\
1 & \text{if } \tau \geq 1.
\end{cases}
\]
5.2 Identification

Let $H(u)$ be a c.d.f. on $[0, 1]$ with density $h(u)$ satisfying $h(1) = 1$. Then

$$\Psi(Y|X, \theta_0, F_0, H_0) = \Psi(Y|X, \theta, F, H) \text{ a.s.}$$

implies

- $\theta = \theta_0$,
- $F = F_0$,
- $H(u) = H_0(u)$ for $u \in (\underline{u}, 1]$, where $\underline{u}$ is the lower bound of the support of

$$\exp(-\mu(\theta_0, X)) = \exp(-\exp(\alpha_0 + \beta_0'X))$$
Lemma:

\[ E \left[ \exp (i.t.\Psi (Y|X, \theta_0, F_0, H_0)) \mid X \right] = \varphi (t|X, \theta_0, H_0) , \]

where \( i = \sqrt{-1} \) and, with \( \mu (\theta, X) = \exp(\alpha + \beta'X) \),

\[ \varphi (t|X, \theta, H) = \exp(i.t)H(\exp(-\mu (\theta, X))) \]

\[ + \frac{\sin (t. (1 - H(\exp(-\mu (\theta, X))))))}{t} \]

\[ + i.\frac{1 - \cos (t. (1 - H(\exp(-\mu (\theta, X))))))}{t} \]

Impose the condition \( h(1) = 1 \), where \( h(u) = H'(u) \).

If

\[ E \left[ \exp (i.t.\Psi (Y|X, \theta, F, H)) \mid X \right] = \varphi (t|X, \theta, H) \text{ a.s.} \]

for all \( t \) in an arbitrary open neighborhood of zero then

- \( \theta = \theta_0 \),
- \( F(u) = F_0(u) \) for all \( u \in [0, 1] \)
- \( H(u) = H_0(u) \) for all \( u \in (u, 1] \)
**Theorem:** Let $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}^k$ be a bounded one-to-one mapping. Denote

$$
\eta(t|Y, X, \theta, F, H) = t_1. (\varphi(t_1|X, \theta, H) - \exp(i.t_1.\Psi(Y|X, \theta, F, H))) \times \exp(i.t'_2\Phi(X)), \ t = (t_1, t'_2) \in \mathbb{R} \times \mathbb{R}^k,
$$

$$
\overline{Q}(\theta, F, H) = \int_T |E[\eta(t|Y, X, \theta, F, H)]|^2 \, dt,
$$

where for example

$$
T = \times_{m=1}^{k+1} [-c_m, c_m], \ c_m > 0
$$

Impose the condition $h(1) = 1$, where $h(u) = H'(u)$. Then

$$
\overline{Q}(\theta, F, H) = 0
$$

if and only if

- $\theta = \theta_0,$
- $F(u) = F_0(u)$ for all $u \in [0, 1]$
- $H(u) = H_0(u)$ for all $u \in (u, 1].$
5.3 Sieve estimation

Let

- $H_0 \in \mathcal{H}$, where $\mathcal{H}$ is a compact metric space of absolutely continuous c.d.f.'s $H$ on $[0, 1]$ with density $h(u)$ satisfying $h(1) = 1$, endowed with the metric
$$\sup_{u \in (u, 1]} |H_1(u) - H_2(u)|$$

- $F_0 \in \mathcal{F}$, where $\mathcal{F}$ is a compact metric space of absolutely continuous c.d.f.'s $F$ on $[0, 1]$, endowed with the metric
$$\sup_{u \in [0,1]} |F_1(u) - F_2(u)|$$

- $\theta_0 = (\alpha_0, \beta_0)' \in \Theta$, where $\Theta \subset \mathbb{R}^{k+1}$ is compact
Then
\[
\bar{Q}(\theta, F, H) = \int_T |E[\eta(t|Y, X, \theta, F, H)]|^2 dt
\]
is continuous on $\Theta \times \mathcal{F} \times \mathcal{H}$,
\[
(\theta_0, F_0, H_0) = \arg \min_{(\theta, F, H) \in \Theta \times \mathcal{F} \times \mathcal{H}} \bar{Q}(\theta, F, H)
\]
is unique, and
\[
p \lim_{N \to \infty} \sup_{(\theta, F, H) \in \Theta \times \mathcal{F} \times \mathcal{H}} \left| \hat{Q}_N(\theta, F, H) - \bar{Q}(\theta, F, H) \right| = 0,
\]
where
\[
\hat{Q}_N(\theta, F, H) = \int_T \left| \frac{1}{N} \sum_{j=1}^N \eta(t|Y_j, X_j, \theta, F, H) \right|^2 dt
\]
Therefore, all the conditions for sieve estimation of $(\theta_0, F_0, H_0)$ are satisfied.
6 First-price auction models

6.1 Simplifying assumptions

- The reservation price is non-binding
- The same auction is repeated independently \( L \) times
- The number of bidders, \( I \geq 2 \), is the same across auctions
- The values \( V_j \) of the bidders are i.i.d. within and across auctions
- \( E[V_j] < \infty \)

Then the optimal bids are

\[
B_j = V_j - \frac{\int_0^{V_j} F_0(x)^{I-1} dx}{F_0(V_j)^{I-1}}, \quad j = 1, 2, \ldots, N = L \times I
\]

where the values \( V_j \) are random drawings from the unknown value distribution \( F_0(v) \).

The bid distribution has bounded support.

propose the following sieve estimation method:

Let $F_0 \in \mathcal{F}$, where $\mathcal{F}$ is a compact metric space of absolutely continuous c.d.f.’s $F$ on $(0, \infty)$, endowed with the metric $\sup_{v>0} |F_1(v) - F_2(v)|$.

For each $F \in \mathcal{F}$ draw simulated values $\tilde{V}_j(F')$ from $F$, with corresponding simulated bids

$$\tilde{B}_j(F) = V_j - \frac{\int_0^{\tilde{V}_j(F)} F(x)^{I-1} dx}{F\left(\tilde{V}_j(F')\right)^{I-1}}$$
Let for some $c > 0$,
\[
\overline{Q}(F) = \int_{-c}^{c} |\psi(t, F)|^2 \, dt, \quad \widehat{Q}_N(F) = \int_{-c}^{c} |\widehat{\psi}_N(t, F)|^2 \, dt
\]
where
\[
\psi(t, F) = E[\exp(i.t.B)] - E[\exp(i.t.\widetilde{B}(F))]
\]
\[
\widehat{\psi}_N(t, F) = \frac{1}{N} \sum_{j=1}^{N} \exp(i.t.B_j) - \frac{1}{N} \sum_{j=1}^{N} \exp(i.t.\widetilde{B}_j(F))
\]

Then
- $\overline{Q}(F)$ is continuous on $\mathcal{F}$
- $F_0 = \arg\min_{F \in \mathcal{F}} \overline{Q}(F)$ is unique
- $p \lim_{N \to \infty} \sup_{F \in \mathcal{F}} \left| \widehat{Q}_N(F) - \overline{Q}(F) \right| = 0$

Hence, $F_0$ can be estimated consistently by sieve estimation.
6.2 Some numerical experiments

Number of bidders per auction: \( I = 5 \)
Number of auctions: \( L = 200 \)
True value distribution \( F_0: \chi^2_r, \) for \( r = 3, 4, 5 \)
Sieve estimator of \( F_0: \)

\[
F_n(v) = H_n (G(v)) = \int_0^{G(v)} \frac{(1 + \sum_{k=1}^n \delta_k \varphi_k(u))^2}{1 + \sum_{m=1}^n \delta_m^2} du
\]

where

- The \( \varphi_k(u) \)'s are Legendre polynomials
- \( G(v) = 1 - \exp(-v/3) \) is the initial guess of \( F_0(v) \)
Figure 1. $g(v) = \exp(-v/3)/3$ (dashed curve) compared with the $\chi_r^2$ densities for $r = 3, 4, 5$
The sieve orders \( n \) have been determined via an information criterion:

\[
\tilde{C}_N(n) = \inf_{F \in \mathcal{F}_n} \hat{Q}(F) + \left(1 - (n + 1)^{-1/3}\right) \frac{\ln(\ln(N))}{N},
\]

\[
n_N = \arg \min_n \tilde{C}_N(n)
\]

The following estimates were obtained:

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r ):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n_N ):</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>
Figure 2. $\tilde{f}_4(v)$ (dashed curve) compared with the true $\chi^2_3$ density $f_0(v)$
Figure 3. $\tilde{f}_2(v)$ (dashed curve) compared with the true $\chi^2_4$ density $f_0(v)$
Figure 4. $\tilde{f}_4(v)$ (dashed curve) compared with the true $\chi^2_5$ density $f_0(v)$