Nonparametric Identification and Estimation of a Transformation Model

Hidehiko Ichimura and Sokbae Lee

University of Tokyo and Seoul National University

15 February, 2012

Outline

1. The Model and Motivation
2. Identification
3. Consistency
4. Asymptotic Distribution
5. Computation
6. Further Research

The model

- Assume X and Y are observed, where X is a vector of covariates and Y is the dependent variable, (Y, X) being generated by

$$\Lambda_0(Y) = m_0(X) + U,$$

  where $\Lambda_0(\cdot)$ is a strictly increasing unknown function, $m_0(\cdot)$ is an unknown function, U is an unobserved random variable that is independent of X with the cumulative distribution function $F_0(\cdot)$.
- The objective of this project is to identify and estimate unknown functions Λ_0, m_0, and F_0.

Related Literature (1)

- Box and Cox (1964) specify all three functions parametrically; $m_0(X) = X'\beta$, $U \sim N(0, \sigma^2)$ and for $y > 0$,

$$y^\lambda = \begin{cases} (y^\lambda - 1)/\lambda & \text{if } \lambda \neq 0 \\ \log y & \text{if } \lambda = 0. \end{cases}$$

- Horowitz (1996), Ye and Duan (1997), and Chen (2002) specify $m_0(X) = X'\beta$.
- Linton, Sperlich and Van Keilegom (2008) specify $m_0(X) = G(m_1(X_1), \ldots, m_K(X_K))$, where G is a known function.

Related Literature (2)

- Jacho-Chavez, Lewbel, and Linton (2010) studies a model:

$$r(x, z) = H(G(x) + F(z))$$

where H is a strictly monotonic function and $r(x, z)$ is nonparametrically estimatable model assuming continuous regressors.

Related Literature (3)

- The nonparametric transformation model includes the mixed proportional hazard model as a special case (Ridder, 1990). For the proportional hazard model of duration Y with unobserved heterogeneity v,

$$\lambda(t|x, v) = \lambda_0(t) \exp[-(m_0(x) + v)],$$

Let $T(t) = \int_0^t \lambda_0(s) ds$. Then

$$\Lambda_0(Y) = \log T(Y) = m_0(X) + V + \epsilon$$

where ϵ is independent of (X, U) and has the CDF $1 - \exp(-\exp(t))$ (a Gompertz distribution).

Related Literature (4)

- Ekeland, Heckman, and Nesheim (2004) present a hedonic model which, under suitable restrictions placed on utility functions and production functions gives rise to a nonparametric transformation model.

  For example they assume production function $F(z, x, \epsilon)$ with hedonic element z, observed and unobserved control variables x and ϵ, cost function $C(z)$.

Related Literature (5)

The first order condition is

$$F_z(z, x, \epsilon) = C'(z).$$

They assume an existence of a monotonic transformation ψ such that

$$\psi(F_z(z, x, \epsilon)) = \tau(z) + M(\eta(x) + \epsilon),$$

where M is also monotonic. Then

$$\psi(C'(z)) = \tau(z) + M(\eta(x) + \epsilon),$$

so that

$$M^{-1}(\psi(C'(z)) - \tau(z)) = \eta(x) + \epsilon.$$

Related Literature (5)

- Guerre, Perrigne, and Vuong (2009) study the identification of a private value first-price auction model with risk averse bidders and showed that, writing $\lambda(\cdot) = U(\cdot)/U'(\cdot)$, denoting the number of bidders by n, and the density of the equilibrium bid by $g(\cdot|n)$, and bids's α-quantile by $b(\alpha|n)$ we have

$$\lambda^{-1}\left(\frac{\alpha}{(n_2-1)g(b(\alpha|n_2))}\right) = b(\alpha|n_1) - b(\alpha|n_2) + \lambda^{-1}\left(\frac{\alpha}{(n_1-1)g(b(\alpha|n_1))}\right)$$

Objectives

All previous studies approach the identification via derivatives with respect to regressors excluding discrete regressors.

In this paper we

- construct an identification result which does not depend on taking derivatives with respect to regressors,
- consider identification issues when regressors are discrete, and
- based on the identification results develop estimators for functions Λ_0, m_0, and F_0,
- develop computatinal method for the estimator,
- establish consistency and $1/\sqrt{n}$-consistency and outline the distribution theory.
- Our study helps isolate the variation intrinsic to the parameter identification.

Point Identification (1)

- We first provide a point identification result when X includes a continuous regressor.
- Note that for any constants a, b, and $c > 0$, we have

$$c[\Lambda_0(Y) + b] = c[m_0(X) + b + a] + c(U - a),$$

- Thus, the model is not identified without location and scale normalization.
- Our location and scale normalization is achieved by assuming that $F_0(0) = 0.5$, $m_0(x_0) = 0$, and $m_0(x_1) = 1$ for some (x_0, x_1) such that $x_1 \neq x_0$.

Point Identification (2)

- Since $Y = \Lambda_0^{-1}[m_0(x) + U]$,

$$Q_{Y|X}(\alpha|x) = \Lambda_0^{-1}[m_0(x) + F_0^{-1}(\alpha)].$$

- As $F_0^{-1}(0.5) = 0$, $m_0(x) = \Lambda_0[Q_{Y|X}(0.5|x)]$.
- As $m_0(x_0) = 0$, $F_0^{-1}(\alpha) = \Lambda_0[Q_{Y|X}(\alpha|x_0)]$.
- Thus

$$\Lambda_0[Q_{Y|X}(\alpha|x)] = \Lambda_0[Q_{Y|X}(0.5|x)] + \Lambda_0[Q_{Y|X}(\alpha|x_0)].$$

Point Identification (3)

- We seek sufficient conditions under which the unique solution (a.s.) to the following equation is Λ_0 on a compact set $[y_1, y_2] \supset [0, 1]$:

$$\Lambda[Q_{Y|X}(\alpha|x)] = \Lambda[Q_{Y|X}(0.5|x)] + \Lambda[Q_{Y|X}(\alpha|x_0)].$$

- Substituting the above relationships, we have

$$\begin{aligned}&\Lambda[\Lambda_0^{-1}(m_0(x) + F_0^{-1}(\alpha))]\\ &\quad = \Lambda[\Lambda_0^{-1}(m_0(x))] + \Lambda[\Lambda_0^{-1}(F_0^{-1}(\alpha))].\end{aligned}$$

Point Identification (4)

- Examining this with $s = m_0(x)$ and $t = F_0^{-1}(\alpha)$ and $T(z) := \Lambda[\Lambda_0^{-1}(z)]$, we have

$$T(s+t) = T(s) + T(t).$$

- When s and t vary over a common interval $[y_1, y_2]$, and T is continuous, $T(z) = Cz$.
- Since $\Lambda_0(Q_{Y|X}(\alpha|x)) = m_0(x) + F_0^{-1}(\alpha)$, the scale normalization implies $\Lambda_0(Q_{Y|X}(0.5|x_1)) = 1$ so that $\Lambda(Q_{Y|X}(0.5|x_1)) = 1$ always via the scale normalization. This implies $T(1) = \Lambda(\Lambda_0^{-1}(1)) = 1$ so that $C = 1$. Thus $T(z) = z$.
- Therefore, Λ_0 is identified on $[y_1, y_2]$.

Point Identification (5)

- Note that once $\Lambda_0(y)$ is identified, then $m_0(x)$ and $F_0^{-1}(\alpha)$ are identified by $m_0(x) = \Lambda_0[Q_{Y|X}(0.5|x)]$ and $F_0^{-1}(\alpha) = \Lambda_0[Q_{Y|X}(\alpha|x_0)]$.
- Hence, $m_0(x)$ is identified for any x satisfying $Q_{Y|X}(0.5|x) \in [y_1, y_2]$.
- Likewise, $F_0^{-1}(\alpha)$ is identified for any α satisfying $Q_{Y|X}(\alpha|x_0) \in [y_1, y_2]$.

Identification Theorem

- If the distribution of Y conditional on X is given, (Y, X) being generated by the model specified above,
- Λ_0 is a strictly increasing and continuous function,
- $F_0(0) = 0.5$ and $m_0(x_0) = 0$ and $m_0(x_1) = 1$ for two distinct points x_0 and x_1, and
- the marginal distribution of $m_0(X)$ and U are continuous and their joint support include $[0, 1]^2$.
- Then there exists a non-empty set $[y_1, y_2]$, which is a strict subset of the common support of $Q_{Y|X}(0.5|x)$ and $Q_{Y|X}(\alpha|x_0)$ over which Λ_0 is identified. $m_0(x)$ is identified for any x satisfying $Q_{Y|X}(0.5|x) \in [y_1, y_2]$. Likewise, $F_0^{-1}(\alpha)$ is identified for any α satisfying $Q_{Y|X}(\alpha|x_0) \in [y_1, y_2]$.

Partial Identification (1)

- We next consider a discrete X taking a finite number of probability mass points.
- In this case s takes on a finite number of points which includes 0 and 1 from our normalization.
- Note that

$$T(s_i + t) = T(s_i) + T(t)$$

and

$$T(s_j + t) = T(s_j) + T(t)$$

so that

$$T(s_i + t) - T(s_j + t) = T(s_i) - T(s_j).$$

Partial Identification (2)

- Differentiating both sides we have

$$T'(s_i + t) - T'(s_j + t) = 0$$

or writing $\Delta_{ij} = s_j - s_i$ we have

$$T'(t) = T'(t + \Delta_{ij}).$$

- Thus $T'(z)$ is a periodic function with periodicity Δ_{ij}.
- Note also that since

$$T'(t + \Delta_{ij}) = T'(t + \Delta_{i'j'}),$$

we also have

$$T'(t) = T'(t + \Delta_{ij} - \Delta_{i'j'}),$$

which implies that $T'(z)$ is a periodic function with periodicity $\Delta_{ij} - \Delta_{i'j'}$.

Partial Identification (3)

- We consider the case in which there is a $\Delta > 0$ in all Δ_{ij} and differences among them and so on so that each of these terms can be written as an integer multiple of Δ.
- Clearly this is not necessarily the case, but it is also clear that this is always the case if all s_i are rational.
- We show that, in this case, we cannot point identify Λ_0 but can provide a bound and the bound becomes tight as Δ becomes small.
- Let ψ to be a non-constant periodicity Δ function with $\int_0^\Delta \psi(t)dt = 0$.
- Then $T'(s) = \psi(s) + A$ for some constant A so that

$$T(s) = \int_0^s \psi(u)du + A \cdot s + B$$

for some constants A and B.

Partial Identification (4)

- Clearly $B = 0$ and that $T(1) = 1$ and the fact the location normalization implies $\int_0^1 \psi(u)du = 0$ implies $A = 1$ so that

$$T(s) = \int_0^s \psi(u)du + s.$$

Recall that $T(s) = \Lambda(\Lambda_0^{-1}(s))$, taking $v = \Lambda_0^{-1}(s)$,

$$\Lambda(v) = \Lambda_0(v) + \int_0^{\Lambda_0(v)} \psi(u)du.$$

Partial Identification (5)

- Let $h(z) = \int_0^z \psi(t)dt$. Then

$$
\begin{aligned}
h(z+\Delta) &= \int_0^{z+\Delta} \psi(t)dt \\
&= h(z) + \int_z^{z+\Delta} \psi(t)dt, \\
&= h(z).
\end{aligned}
$$

so that h is also a periodicity Δ function and

$$\Lambda(u) = \Lambda_0(u) + h(\Lambda_0(u)).$$

Partial Identification (6)

- Differentiating $\Lambda(u)$ we observe that

$$\Lambda'(u) = \Lambda_0'(u)[1 + h'(\Lambda_0(u))]$$

so that the minimum slope of h' is greater than -1.
- Thus the supremum value h can take is Δ.
- Thus the supremum deviation of $\Lambda(u)$ from $\Lambda_0(u)$ is $|\Delta|$.

Estimating Λ_0 (1)

- To construct a root-n-consistent estimator of Λ_0, without the loss of generality, we normalize m_0 by

$$\int w_X(x) m_0(x) dx = 0$$

for some weight function w_X that has a compact subset on the support of X and satisfies $\int w_X(x)dx = 1$. Analogously we normalize $F_0(\alpha)$ by

$$\int w_\alpha(a) F_0^{-1}(a) da = 0$$

for some weight function w_α that has a compact subset on the support of α and satisfies $\int w_\alpha(a)da = 1$.

Estimating Λ_0 (1)

- Then $F_0^{-1}(\alpha) = \int w_X(x)\Lambda_0[Q_{Y|X}(\alpha|x)]dx$ and $m_0(x) = \int w_\alpha(a)\Lambda_0[Q_{Y|X}(\alpha|x)]da$, so that

$$\Lambda_0[Q_{Y|X}(\alpha|x)] = \int w_\alpha(a)\Lambda_0[Q_{Y|X}(\alpha|x)]da + \int w_X(u)\Lambda_0[Q_{Y|X}(\alpha|u)]du.$$

This equation is the basis for our estimator of Λ_0.

Estimating Λ_0 (2)

- Our estimation procedure for Λ_0 consists of two steps.
- The first step is nonparametric estimation of $Q_{Y|X}(\alpha|x)$.
- Then the second step is to fit the data by minimizing an objective function based on the identification result. We use the least squares criterion.

Estimating Λ_0 (3)

- We minimize a least-squares criterion function

$$M_n(\Lambda) := \int w_X(x) w_\alpha(\alpha) \Big\{ \Lambda[\hat{Q}_{Y|X}(\alpha|x)] - \int w_\alpha(a) \Lambda[\hat{Q}_{Y|X}(a|x)] da - \int w_X(u) \Lambda[\hat{Q}_{Y|X}(\alpha|u)] du \Big\}^2 dx\, da$$

  over a set of possible functions for Λ, $\mathbb{L}_n$, where
- $\hat{Q}_{Y|X}(\alpha|x)$ is the first-step estimator,

Estimating m_0 and F_0

- m_0 can be estimated using the normalization $\int w_\alpha(a) F_0^{-1}(a) da = 0$:

$$\hat{m}(x) := \int w_\alpha(a) \hat{\Lambda}[\hat{Q}_{Y|X}(a|x)] da.$$

- F_0 can be estimated using the normalization $\int w_X(x) m_0(x) dx = 0$:

$$\hat{F}^{-1}(\alpha) := \int w_X(x) \hat{\Lambda}[\hat{Q}_{Y|X}(\alpha|x)] dx.$$

Lemma for Showing Consistency

- Denote the probability limit of the objective function by $M(\Lambda, Q_{Y|X}(\alpha|x))$, the support of (X, α) by $\mathcal{S}$, and for some $L > 0$, define

$$\mathbf{L} = \{\Lambda; |\Lambda'| \leq L, \int\int w_X(u) w_\alpha(a) \Lambda(Q_{Y|X}(a|u) dadu = 0$$
$$\int\int w_X(u) w_\alpha(a) \Lambda^2(Q_{Y|X}(a|u) dadu = 1\}.$$

- Then there is a universal constant C such that for any $\Lambda \in \mathbf{L}$,

$$M(\Lambda, Q_{Y|X}(\alpha|x)) \geq C \sup_{(x,\alpha)\in\mathcal{S}} |\Lambda(Q_{Y|X}(\alpha|x)) - \Lambda_0(Q_{Y|X}(\alpha|x))|^2.$$

Asymptotic Normality: Outline of the Proof I

- We assume that the first order condition of the optimization problem is satisfied with equality so that

$$\hat{\Lambda}(\hat{Q}(\alpha|x)) = \int w_X(u)\hat{\Lambda}(\hat{Q}(\alpha|u))du + \int w_\alpha(a)\hat{\Lambda}(\hat{Q}(a|x))da.$$

- Substitute $t = \hat{Q}(\alpha|x)$, we obtain,

$$\hat{\Lambda}(t) = \int w_X(u)\hat{\Lambda}(\hat{Q}(\hat{Q}^{-1}(t|x)|u))du + \int w_\alpha(a)\hat{\Lambda}(\hat{Q}(a|x))da.$$

- Then by location normalization,

$$\hat{\Lambda}(t) = \int w_X(x) \int w_X(u)\hat{\Lambda}(\hat{Q}(\hat{Q}^{-1}(t|x)|u))dudx.$$

Asymptotic Normality: Outline of the Proof II

- Define

$$T_0(\Lambda)(t) := \int w_X(x) \int w_X(u)\Lambda(Q_0(Q_0^{-1}(t|x)|u))dudx,$$
$$\hat{T}(\Lambda)(t) := \int w_X(x) \int w_X(u)\Lambda(\hat{Q}(\hat{Q}^{-1}(t|x)|u))dudx.$$

- Both T_0 and $\hat{T}$ are linear operators and the latter above can be seen as empirical approximation to T_0 using the nonparametrically estimated conditional quantile function.
- Note that $\Lambda_0 = T_0\Lambda_0$, that is $(I - T_0)\Lambda_0 = 0$. Since Λ_0 is different from zero, $(I - T_0)$ must not be invertible.

Asymptotic Normality: Outline of the Proof III

- However, we write as

$$\begin{aligned}\hat{\Lambda} - \Lambda_0 &= \hat{T}\hat{\Lambda} - T_0\Lambda_0 \\ &= T_0(\hat{\Lambda} - \Lambda_0) + (\hat{T} - T_0)\Lambda_0 + (\hat{T} - T_0)(\hat{\Lambda} - \Lambda_0) \\ &= T_0^{m+1}(\hat{\Lambda} - \Lambda_0) + [I + T_0 + \cdots + T_0^m](\hat{T} - T_0)\Lambda_0 \\ &+ [I + T_0 + \cdots + T_0^m](\hat{T} - T_0)(\hat{\Lambda} - \Lambda_0),\end{aligned}$$

  where the last equality holds for any positive integer m.
- Note that in view of the identification result, $\text{Null}(I - T_0) := \{\phi : T_0\phi = \phi\}$ has the form

$$\text{Null}(I - T_0) = \{c_0 + c_1\Lambda_0 : (c_0, c_1) \in \mathbb{R}^2\}.$$

Asymptotic Normality: Outline of the Proof IV

- In general, $\|T_0\| = 1$; however, in view of the identification result,

$$\sup\{\|T_0\phi\| : \|\phi\| \leq 1 \text{ and } \phi \notin \text{Null}(I - T_0)\} < 1.$$

- Hence, if $\sqrt{n}(\hat{\Lambda} - \Lambda_0) \notin \text{Null}(I - T_0)$ with probability approaching one, then

$$\lim_{m\to\infty} \text{plim}_{n\to\infty} T_0^{m+1}\left[\sqrt{n}(\hat{\Lambda} - \Lambda_0)\right] = 0.$$

- Suppose that we can show that

$$\left\|\hat{T} - T_0\right\| = o_p(1).$$

Asymptotic Normality: Outline of the Proof V

- Then

$$\text{plim}_{n\to\infty}(\hat{T} - T_0)[\sqrt{n}(\hat{\Lambda} - \Lambda_0)] = 0.$$

Let

$$\hat{R}_m := [I + T_0 + \cdots + T_0^m](\hat{T} - T_0)[\sqrt{n}(\hat{\Lambda} - \Lambda_0)].$$

Now $\text{plim}_{n\to\infty}\hat{R}_m = 0$ for any integer $m \geq 1$.

Asymptotic Normality: Outline of the Proof VI

- Suppose that the following strong approximation holds:

$$\sqrt{n}(\hat{T} - T_0)\Lambda_0 = \mathbb{G}_0 + o_p(1),$$

where $\mathbb{G}_0$ is a centered Gaussian process with almost sure continuous sample paths. Then $\mathbb{G} \notin \text{Null}(I - T_0)$ with probability one, so that

$$\mathbb{H} := \lim_{m\to\infty} [I + T_0 + \cdots + T_0^m]\mathbb{G}_0$$

is well defined and also is a centered Gaussian process. In summary,

$$\sqrt{n}(\hat{\Lambda} - \Lambda_0) \to_d \mathbb{H}.$$

Computation

- The second stage requires a constrained optimization since Λ is a strictly monotone increasing function.
- We build on Ramsay (1998) who proposes a method for estimating an arbitrary twice differentiable strictly monotone function.

Ramsay's method (1)

- To describe Ramsay (1998)'s method, consider a class of monotone functions

$$\mathcal{F} = \{\text{monotone } f : \ln(Df) \text{ is differentiable}$$
$$\text{and } D\{\ln(Df)\} \text{ is Lebesgue square integrable}\},$$

where the notation Df refers to the first-order derivative of f.

- Thus, each element of $\mathcal{F}$ is strictly increasing and its first derivative is smooth and bounded almost everywhere.

Ramsay's method (2)

- Using the same notation as in Ramsay (1998), define a partial integration operator $D^{-1}f$ by

$$D^{-1}f(t) = \int_0^t f(s)ds. \tag{1}$$

Theorem 1 of Ramsay (1998) states that every function $f \in \mathcal{F}$ can be represented as

$$f(x) = C_0 + C_1 D^{-1}\left\{\exp(D^{-1}w)\right\}(x), \tag{2}$$

where w is a Lebesgue square integrable function and C_0 and C_1 are arbitrary constants.

Sieve Space (1)

- Assume that Λ_0 belongs to $\mathcal{F}$ and that w can be written as $w(t) = \sum_{k=1}^{\infty} c_k \phi_k(t)$ with some basis functions $\{\phi_k : k = 1, 2, \ldots\}$.

Sieve Space (2)

- Then we consider a natural sieve space $\mathbb{L}_n$ by imposing the scale normalization $C_1 = 1$ and the location normalization $C_0(\mathbf{c})$:

$$\mathbb{L}_n = \left\{ C_0(\mathbf{c}) + D^{-1} \left[\exp \left(\sum_{k=1}^{K_n} c_k D^{-1} \phi_k \right) \right] \right\},$$

where

$$C_0(\mathbf{c}) = -\int w_0(u) \\ D^{-1} \left\{ \exp \left(\sum_{k=1}^{K_n} c_k D^{-1} \phi_k \right) \right\} \left(\left[\hat{Q}_{Y|X}(0.5|u) \right] \right) du$$

and $\mathbf{c} = (c_1, \ldots, c_K)$ for some K_n that converges to infinity as $n \to \infty$.

The Second Step with Sieve Space (1)

- Note that the location normalization ensures that $\hat{F}^{-1}(0.5) = 0$.
- Thus, for each n, minimizing the objective function can be now viewed as an unconstrained optimization problem with K_n-dimensional $\mathbf{c}$.
- Specifically, our second-stage estimation consists of

$$\min_{\Lambda \in \mathbb{L}_n} \left\{ M_n(\Lambda) + \lambda \int_{y_1}^{y_2} w_{K_n}^2(t) dt \right\}, \tag{3}$$

where λ is a regularization parameter that converges to zero and $w_{K_n}(t) = \sum_{k=1}^{K_n} c_k \phi_k(t)$.

The Second Step with Sieve Space (2)

- Then (3) can be solved by some numerical optimization algorithm, along with the location normalization imposed on Λ.

We use the following model to conduct a simple Monte Carlo simulation:

$$\Lambda_0(Y) = X + U$$

where X and U are both standard normal random variables and

$$\Lambda_0(t) = \log t$$

Table 1. *The Finite Sample Performance of the Estimator of* Λ_0

y	New Estimator BIAS	SD	RMSE	CS RMSE	HJ RMSE	KS RMSE	YD RMSE
0.050							
0.240	0.004	0.173	0.173	0.094	0.124	0.102	0.104
0.430	0.013	0.129	0.130	0.069	0.069	0.069	0.076
0.620	0.024	0.094	0.097	0.051	0.046	0.048	0.058
0.810	0.026	0.075	0.079	0.036	0.022	0.033	0.040
4.817	-0.018	0.048	0.051	0.103	0.246	0.129	0.106
8.634	-0.062	0.102	0.119	0.134	0.406	0.586	0.166
12.451	-0.054	0.149	0.158	0.156	0.475	0.644	0.260
16.268	0.000	0.200	0.200	0.181	0.541	1.014	0.372
20.086	0.069	0.247	0.256	0.195	0.572	1.332	0.489

CS:Songnian Chen's estimator, HJ: Joel Horowitz's estimator
KS: Klein-Sherman's estimator, YD: Ye-Duan's estimator

Summary

- We obtained new identification results, proposed a sample analog estimation method for the nonparametric transformation model, and obtained some asymptotic distribution results.
- Things to be done:
- Complete the asymptotic analysis for Λ_0 and
- obtain results for estimation of m_0 and F_0.
- Monte Carlo experiments.
- Use the results for empirical illustration.