IDENTIFICATION AND ESTIMATION

OF THE DISTRIBUTION OF UNOBSERVABLES IN

DYNAMIC DISCRETE CHOICE STRUCTURAL MODELS

(ECO 2403)

Victor Aguirregabiria

Winter 2017

INTRODUCTION & MOTIVATION

• Overview of recent literature on the identification and estimation of the **distribution of unobservables** in **dynamic discrete choice structural models**.

• In this class of models, distribution of the unobservables has **two main implications**.

• **[1]** As in any discrete choice model (static or dynamic), this distribution determines how individual treatment effects are aggregated to obtain average treatment effects for the population (or a subpopulation).

• **[2]** This distribution has important implications on the risk that agents face and therefore on their behavior.

EXAMPLE

- Two-period binary choice model [firm market entry]. For $t = 1, 2$, the binary variable $Y_{it} \in \{0, 1\}$ represents entry decision of firm i at period t.

- The optimal decision at the last period $t = 2$ is: $\{Y_{i2} = 1\}$ iff

$$X_{i2}\,\theta + \alpha\, Y_{i1} + \varepsilon_{i2} \geq 0$$

- The optimal decision at period $t = 1$ is: $\{Y_{i1} = 1\}$ iff

$$\{X_{i1}\,\theta + \varepsilon_{i1} + \beta\,\mathbb{E}\left[\max\{X_{i2}\,\theta + \alpha + \varepsilon_{i2}\ ,\ 0\}\right] \geq \beta\,\mathbb{E}\left[\max\{X_{i2}\,\theta + \varepsilon_{i2}\ ,\ 0\}\right]\}$$

- The option values $\mathbb{E}\left[\max\{X_{i2}\,\theta + \alpha + \varepsilon_{i2}\ ,\ 0\}\right]$ and $\mathbb{E}\left[\max\{X_{i2}\,\theta + \varepsilon_{i2}\ ,\ 0\}\right]$, and the difference between them, depend on the shape of the distribution of ε_{i2}.

OUTLINE

1. Static binary choice models

   (a) Identification

   (b) Estimation methods

2. Dynamic binary choice models

   (a) Identification

   (b) Estimation methods

1. STATIC BINARY CHOICE MODELS

- Consider the binary choice model:

$$Y = \mathbf{1}\{Z + \pi(X) - \varepsilon \geq 0\}$$

- $\mathbf{1}\{.\}$ is the indicator function
- Z and X are observable; ε is unobservable.
- ε is continuous with support $\mathbb{R}$.
- $\pi(.)$ is an unknown (nonparametric) function.
- $F(.)$ is the distribution function of ε that is nonparametric.
- Researcher observes random sample: $\{y_i, z_i, x_i : i = 1, 2, ..., N\}$

1.1. IDENTIFICATION [1]

• Define the *Conditional Choice Probability (CCP)* function:

$$P(z,x) = \Pr(Y = 1 \mid Z = z,\ X = x)$$

• Under mild regularity conditions, function $P(z,x)$ is NP identified for every value (z,x) in the support set $\mathcal{Z} \times \mathcal{X}$.

• All the information in the data about $\pi(.)$ and $F(.)$ is contained in the CCP function $P(z,x)$. The model constraints are:

$$P(z,x) \;=\; F_{\varepsilon|(z,x)}(z + \pi(x))$$

• We are interested in the identification of $\pi(.)$ and $F(.)$ from $P(.,.)$.

IDENTIFICATION Matzkin (ECMA, 1992)

Theorem [Matzkin (ECMA, 1992)]. Suppose that:

(a) ε and Z are independent;

(b) ε has median zero and median independent of X;

(c) $\pi(.)$ is bounded on $\mathcal{X}$: $\pi(\mathcal{X}) \in [\pi_L, \pi_H] \subset \mathbb{R}$;

(d) for any $X = x$, variable Z has continuous support on $[z_L, z_H]$ with $z_L + \pi_H < 0$ and $z_H + \pi_L > 0$.

Then, $\pi(x)$ and $F_{\varepsilon|x}(\varepsilon)$ are NP identified for every $x \in \mathcal{X}$ and $\varepsilon \in [z_L + \pi(x_0), z_H + \pi(x_0)]$.

Matzkin (ECMA, 1992)   Proof

• Let $(x_0, \varepsilon_0) \in \mathcal{X} \times \mathbb{R}$ be arbitrary values of X and ε. Define $z^*(x_0)$ as the value of z that solves the equation:

$$P(z, x_0) = \frac{1}{2}$$

Condition (a) implies that $P(z, x_0)$ is strictly increasing in z; and condition (d) implies that the solution $z^*(x_0)$ always exists. Therefore, function $z^*(x_0)$ is NP identified everywhere on $\mathcal{X}$.

• Given that $P(z, x) = F_{\varepsilon|x}(z + \pi(x))$ and under condition (a), we have that $P(z, x) = \frac{1}{2}$ is equivalent to $z + \pi(x) = 0$. Therefore, by construction, $z^*(x_0) + \pi(x_0) = 0$, such that function $\pi(x_0)$ is NP identified everywhere on $\mathcal{X}$ as

$$\pi(x_0) = -z^*(x_0)$$

Matzkin (ECMA, 1992) Proof [2]

• Now, given any pair $(x_0, \varepsilon_0) \in \mathcal{X} \times [z_L + \pi(x_0), z_H + \pi(x_0)]$ we can construct the value $\widetilde{z}(x_0, \varepsilon_0) \equiv \varepsilon_0 - \pi(x_0)$. By construction,

$$\begin{aligned} F_{\varepsilon|x_0}(\varepsilon_0) &= F_{\varepsilon|x_0}\left[\widetilde{z}(x_0, \varepsilon_0) + \pi(x_0)\right] \\ &= P\left[\widetilde{z}(x_0, \varepsilon_0), x_0\right] \end{aligned}$$

such that $F_{\varepsilon|x}(\varepsilon)$ is NP identified on $\mathcal{X} \times [z_L + \pi(x), z_H + \pi(x)]$.

• **Remark 1:** If $[z_L, z_H] = \mathbb{R}$, then $F_{\varepsilon|x}(\varepsilon)$ is identified everywhere.

• **Remark 2:** Median independence between Z and ε can be replaced by other quantile independence.

Example 1 (Binary choice demand model)

• Suppose that we have daily consumer-level supermarket scanner data with information on consumer purchasing decisions of some product.

• Y_{it} is the indicator for "consumer i purchases the product at period t". Model:

$$Y_{it} = \mathbf{1}\{-P_t + \pi(X_{it}) - \varepsilon_{it} \geq 0\}$$

where P_t is the price of the product at day t; $\pi(X_{it}) - \varepsilon_{it}$ represents the consumer willingness to pay that depend on observable and unobservable consumer characteristics.

• Key assumptions: independence between ε_{it} and P_t; and P_t has continuous variation.

Relaxing the linearity of the payoff function in Z

• This restriction can be quite strong in some empirical applications. In some applications, it implies that agents are risk neutral (see below the example on retirement).

• We can relax this restriction in the following semiparametric model:

$$Y \;=\; \mathbf{1}\left\{Z+\beta_2\, Z^2+...+\beta_q\, Z^q+\pi(X)-\varepsilon\geq 0\right\}$$

where $\beta' s$ are unknown parameters and $Z+\beta_2\, Z^2+...+\beta_q\, Z^q$ is strictly increasing in Z.

Relaxing the linearity of the payoff function in Z [2]

Theorem. Consider the model $Y = \mathbf{1}\left\{Z + \beta_2\, Z^2 + ... + \beta_q\, Z^q + \pi(X) - \varepsilon \geq 0\right\}$, and assume that:

(a) ε is independent of Z and X and has median zero;

(b) X has at least two $q+1$ points in its support set;

(c) and (d) from previous Theorem.

Then, $\{\pi, F_\varepsilon, \beta\}$ are NP identified.

Relaxing the linearity of the payoff function in Z [3]

• **Proof** (Sketch): For any $p \in (0,1)$, define $z_p^*(x_0)$ the unique solution in z to $P(z, x_0) = p$. Then,

$$z_p^*(x_0) + \beta_2 \, z_p^*(x_0)^2 + ... + \beta_q \, z_p^*(x_0)^q + \pi(x_0) = Q_\varepsilon(p)$$

where $Q_\varepsilon(p)$ is the quantile function of ε.

• Given p and p', with $p \neq p'$:

$$\left[z_p^*(x_0) - z_{p'}^*(x_0)\right] =$$

$$Q_\varepsilon(p) - Q_\varepsilon(p') + \beta_2 \left[z_{p'}^*(x_0)^2 - z_p^*(x_0)^2\right] + ... + \beta_q \left[z_{p'}^*(x_0)^q - z_p^*(x_0)^q\right]$$

And given $q+1$ different values for x_0, we have a system of $q+1$ equations and $q+1$ unknowns that identifies β's.

• The identification of π and β proceeds in the same way as in previous Theorem.

Example 2 (Retirement from the labor force)

• We have panel data where we observe individual decision of retiring (collecting pension benefits) or keeping working, and their earnings if working (salary) or if retired (pension benefits).

• Y_{it} is the indicator for "individual i retires at period t". Model:

$$Y_{it} = \mathbf{1}\left\{ \begin{array}{c} [B_{it} - W_{it}] + \beta_2 \left[B_{it}^2 - W_{it}^2\right] + ... + \beta_q \left[B_{it}^q - W_{it}^q\right] \\ +\pi(X_{it}) - \varepsilon_{it} \geq 0 \end{array} \right\}$$

where W_{it} and B_{it} represent earnings when working and retired, resp; and $\pi(X_{it}) - \varepsilon_{it}$ represents the additional non-pecuniary utility from being retired, that depends on observable and unobservable characteristics.

• Key assumption: independence between ε_{it} and B_{it}, W_{it}, and X_{it}.

PARTIAL IDENTIFICATION

• In some empirical applications Z and X are discrete. This implies that the distribution of the unobservables cannot be point-identified. It is still possible to obtain **informative bounds on π and the distribution function**.

• Consider the model: $Y = \mathbf{1}\{Z + \pi(X) - \varepsilon \geq 0\}$, where both Z and X have discrete and finite supports. We maintain the same assumptions as above.

• For arbitrary x_0, define:

$$z^{+}(x_0) \equiv \inf_{z \in \mathcal{Z}} \left[\max \left\{ P(z, x_0) \; ; \frac{1}{2} \right\} \right]$$

$$z^{-}(x_0) \equiv \sup_{z \in \mathcal{Z}} \left[\min \left\{ P(z, x_0) \; ; \frac{1}{2} \right\} \right]$$

With discrete support, we have that, in general: $z^{+}(x_0) \geq z^{-}(x_0)$

PARTIAL IDENTIFICATION [2]

• By construction, we have that:

$$\begin{aligned} z^{+}(x_0) + \pi(x_0) &\geq 0 \\ z^{-}(x_0) + \pi(x_0) &\leq 0 \end{aligned}$$

such that:

$$\pi(x_0) \in \left[-z^{+}(x_0) \ , \ - z^{-}(x_0) \right]$$

• This interval provides the sharp bounds for the identification of $\pi(x_0)$.

• We can denote these bounds $\pi^L(x_0)$ (that is equal to $-z^+(x_0)$) and $\pi^H(x_0)$ (that is equal to $-z^-(x_0)$).

PARTIAL IDENTIFICATION [3]

- Given a pair (x_0, ε_0), we can construct the values:

$$\widetilde{z}^H(x_0, \varepsilon_0) \equiv \varepsilon_0 - \pi^L(x_0) \;=\; \varepsilon_0 + z^+(x_0)$$

$$\widetilde{z}^L(x_0, \varepsilon_0) \equiv \varepsilon_0 - \pi^H(x_0) \;=\; \varepsilon_0 + z^-(x_0)$$

- By construction, we have that $\widetilde{z}^H(x_0, \varepsilon_0) \geq \widetilde{z}^L(x_0, \varepsilon_0)$ and:

$$\begin{aligned} P\left[\widetilde{z}^H(x_0, \varepsilon_0), x_0\right] &= F_{\varepsilon|x_0}\left[\widetilde{z}^H(x_0, \varepsilon_0) + \pi(x_0)\right] \\ &= F_{\varepsilon|x_0}\left[\varepsilon_0 + \left[\pi(x_0) - \pi^L(x_0)\right]\right] \\ &\geq F_{\varepsilon|x_0}(\varepsilon_0) \end{aligned}$$

$$\begin{aligned} P\left[\widetilde{z}^L(x_0, \varepsilon_0), x_0\right] &= F_{\varepsilon|x_0}\left[\widetilde{z}^L(x_0, \varepsilon_0) + \pi(x_0)\right] \\ &= F_{\varepsilon|x_0}\left[\varepsilon_0 + \left[\pi(x_0) - \pi^H(x_0)\right]\right] \\ &\leq F_{\varepsilon|x_0}(\varepsilon_0) \end{aligned}$$

- Therefore,

$$F_{\varepsilon|x_0}(\varepsilon_0) \in \left[P\left[\widetilde{z}^L(x_0, \varepsilon_0), x_0 \right] \ , \ P\left[\widetilde{z}^H(x_0, \varepsilon_0), x_0 \right] \right]$$

This interval provides the sharp bounds for the identification of $F_{\varepsilon|x_0}(\varepsilon_0)$.

1.2. ESTIMATION METHODS

- I'll discuss the following estimation methods:

  (a) A simple Kernel method for the (just-identified) NP model

  (b) Two-step method for semiparametric model

  (c) Lewbel's method

  (d) Klein-Spady method

(a) A Simple Kernel Estimator for the (just-identified) NP model

- The constructive proof of identification provides a simple estimator.

- **[Step 1]** We estimate $P(z,x)$ using a Kernel estimator:

$$\widehat{P}(z,x) = \frac{\sum_{i=1}^{N} y_i \, \mathbf{1}\{x_i = x\} \;\; K\left(\frac{z_i - z}{b}\right)}{\sum_{i=1}^{N} \mathbf{1}\{x_i = x\} \;\; K\left(\frac{z_i - z}{b}\right)}$$

Note: Imposing monotonicity in z is very important. If the kernel method does not satisfy monotonicity at some values of z, then we need Isotonic-Kernel methods

- **[Step 2]** Newton's method to obtain $\widehat{z^*}(x_0)$ as the unique solution in z to $\widehat{P}(z,x_0) = 1/2$.

• **[Step 3]** Estimate distribution as:

$$\widehat{F}_{\varepsilon|x_0}(\varepsilon_0) = \widehat{P}\left(\varepsilon_0 + \widehat{z^*}(x_0), x_0\right) = \frac{\displaystyle\sum_{i=1}^{N} y_i \ \mathbf{1}\{x_i = x_0\} \ K\left(\frac{z_i - \left[\varepsilon_0 + \widehat{z^*}(x_0)\right]}{b}\right)}{\displaystyle\sum_{i=1}^{N} \mathbf{1}\{x_i = x_0\} \ K\left(\frac{z_i - \left[\varepsilon_0 + \widehat{z^*}(x_0)\right]}{b}\right)}$$

• For this just-identified NP model, this estimator exploits all the restrictions of the model.

(b) Two-step methods for semiparametric model

• In many applications without very large sample sizes or/and relatively large number of variables X, it can be impractical to estimate with enough precision a model that is nonparametric in both F and π.

• The researcher may be willing to consider a parametric model for $\pi(X)$ and to impose some restrictions about how the distribution of ε depends on X.

• Common restrictions are:

(a) $\pi(X) = X'\beta$

(b) $\varepsilon = \sigma(X)\,\widetilde{\varepsilon}$, where $\widetilde{\varepsilon}$ is independent of X

(b) Two-step method for semiparametric model

• First step consists of an estimator of β that is robust to the specification of F_ε.

• For instance, we can project the NP estimator of $\pi(X)$ above on the linear space $X'\beta$ to estimate β by OLS.

• Other estimators are Manski's Maximum Score estimator (MSE) and Horowitz's Smooth MSE. These estimators require optimization with respect to β.

• Lewbel's method does not require optimization.

(c) Lewbel's method

• Lewbel shows that in this BC model:

$$\beta = \left[\mathbb{E}\left(X\ X'\right)\right]^{-1} \mathbb{E}\left(X\ \widetilde{Y}\right)$$

where $\widetilde{Y} = \dfrac{Y - \mathbf{1}\{Z > 0\}}{f_{Z|X}(Z|X)}$.

• This expression shows that we can estimate consistently β by an OLS regression of $\widetilde{Y}$ on X. Variable $\widetilde{Y}$ should be constructed and requires estimating the density $f_{Z|X}(Z|X)$.

• Since the density $f_{Z|X}(Z|X)$ appears in the denominator, $\sqrt{N}-$consistency of the estimator (and good finite sample properties) requires trimming observations where $\widehat{f}_{Z|X}(z_i|x_i) < h_N$.

(d) Klein-Spady method

• Klein & Spady (ECMA 1993) propose an asymptotically efficient method to estimate jointly β and the CDF of ε. However, an important restriction of their model/method is that $Var(\varepsilon|Z, X) = \sigma^2(Z + X'\beta)$, i.e., the conditional variance depends on Z and X but only through the index $Z + X'\beta$.

• Under this restriction, $P(Z, X) = F_\varepsilon\left(\frac{Z + X'\beta}{\sigma(Z + X'\beta)}\right) = G(Z + X'\beta)$.

• They propose a **semiparametric maximum likelihood estimator** of β and the function $G(.)$. The log-likelihood function is:

$$l(\beta, G) = \sum_{i=1}^{n} y_i \; \ln G\left(x_i'\beta\right) + (1 - y_i) \; \ln\left[1 - G\left(x_i'\beta\right)\right]$$

• And KS estimator is defined as:

$$\left(\widehat{\beta}_{KS}, \widehat{G}_{KS}\right) = \arg\max_{\{\beta, G\}} \; l(\beta, G)$$

- Let $\widehat{\beta}_0$ be an initial consistent estimator of β, e.g., Lewbel's estimator.

Step 1: Given that $G(\varepsilon_0) = \mathbb{E}(Y \mid Z + X'\beta = \varepsilon_0)$, we estimate $G(\varepsilon_0)$ using a Kernel regression of y_i on $z_i + x_i'\widehat{\beta}_0$:

$$\widehat{G}_1(\varepsilon_0) = \frac{\sum_{i=1}^{n} y_i \; K\left(\frac{z_i + x_i'\widehat{\beta}_0 - \varepsilon_0}{b_n}\right)}{\sum_{i=1}^{n} K\left(\frac{z_i + x_i'\widehat{\beta}_0 - \varepsilon_0}{b_n}\right)}$$

Step 2: Obtain a new $\widehat{\beta}$ as:

$$\widehat{\beta}_1 = \arg\max_{\beta} \;\; l\left(\beta, \widehat{G}_1\right)$$

- The algorithm iterates in steps 1 &2 until convergence in $\left\|\widehat{\beta}_K - \widehat{\beta}_{K-1}\right\|$.

2. DYNAMIC BINARY CHOICE MODELS

• $Y_{it} \in \{0, 1\}$ is the decision of agent i at period t. The one-period payoff is:*

$$\begin{cases} Z_{it} + \pi_t(X_{it}) - \varepsilon_{it} & \text{if} \quad Y_{it} = 1 \\ \\ 0 & \text{if} \quad Y_{it} = 0 \end{cases}$$

• The choice at period t has implications on future profits. State variables (Z_{it}, X_{it}) follow a controlled first order Markov process that depends on the choice variable:

$$\Pr(Z_{it+1}, X_{it+1} \mid Y_{it}, Z_{it}, X_{it}) \equiv f_{Z,X,t}(Z_{it+1}, X_{it+1} \mid Y_{it}, Z_{it}, X_{it})$$

• The unobservable ε_{it} can have a distribution that changes over time, but it is independently distributed over time:*

$$\varepsilon_{it} \text{ is } i.i.d. \text{ over } i \text{ and indep. over } t \text{ with CDF } F_t(.)$$

- Agent chooses Y_{it} to maximize expected & discounted intertemp. payoff:

$$\mathbb{E}\left(\sum_{s=t}^{T} \beta^{s-t}\, Y_{is}\ \left[Z_{is} + \pi_t(X_{is}) - \varepsilon_{is}\right]\ \mid Y_{it}, Z_{it}, X_{it}\right)$$

where $\beta \in [0,1)$ is the discount factor.

- The optimal choice of agent i at period t can be represented as:

$$\begin{aligned} Y_{it} &= \mathbf{1}\left\{Z_{it} + \pi_t(X_{it}) - \varepsilon_{it} + v_t\left(1, Z_{it}, X_{it}\right) \geq v_t\left(0, Z_{it}, X_{it}\right)\right\} \\ &= \mathbf{1}\left\{\varepsilon_{it} \leq Z_{it} + \pi_t(X_{it}) + v_t\left(1, Z_{it}, X_{it}\right) - v_t\left(0, Z_{it}, X_{it}\right)\right\} \end{aligned}$$

- $v_t\left(Y, Z_{it}, X_{it}\right)$ is the present value of future payoffs if current choice is Y.

• By Bellman's principle, this value function has a recursive structure:

$$v_t(Y, Z_{it}, X_{it}) =$$

$$\beta\ \mathbb{E}_t \left(\max \left\{ \begin{array}{c} Z_{it+1} + \pi_{t+1}(X_{it+1}) - \varepsilon_{it+1} + v_{t+1}(1, Z_{it}, X_{it}) \\ \\ ;\ v_{t+1}(0, Z_{it+1}, X_{it+1}) \end{array} \right\} \right)$$

• According to the model, the conditional choice probability (CCP) function at period t is:

$$P_t(Z_{it}, X_{it}) = F_t\left[Z_{it} + \pi_t(X_{it}) + v_t(1, Z_{it}, X_{it}) - v_t(0, Z_{it}, X_{it})\right]$$

• The primitives or structural parameters of the mode are:

$$\{f_{Z,X,t}, \pi_t, F_t, \beta : t = 1, 2, ..., T\}$$

IDENTIFICATION OF DYNAMIC MODEL

• Suppose that we have panel data on choices and state variables:

$$\{Y_{it}, Z_{it}, X_{it} : i = 1, 2, ..., N;\ t = 1, 2, ..., T_{data}\}$$

where T_{data} may be different than the T of the model.

• We are interested in using these data to identify $\{f_{Z,X,t},\ \pi_t,\ \sigma_t,\ F_{\widetilde{\varepsilon}},\ \beta\}$.

• Under mild regularity conditions, the CCP function $P_t\left(Z_{it}, X_{it}\right)$ and the transition probability functions $f_{Z,X,t}(Z_{it+1}, X_{it+1} \mid Y_{it}, Z_{it}, X_{it})$ are non-parametrically identified from the data. Therefore, the relevant problem is the identification of payoff functions and distribution of the unobservables, $\{\pi_t,\ \sigma_t,\ F_{\widetilde{\varepsilon}},\ \beta\}$, using CCPs and transitions.

IDENTIFICATION OF DYNAMIC MODEL [2]

- In the identification of these models, it is useful to distinguish two cases:

  (a) Finite horizon models where the researcher observes agents' decision at the last period: $T_{data} = T$.

  (b) Applications where $T_{data} < T$ (either because infinite horizon models or because $T_{data} < T < \infty$).

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} = \mathbf{T}$

- For simplicity, but really w.l.o.g. suppose that $T_{data} = T = 2$.

- The model can be described in two equations. Optimal decision at last period:

$$Y_2 = \mathbf{1}\{\varepsilon_2 \leq Z_2 + \pi_2(X_2)\}$$

And the optimal decision at period 1,

$$Y_1 = \mathbf{1}\{\varepsilon_1 \leq Z_1 + \pi_1(X_1) + v_1(1, Z_1, X_1) - v_1(0, Z_1, X_1)\}$$

where

$$v_1(Y_1, Z_1, X_1) = \beta\, \mathbb{E}\left(\max\{Z_2 + \pi_2(X_2) - \varepsilon_2\,;\ 0\} \mid Y_1, Z_1, X_1\right)$$

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} = \mathbf{T}$ [2]

• Under similar conditions as those in Matzkin's Theorem, we can apply a recursive argument to show the NP identification of the functions $\{\pi_1$, π_2, F_1, $F_2\}$ when the discount factor β is known to the researcher.

Theorem [Aguirregabiria (JBES, 2010)]. Suppose that, for $t = 1, 2$,

(a) ε_t is independent of Z_t; and ε_1 and ε_2 are independent;

(b) ε_t has zero median and is median independent of X_t;

(c) Function $\pi_t(.)$ is bounded on $\mathcal{X}_t$, i..e., $\pi_t(\mathcal{X}_t) \in [\pi_t^L, \pi_t^H] \subset \mathbb{R}$;

(d) Conditional on any value of X_t, variable Z_t has support over the whole real line $\mathbb{R}$;

(e) The differential value function $\widetilde{v}_1 (Z_1, X_1) \equiv v_1 (1, Z_1, X_1) - v_1 (0, Z_1, X_1)$ is non-decreasing in Z_1 [it can weaken to $\partial \widetilde{v}_1 (Z_1, X_1) / \partial Z_1 > -1$;

(f) The discount factor β is known to the researcher.

Then, $\{\pi_t(x_t) : t = 1, 2\}$ and $\{F_{t,x_t}(\varepsilon_t) : t = 1, 2\}$ are NP identified for every $x_t \in \mathcal{X}_t$ and $\varepsilon_t \in \mathbb{R}$.

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} = \mathbf{T}$ [3]

• The model at $t = 2$ is a static model, and therefore identification of $\pi_2(.)$ and $F_{\varepsilon_2|X_2}(.)$ follows form Matzkin's Theorem. Now, define the function:

$$\begin{aligned} e_2(Z_2, X_2) &\equiv \mathbb{E}_{\varepsilon_2}\left(\max\left\{Z_2 + \pi_2(X_2) - \varepsilon_2\,;\ 0\right\}\right) \\ &= \int_{-\infty}^{Z_2+\pi_2(X_2)} \left[Z_2 + \pi_2(X_2) - u\right]\ dF_{\varepsilon_2|X_2}(u) \end{aligned}$$

• Since $\pi_2(.)$ and $F_{\varepsilon_2|X_2}(.)$ are identified, it is clear that function $e_2(Z_2, X_2)$ is NP identified everywhere in the support of (Z_2, X_2).

• By definition, we have that

$$v_1(Y_1, Z_1, X_1) = \beta\, \mathbb{E}\left(e_2(Z_2, X_2) \mid Y_1, Z_1, X_1\right)$$

Since $e_2(Z_2, X_2)$ is identified everywhere, it is clear that conditional expectation function $\mathbb{E}\left(e_2(Z_2, X_2) \mid Y_1, Z_1, X_1\right)$ is NP identified everywhere in the support of (Y_1, Z_1, X_1). Then, for β known, the value function $v_1(Y_1, Z_1, X_1)$ is identified everywhere.

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} = \mathbf{T}$ [4]

- Then, we have:

$$P_1(Z_1, X_1) = F_{\varepsilon_1|X_1}\left[Z_1 + \pi_1(X_1) + \widetilde{v}_1(Z_1, X_1)\right]$$

Conditions (a) and (e) imply that $P_1(Z_1, X_1)$ is strictly increasing in Z_1. Therefore, for any value $X_1 = x_1$ we have that there exist a unique value Z_1 that solves $P_1(Z_1, x_1) = 1/2$. Let $z_1^*(x_1)$ be the solution of that equation.

- By the zero median of ε_1, we have that:

$$\pi_1(x_1) = -z_1^*(x_1) - \widetilde{v}_1(z_1^*(x_1), x_1)$$

such that $\pi_1(.)$ is identified everywhere in $\mathcal{X}_1$.

- For any pair $(x_1, \varepsilon_1) \in \mathcal{X}_1 \times \mathbb{R}$, consider the following equation in Z_1:

$$Z_1 + \pi_1(x_1) + \widetilde{v}_1(Z_1, x_1) = \varepsilon_1$$

For any (x_1, ε_1) this equation has a always a solution and the solution is unique.

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} = \mathbf{T}$ [5]

• Let $\widetilde{z}_1(x_1, \varepsilon_1)$ be that solution such that $\widetilde{z}_1(x_1, \varepsilon_1) + \pi_1(x_1) + \widetilde{v}_1\left(\widetilde{z}_1(x_1, \varepsilon_1), x_1\right) = \varepsilon_1$. By construction,

$$\begin{aligned} F_{\varepsilon_1|x_1}(\varepsilon_1) &= F_{\varepsilon_1|x_1}\left[\widetilde{z}_1(x_1, \varepsilon_1) + \pi_1(x_1) + \widetilde{v}_1\left(\widetilde{z}_1(x_1, \varepsilon_1), x_1\right)\right] \\ &= P_1\left[\widetilde{z}_1(x_1, \varepsilon_1), x_1\right] \end{aligned}$$

Therefore, $F_{\varepsilon_1|x_1}(\varepsilon_1)$ is NP identified on $\mathcal{X}_1 \times \mathbb{R}$. ■

• **Remark 1:** This argument can be applied recursively to prove the NP identification of $\pi_t(.)$ and $F_{\varepsilon_t|x_t}(.)$ at every period t in the sample.

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} = \mathbf{T}$ [6]

• **Remark 2:** Note that we need a stronger condition on the support of Z_t: this variable should have support over the whole real line.

• The main reason is the identification of the function:

$$e_2(Z_2, X_2) = \int_{-\infty}^{Z_2+\pi_2(X_2)} [Z_2 + \pi_2(X_2) - u] \ dF_{\varepsilon_2|X_2}(u)$$

• To obtain this function, we need to know the whole left-tail of the distribution of $F_{\varepsilon_2|X_2}(u)$. This can be problematic because identification of distribution tails can be very imprecise.

• A possible "solution" is imposing the restriction that ε_2 has bounded support.

IDENTIFICATION WITH $\mathsf{T}_{\mathsf{data}} < \mathsf{T}$ [Based on Aguirregabiria & Tang, 2017]

- To study identification in this case, it is convenient to present an Euler equation representation of the optimal decision in this model.

- The intuition behind the Euler equation is quite simple: at the optimal solution, it is not possible to perturb marginally the CCPs P_t and P_{t+1} to improve expected intertemporal values.

- The particular form of the Euler equations depends on which are the endogenous state variables of the problem and on their transition probabilities. For the sake of concreteness, I consider here I simple model where the endogenous state variables is the lagged decision, Y_{t-1}, e.g., market entry-exit model.

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} < \mathbf{T}$ [2]

• The payoff function at period t is:

$$\begin{cases} Z_t + \pi_t(X_t, Y_{t-1}) - \varepsilon_t & \text{if} \quad Y_t = 1 \\ \\ 0 & \text{if} \quad Y_t = 0 \end{cases}$$

where now I make explicit the endogenous state variable Y_{t-1} in the payoff function. Note that now (Z_t, X_t) are exogenous state variables.

• The Euler equation for this entry-exit model is (see Aguirregabiria and Magesan, 2013, 2016):

$$\begin{aligned} Z_t + \pi_t(X_t, Y_{t-1}) + e_t(P_t \,|X_t, Y_{t-1}) \quad &+ \\ \beta \, \mathbb{E}_t \left[Z_{t+1} + \pi_{t+1}(X_{t+1}, 1) + e_{t+1}(P_{t+1} \,|X_{t+1}, 1) \right] \quad &= \\ \beta \, \mathbb{E}_t \left[Z_{t+1} + \pi_{t+1}(X_{t+1}, 0) + e_{t+1}(P_{t+1} \,|X_{t+1}, 0) \right] \quad & \end{aligned}$$

IDENTIFICATION WITH $\mathbf{T}_{\mathsf{data}} < \mathbf{T}$ [3]

• The function $e_t(p \,|X_t, Y_{t-1})$ is defined as follows: for any probability p:

$$e_t(p|X_t, Y_{t-1}) \;\equiv\; \mathbb{E}_{\varepsilon_t}\left[\, \varepsilon_t \mid \varepsilon_t \leq Q_{\varepsilon_t|X_t,Y_{t-1}}(p)\right]$$

and $Q_{\varepsilon_t|X_t,Y_{t-1}}(p)$ is the quantile associated to the distribution $F_{\varepsilon_t|X_t,Y_{t-1}}(\varepsilon_t)$, i.e., the inverse function of $F_{\varepsilon_t|X_t,Y_{t-1}}$.

• It is straightforward to show that this function is the *Integrated Quantile Function* (IQF) of the distribution $F_{\varepsilon_t|X_t,Y_{t-1}}(\varepsilon_t)$.

• In general, if a function $e(p)$ on $[0, 1]$ is defined as $\mathbb{E}\,[\,\varepsilon \mid \varepsilon \leq Q(p)]$ where $Q(p)$ is the quantile function of the distribution of ε, then:

$$\frac{d\ e(p)}{dp} = Q(p)$$

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} < \mathbf{T}$ [4]

• To identify this model, we need to impose a time-homogeneity assumption on the distribution of ε_t. The distribution of ε_t may depend on (X_t, Y_{t-1}), but conditional on these variables the distribution is time invariant.

• Here I present an identification result for a simplified version of the model where all the primitive functions are time-homogeneous and the distribution of ε_t does not depend on (X_t, Y_{t-1}) such that it is $i.i.d.$

• This simplified version of the model has the following Euler equation:

$$
\begin{aligned}
Z_t + \pi(X_t, Y_{t-1}) + e(P[Z_t, X_t, Y_{t-1}]) &\ + \\
\beta\, \mathbb{E}_t\left[\pi(X_{t+1}, 1) + e(P[Z_{t+1}, X_{t+1}, 1])\right] &\ = \\
\beta\, \mathbb{E}_t\left[\pi(X_{t+1}, 0) + e(P[Z_{t+1}, X_{t+1}, 0])\right] &
\end{aligned}
$$

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} < \mathbf{T}$ [5]

• Here I show identification for a particular specification for the stochastic process of $\{Z_t, X_t\}$. The result can be easily extended, though the expressions are more complicated.

• Suppose that the stochastic process of $\{Z_t, X_t\}$ is such that:

(a) [Conditional independence]

$$f_{Z,X}(Z_{t+1}, X_{t+1}|Z_t, X_t) = f_Z(Z_{t+1}|Z_t)\ f_X(X_{t+1}|X_t)\ ;$$

(b) [Autoregressive Z_t] $Z_{t+1} = \rho(Z_t) + U_{t+1}$, with U_{t+1} i.i.d. and independent of Z_t. Function $\rho(Z_t)$ is continuously differentiable and $\rho'(Z_t) \equiv d\rho(Z_t)/dZ_t$ is such that $|\rho'(Z_t)| < 1$.

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} < \mathbf{T}$ [6]

- Under these conditions, we have that:

$$\mathbb{E}_t \left[e(P[Z_{t+1}, X_{t+1}, Y]) \right] =$$

$$\int\limits_{X_{t+1}, u_{t+1}} e(P[\rho(Z_t) + U_{t+1}, X_{t+1}, Y]) \ f(U_{t+1}) \ f_X(X_{t+1}|X_t) dU_{t+1} dX_{t+1}$$

- This function is continuous differentiable in Z_t and it is simple to show that:

$$\frac{\partial \ \mathbb{E}_t \left[e(P[Z_{t+1}, X_{t+1}, Y]) \right]}{\partial \ Z_t} = \rho(Z_t) \ \mathbb{E}_t \left[Q_\varepsilon \left(P_{t+1} \right) \ \frac{\partial P_{t+1}}{\partial Z_{t+1}} \right]$$

Note also $\dfrac{\partial \ e(P_t)}{\partial Z_t} = Q_\varepsilon \left(P_t \right) \dfrac{\partial P_t}{\partial Z_t}$.

IDENTIFICATION WITH $\mathbf{T}_{\text{data}} < \mathbf{T}$ [7]

Theorem [Aguirregabiria and Tang (2017)]. Under conditions (a) and (b), differencing the Euler equation with respect to Z_t implies a contraction mapping in the space of the quantile function $Q_\varepsilon(P)$. This contraction mapping uniquely identifies the distribution of ε.

Proof: Differentiating the Euler equation with respect to Z_t, we can get the following expression:

$$Q_\varepsilon(P_t) = \left[-\frac{\partial P_t}{\partial Z_t}\right]^{-1} \left\{1 + \beta\ \rho(Z_t)\ \mathbb{E}_t \left[Q_\varepsilon(P_{t+1}(1))\ \frac{\partial P_{t+1}(1)}{\partial Z_{t+1}} \right.\right.$$

$$\left.\left. - Q_\varepsilon(P_{t+1}(0))\ \frac{\partial P_{t+1}(0)}{\partial Z_{t+1}}\right]\right\}$$

• All the elements in this expression, except the quantile function $Q_\varepsilon(.)$ are known to the researcher.

• This mapping is a contraction in Q_ε. Therefore, $Q_\varepsilon(.)$ is uniquely identified.

Further references:

Aguirregabiria (JBES, 2010)

Norets and Tang (REStud, 2013)

Blevins (QE, 2014)

Buchholz, Shum, and Hu (Working Paper, 2016)

Aguirregabiria and Tang (Working Paper, 2017)