Identification and Estimation of Nonstationary Dynamic Binary Choice Models

Cheng Chou — Univ. of Leicester
Geert Ridder — USC
Ruoyao Shi — UC Riverside

California Econometrics Conference (U Washington)
September 16, 2023

Estimation of Dynamic Discrete Choice (DDC) Models

- DDC models are widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, etc.
- Estimation:
  - Solve the full dynamic programming problem $\rightarrow$ MLE (Rust, 1987).
  - Match conditional choice probabilities (CCPs) $\rightarrow$ GMM (Hotz & Miller, 1993; Hotz et al., 1994).
- Disadvantages:
  - **Implementation**: both require estimating & simulating from state transitions, except special cases (e.g., terminal choice, renewal choice).
  - **Robustness**: NP state transitions can be difficult $\rightarrow$ parametric specification is often assumed.

Estimation of Dynamic Discrete Choice (DDC) Models

- DDC models are widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, etc.
- Estimation:
  - Solve the full dynamic programming problem $\rightarrow$ MLE (Rust, 1987).
  - Match conditional choice probabilities (CCPs) $\rightarrow$ GMM (Hotz & Miller, 1993; Hotz et al., 1994).
- Disadvantages:
  - **Implementation**: both require estimating & simulating from state transitions, except special cases (e.g., terminal choice, renewal choice).
  - **Robustness**: NP state transitions can be difficult $\rightarrow$ parametric specification is often assumed.

Estimation of Dynamic Discrete Choice (DDC) Models

- DDC models are widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, etc.
- Estimation:
  - Solve the full dynamic programming problem $\rightarrow$ MLE (Rust, 1987).
  - Match conditional choice probabilities (CCPs) $\rightarrow$ GMM (Hotz & Miller, 1993; Hotz et al., 1994).
- Disadvantages:
  - **Implementation**: both require estimating & simulating from state transitions, except special cases (e.g., terminal choice, renewal choice).
  - **Robustness**: NP state transitions can be difficult $\rightarrow$ parametric specification is often assumed.

This Paper: ID & Est. of Nonstationary DBC Model

- We derive a **Markovian** property for observed state variables under typical assumptions in DDC literature (Aguirregabiria & Mira, 2010).
- It allows us to:
  - simplify the log odds ratio equation of a general nonstationary DBC model, **eliminating** the necessity of dealing with the state transitions;
  - transform log odds ratio equation of the DBC model into a linear system under mild additional assumptions;
  - find clear sufficient identification conditions for flow utility parameters;
  - develop a CCP-based semiparametric estimator that avoids estimating & simulating from state transitions; and
  - quantify & reduce sensitivity of estimate to the additional assumption.

This Paper: ID & Est. of Nonstationary DBC Model

- We derive a **Markovian** property for observed state variables under typical assumptions in DDC literature (Aguirregabiria & Mira, 2010).
- It allows us to:
  - simplify the log odds ratio equation of a general nonstationary DBC model, **eliminating** the necessity of dealing with the state transitions;
  - transform log odds ratio equation of the DBC model into a **linear** system under mild additional assumptions;
  - find clear sufficient **identification** conditions for flow utility parameters;
  - develop a CCP-based semiparametric estimator that **avoids** estimating & simulating from state transitions; and
  - quanfity & reduce sensitivity of estimate to the additional assumption.

This Paper: ID & Est. of Nonstationary DBC Model

- We derive a **Markovian** property for observed state variables under typical assumptions in DDC literature (Aguirregabiria & Mira, 2010).
- It allows us to:
  - simplify the log odds ratio equation of a general nonstationary DBC model, **eliminating** the necessity of dealing with the state transitions;
  - transform log odds ratio equation of the DBC model into a **linear** system under mild additional assumptions;
  - find clear sufficient **identification** conditions for flow utility parameters;
  - develop a CCP-based semiparametric estimator that **avoids** estimating & simulating from state transitions; and
  - quanfity & reduce sensitivity of estimate to the additional assumption.

This Paper: ID & Est. of Nonstationary DBC Model

- We derive a **Markovian** property for observed state variables under typical assumptions in DDC literature (Aguirregabiria & Mira, 2010).
- It allows us to:
  - simplify the log odds ratio equation of a general nonstationary DBC model, **eliminating** the necessity of dealing with the state transitions;
  - transform log odds ratio equation of the DBC model into a **linear** system under mild additional assumptions;
  - find clear sufficient **identification** conditions for flow utility parameters;
  - develop a CCP-based semiparametric estimator that **avoids** estimating & simulating from state transitions; and
  - quanfity & reduce sensitivity of estimate to the additional assumption.

This Paper: ID & Est. of Nonstationary DBC Model

- We derive a **Markovian** property for observed state variables under typical assumptions in DDC literature (Aguirregabiria & Mira, 2010).
- It allows us to:
  - simplify the log odds ratio equation of a general nonstationary DBC model, **eliminating** the necessity of dealing with the state transitions;
  - transform log odds ratio equation of the DBC model into a **linear** system under mild additional assumptions;
  - find clear sufficient **identification** conditions for flow utility parameters;
  - develop a CCP-based semiparametric estimator that **avoids** estimating & simulating from state transitions; and
  - quanfity & reduce sensitivity of estimate to the additional assumption.

This Paper: ID & Est. of Nonstationary DBC Model

- We derive a **Markovian** property for observed state variables under typical assumptions in DDC literature (Aguirregabiria & Mira, 2010).
- It allows us to:
  - simplify the log odds ratio equation of a general nonstationary DBC model, **eliminating** the necessity of dealing with the state transitions;
  - transform log odds ratio equation of the DBC model into a **linear** system under mild additional assumptions;
  - find clear sufficient **identification** conditions for flow utility parameters;
  - develop a CCP-based semiparametric estimator that **avoids** estimating & simulating from state transitions; and
  - quanfity & reduce sensitivity of estimate to the additional assumption.

This Paper: ID & Est. of Nonstationary DBC Model

- We derive a **Markovian** property for observed state variables under typical assumptions in DDC literature (Aguirregabiria & Mira, 2010).
- It allows us to:
  - simplify the log odds ratio equation of a general nonstationary DBC model, **eliminating** the necessity of dealing with the state transitions;
  - transform log odds ratio equation of the DBC model into a **linear** system under mild additional assumptions;
  - find clear sufficient **identification** conditions for flow utility parameters;
  - develop a CCP-based semiparametric estimator that **avoids** estimating & simulating from state transitions; and
  - quanfity & reduce sensitivity of estimate to the additional assumption.

Outline

1. A Dynamic Binary Choice Model w. General Nonstationarity
   - Model Setup
   - Brief Recount of the Hotz & Miller Estimators
2. Our Approach
   - Transformation into a Linear System
   - Identification
   - Estimation
3. Concluding Remarks

Outline

Basic Dynamic Binary Choice Model Setup

- Each agent makes a binary choice $a_t \in \{0, 1\}$ in each period t.
- Decision horizon: $t \in \mathcal{T} \equiv \{T_{start}, \ldots, T_{end}\}$, $T_{end} = \infty$ allowed.
- Flow utility: $u_t(a_t, x_t) + \varepsilon_{a_t t}$, where
  - state variables $\Omega_t \equiv (s_t', \varepsilon_t')'$;
  - $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{0t})'$: unobserved flow utility shocks;
  - $s_t \equiv (x_t', z_t')'$: $(d_x + d_z) \times 1$ observed state variables;
  - x_t affects the current flow utility;
  - z_t are **excluded variables**, which do not affect current flow utility, but affect future payoff through impact on distribution of x_{t+1}.

Basic Dynamic Binary Choice Model Setup

- Each agent makes a binary choice $a_t \in \{0, 1\}$ in each period t.
- Decision horizon: $t \in \mathcal{T} \equiv \{T_{start}, \ldots, T_{end}\}$, $T_{end} = \infty$ allowed.
- Flow utility: $u_t(a_t, x_t) + \varepsilon_{a_t t}$, where
  - state variables $\Omega_t \equiv (s_t', \varepsilon_t')'$;
  - $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{0t})'$: unobserved flow utility shocks;
  - $s_t \equiv (x_t', z_t')'$: $(d_x + d_z) \times 1$ observed state variables;
  - x_t affects the current flow utility;
  - z_t are **excluded variables**, which do not affect current flow utility, but affect future payoff through impact on distribution of x_{t+1}.

Basic Dynamic Binary Choice Model Setup

- Each agent makes a binary choice $a_t \in \{0, 1\}$ in each period t.
- Decision horizon: $t \in \mathcal{T} \equiv \{T_{start}, \ldots, T_{end}\}$, $T_{end} = \infty$ allowed.
- Flow utility: $u_t(a_t, x_t) + \varepsilon_{a_t t}$, where
  - state variables $\Omega_t \equiv (s_t', \varepsilon_t')'$;
  - $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{0t})'$: unobserved flow utility shocks;
  - $s_t \equiv (x_t', z_t')'$: $(d_x + d_z) \times 1$ observed state variables;
  - x_t affects the current flow utility;
  - z_t are **excluded variables**, which do not affect current flow utility, but affect future payoff through impact on distribution of x_{t+1}.

We Maintain Common Assumptions in DDC Literature

Assumption 1 (Controlled Markov process)

Suppose $\Omega_{t+1} \perp\!\!\!\perp (\Omega_{t-j}, a_{t-j}) \mid (\Omega_t, a_t)$ *for* $j \in \mathbb{N}^+$.

Assumption 2 (Flow utility shocks)

Suppose (i) $\varepsilon_t \perp\!\!\!\perp s_t$*; (ii)* $\varepsilon_t \perp\!\!\!\perp s_{t-1}$*;*
(iii) ε_t *is serially independent; and*
(iv) $\varepsilon_{0t} \perp\!\!\!\perp \varepsilon_{1t}$*, and they both follow zero-mean standard type I extreme value distribution.*

Assumption 3 (Conditional independence)

Suppose $s_{t+1} \perp\!\!\!\perp \varepsilon_t \mid (s_t, a_t)$.

Standard Preliminary Results

- General nonstationarity is allowed in this paper:
  - $T_{end} = \infty$, $T_{end} \neq \infty$, or T_{end} unknown;
  - time-varying flow utility function $u_t(a, x)$;
  - time-varying state transition $f(\Omega_{t+1}|\Omega_t)$.
- Under Assumptions 1-3, the optimal decision rule leads to

$$\ln\left(\frac{p_t(s_t)}{1-p_t(s_t)}\right) = \underbrace{u_t(1,x_t) - u_t(0,x_t)}_{\text{static logit}} + \underbrace{\beta\Delta\mathbb{E}\left(\bar{V}_{t+1}(s_{t+1}) \mid s_t\right)}_{\text{main difficulty in dynamic model}},$$

where

  - $p_t(s) \equiv Pr.(a_t = 1|s_t = s)$ is the CCP;
  - $\bar{V}_{t+1}(s_{t+1})$ is the integrated value function;
  - $\Delta\mathbb{E}(h_\tau|s_t) \equiv \mathbb{E}(h_\tau|s_t, a_t = 1) - \mathbb{E}(h_\tau|s_t, a_t = 0)$ for $\tau > t$.

Standard Preliminary Results

- General nonstationarity is allowed in this paper:
  - $T_{end} = \infty$, $T_{end} \neq \infty$, or T_{end} unknown;
  - time-varying flow utility function $u_t(a, x)$;
  - time-varying state transition $f(\Omega_{t+1} | \Omega_t)$.
- Under Assumptions 1-3, the optimal decision rule leads to

$$\ln\left(\frac{p_t(s_t)}{1 - p_t(s_t)}\right) = \underbrace{u_t(1, x_t) - u_t(0, x_t)}_{\text{static logit}} + \underbrace{\beta \Delta \mathbb{E}\left(\bar{V}_{t+1}(s_{t+1}) \mid s_t\right)}_{\text{main difficulty in dynamic model}},$$

where

- $p_t(s) \equiv Pr.(a_t = 1 | s_t = s)$ is the CCP;
- $\bar{V}_{t+1}(s_{t+1})$ is the integrated value function;
- $\Delta \mathbb{E}(h_\tau | s_t) \equiv \mathbb{E}(h_\tau | s_t, a_t = 1) - \mathbb{E}(h_\tau | s_t, a_t = 0)$ for $\tau > t$.

Brief Recount of the Hotz & Miller Estimators

The HM estimators tackle it based on the **iteration** of the observation:

$$\bar{V}_{t+1}(s_{t+1}) = U^o_{t+1}(s_{t+1}) + \beta\mathbb{E}\left(\bar{V}_{t+2}(s_{t+2}) \mid s_{t+1}\right),$$

where $U^o_{t+1}(s_{t+1})$ is expected optimal flow utility, leading to

$$\begin{aligned}
&\ln\left(\frac{p_t(s_t)}{1-p_t(s_t)}\right) \\
= \; & u_t(1,x_t) - u_t(0,x_t) + \beta\Delta\mathbb{E}(U^o_{t+1}(s_{t+1})|s_t) \\
&+ \beta^2\Delta\mathbb{E}(\mathbb{E}(U^o_{t+2}(s_{t+2})|s_{t+1})|s_t) + \cdots \\
&+ \beta^{T_{tr}-t-1}\Delta\mathbb{E}(\mathbb{E}(\cdots\mathbb{E}(U^o_{T_{tr}-1}(s_{T_{tr}-1})|s_{T_{tr}-2})\cdots|s_{t+1})|s_t) \\
&+ \beta^{T_{tr}-t}\Delta\mathbb{E}(\mathbb{E}(\cdots\mathbb{E}(\mathbb{E}(\bar{V}_{T_{tr}}(s_{T_{tr}})|s_{T_{tr}-1})|s_{T_{tr}-2})\cdots|s_{t+1}|s_t),
\end{aligned}$$

where $T_{tr} \leq T_{end}$ is some truncation period.

Brief Recount of the Hotz & Miller Estimators (cont'd)

1. To evaluate the RHS under **hypothesized** parameter values, the HM estimators **estimate**, **simulate from**, and **integrate over** state transitions distributions $f_{s_{t+1}|s_t,a_t}$ and $f_{\varepsilon_{t+1}|s_{t+1},a_{t+1}}$, as well as CCPs $p_{t+1}(s_{t+1})$, for each t.
2. It matches the resulting LHS log odds ratios (or CCPs) with **actual** ones from the data, solving for the parameter estimates via GMM.

- Nonparametric can be difficult (d_s large, s_t discrete & continuous);
- Appropriate parametric specification? Robust?
- Computationally demanding (nonstationary model).

Brief Recount of the Hotz & Miller Estimators (cont'd)

1. To evaluate the RHS under **hypothesized** parameter values, the HM estimators **estimate**, **simulate from**, and **integrate over** state transitions distributions $f_{s_{t+1}|s_t,a_t}$ and $f_{\varepsilon_{t+1}|s_{t+1},a_{t+1}}$, as well as CCPs $p_{t+1}(s_{t+1})$, for each t.
2. It matches the resulting LHS log odds ratios (or CCPs) with **actual** ones from the data, solving for the parameter estimates via GMM.

- Nonparametric can be difficult (d_s large, s_t discrete & continuous);
- Appropriate parametric specification? Robust?
- Computationally demanding (nonstationary model).

Brief Recount of the Hotz & Miller Estimators (cont'd)

1. To evaluate the RHS under **hypothesized** parameter values, the HM estimators **estimate**, **simulate from**, and **integrate over** state transitions distributions $f_{s_{t+1}|s_t,a_t}$ and $f_{\varepsilon_{t+1}|s_{t+1},a_{t+1}}$, as well as CCPs $p_{t+1}(s_{t+1})$, for each t.
2. It matches the resulting LHS log odds ratios (or CCPs) with **actual** ones from the data, solving for the parameter estimates via GMM.

- Nonparametric can be difficult (d_s large, s_t discrete & continuous);
- Appropriate parametric specification? Robust?
- Computationally demanding (nonstationary model).

Brief Recount of the Hotz & Miller Estimators (cont'd)

1. To evaluate the RHS under **hypothesized** parameter values, the HM estimators **estimate**, **simulate from**, and **integrate over** state transitions distributions $f_{s_{t+1}|s_t,a_t}$ and $f_{\varepsilon_{t+1}|s_{t+1},a_{t+1}}$, as well as CCPs $p_{t+1}(s_{t+1})$, for each t.
2. It matches the resulting LHS log odds ratios (or CCPs) with **actual** ones from the data, solving for the parameter estimates via GMM.

- Nonparametric can be difficult (d_s large, s_t discrete & continuous);
- Appropriate parametric specification? Robust?
- Computationally demanding (nonstationary model).

Brief Recount of the Hotz & Miller Estimators (cont'd)

1. To evaluate the RHS under **hypothesized** parameter values, the HM estimators **estimate**, **simulate from**, and **integrate over** state transitions distributions $f_{s_{t+1}|s_t,a_t}$ and $f_{\varepsilon_{t+1}|s_{t+1},a_{t+1}}$, as well as CCPs $p_{t+1}(s_{t+1})$, for each t.
2. It matches the resulting LHS log odds ratios (or CCPs) with **actual** ones from the data, solving for the parameter estimates via GMM.

- Nonparametric can be difficult (d_s large, s_t discrete & continuous);
- Appropriate parametric specification? Robust?
- Computationally demanding (nonstationary model).

Outline

Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterated conditional means in $\Delta\mathbb{E}\left(\bar{V}_{t+1}(s_{t+1}) \mid s_t\right)$ under common Assumptions 1-3.
  2. Transform into a partially linear system under common Assumption 4.
  3. Transform into a linear system under mild new Assumption 5.
- Identification of the linear system uses usual linear GMM argument.
- A 3-step CCP-based semiparametric estimator follows immediately.

Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterated conditional means in $\Delta\mathbb{E}\left(\bar{V}_{t+1}(s_{t+1}) \mid s_t\right)$ under common Assumptions 1-3.
  2. Transform into a partially linear system under common Assumption 4.
  3. Transform into a linear system under mild new Assumption 5.
- Identification of the linear system uses usual linear GMM argument.
- A 3-step CCP-based semiparametric estimator follows immediately.

Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterated conditional means in $\Delta\mathbb{E}\left(\bar{V}_{t+1}(s_{t+1}) \mid s_t\right)$ under common Assumptions 1-3.
  2. Transform into a partially linear system under common Assumption 4.
  3. Transform into a linear system under mild new Assumption 5.
- Identification of the linear system uses usual linear GMM argument.
- A 3-step CCP-based semiparametric estimator follows immediately.

Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterated conditional means in $\Delta\mathbb{E}\left(\bar{V}_{t+1}(s_{t+1}) \mid s_t\right)$ under common Assumptions 1-3.
  2. Transform into a partially linear system under common Assumption 4.
  3. Transform into a linear system under mild new Assumption 5.
- Identification of the linear system uses usual linear GMM argument.
- A 3-step CCP-based semiparametric estimator follows immediately.

Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterated conditional means in $\Delta\mathbb{E}\left(\bar{V}_{t+1}(s_{t+1}) \mid s_t\right)$ under common Assumptions 1-3.
  2. Transform into a partially linear system under common Assumption 4.
  3. Transform into a linear system under mild new Assumption 5.
- Identification of the linear system uses usual linear GMM argument.
- A 3-step CCP-based semiparametric estimator follows immediately.

Step 1: Collapse of Iterated Conditional Means

Lemma 1 (Markovian s_t)

Under Assumptions 1 and 2(i)-(iii), s_t is a first order Markov process; that is, $s_{t+1} \perp\!\!\!\perp s_{t-j} \mid s_t$ for $j \in \mathbb{N}^+$.

Lemma 2 (Conditional independence)

Under Assumptions 1 and 2(i)-(iii),

$$\mathbb{E}(\mathbb{E}(g(s_{t+j}) \mid s_{t+1}) \mid s_t, a_t) = \mathbb{E}(g(s_{t+j}) \mid s_t, a_t)$$

for $j \in \mathbb{N}^+$ and measurable function $g(\cdot)$.

Step 1: Collapse of Iterated Conditional Means

Lemma 1 (Markovian s_t)

Under Assumptions 1 and 2(i)-(iii), s_t is a first order Markov process; that is, $s_{t+1} \perp\!\!\!\perp s_{t-j} \mid s_t$ for $j \in \mathbb{N}^+$.

Lemma 2 (Conditional independence)

Under Assumptions 1 and 2(i)-(iii),

$$\mathbb{E}(\mathbb{E}(g(s_{t+j}) \mid s_{t+1}) \mid s_t, a_t) = \mathbb{E}(g(s_{t+j}) \mid s_t, a_t)$$

for $j \in \mathbb{N}^+$ and measurable function $g(\cdot)$.

Step 1: Collapse of Iterated Conditional Means (cont'd)

Theorem 1 (Collapse of iterated conditional means)

Under Assumptions 1-3, the log odds ratio for $t \leq T_{tr} - 1$ simplifies to

$$\begin{aligned}
&\ln\left(\frac{p_t}{1-p_t}\right) = u_t(1, x_t) - u_t(0, x_t) \\
&+ \sum_{\tau=t+1}^{T_{tr}-1} \beta^{\tau-t} \Delta\mathbb{E}(\cancel{\mathbb{E}}(\cdots \cancel{\mathbb{E}}(\cancel{\mathbb{E}}(U^o_\tau(s_\tau) \mid \cancel{s_{\tau-1}}) \mid \cancel{s_{\tau-2}}) \cdots \mid \cancel{s_{t+1}}) \mid s_t) \\
&+ \beta^{T_{tr}-t} \Delta\mathbb{E}(\cancel{\mathbb{E}}(\cdots \cancel{\mathbb{E}}(\cancel{\mathbb{E}}(\bar{V}_{T_{tr}}(s_{T_{tr}}) \mid \cancel{s_{T_{tr}-1}}) \mid \cancel{s_{T_{tr}-2}}) \cdots \mid \cancel{s_{t+1}}) \mid s_t) \\
= &\ u_t(1, x_t) - u_t(0, x_t) \\
&+ \underbrace{\sum_{\tau=t+1}^{T_{tr}-1} \beta^{\tau-t} \Delta\mathbb{E}(U^o_\tau(s_\tau) \mid s_t) + \beta^{T_{tr}-t} \Delta\mathbb{E}(\bar{V}_{T_{tr}}(s_{T_{tr}}) \mid s_t)}_{\beta\Delta\mathbb{E}(\bar{V}_{t+1}(s_{t+1}) \mid s_t)}. \qquad (1)
\end{aligned}$$

Step 2: Transform into a Partially Linear System

Assumption 4 (Linear flow utility)

For each $t \in \mathcal{T}$, suppose $u_t(0, x_t) = x_t'\delta_{0,t}$ and $u_t(1, x_t) = x_t'\delta_{1,t}$ for some $\delta_{0,t}$ and $\delta_{1,t}$. We normalize $\delta_{0,1} = 0_{d_x \times 1}$.

Lemma 3 (Expected optimal flow utility)

Under Assumptions 1-4,

$$U_t^o(s_t) = p_t x_t'\delta_{1,t} + (1 - p_t)x_t'\delta_{0,t} - p_t \ln(p_t) - (1 - p_t)\ln(1 - p_t).$$

- By eq. (3.8) of Hotz & Miller (1993).
- Weaker than assuming $u_t(0, x_t) = 0$ for all x_t values and all $t \in \mathcal{T}$ (i.e., $\delta_{0,t} = 0$ for all $t \in \mathcal{T}$), typical in the DDC literature.
- Inter-temporal variation in x_t can help identify $\delta_{0,t}$.

Step 2: Transform into a Partially Linear System

Assumption 4 (Linear flow utility)

For each $t \in \mathcal{T}$, suppose $u_t(0, x_t) = x_t'\delta_{0,t}$ and $u_t(1, x_t) = x_t'\delta_{1,t}$ for some $\delta_{0,t}$ and $\delta_{1,t}$. We normalize $\delta_{0,1} = 0_{d_x \times 1}$.

Lemma 3 (Expected optimal flow utility)

Under Assumptions 1-4,

$$U_t^o(s_t) = p_t x_t'\delta_{1,t} + (1 - p_t)x_t'\delta_{0,t} - p_t \ln(p_t) - (1 - p_t)\ln(1 - p_t).$$

- By eq. (3.8) of Hotz & Miller (1993).
- Weaker than assuming $u_t(0, x_t) = 0$ for all x_t values and all $t \in \mathcal{T}$ (i.e., $\delta_{0,t} = 0$ for all $t \in \mathcal{T}$), typical in the DDC literature.
- Inter-temporal variation in x_t can help identify $\delta_{0,t}$.

Step 2: Transform into a Partially Linear System

Assumption 4 (Linear flow utility)

For each $t \in \mathcal{T}$, suppose $u_t(0, x_t) = x_t'\delta_{0,t}$ and $u_t(1, x_t) = x_t'\delta_{1,t}$ for some $\delta_{0,t}$ and $\delta_{1,t}$. We normalize $\delta_{0,1} = 0_{d_x \times 1}$.

Lemma 3 (Expected optimal flow utility)

Under Assumptions 1-4,

$$U_t^o(s_t) = p_t x_t'\delta_{1,t} + (1 - p_t)x_t'\delta_{0,t} - p_t \ln(p_t) - (1 - p_t)\ln(1 - p_t).$$

- By eq. (3.8) of Hotz & Miller (1993).
- Weaker than assuming $u_t(0, x_t) = 0$ for all x_t values and all $t \in \mathcal{T}$ (i.e., $\delta_{0,t} = 0$ for all $t \in \mathcal{T}$), typical in the DDC literature.
- Inter-temporal variation in x_t can help identify $\delta_{0,t}$.

Step 2: Transform into a Partially Linear System

Assumption 4 (Linear flow utility)

For each $t \in \mathcal{T}$, suppose $u_t(0, x_t) = x_t' \delta_{0,t}$ and $u_t(1, x_t) = x_t' \delta_{1,t}$ for some $\delta_{0,t}$ and $\delta_{1,t}$. We normalize $\delta_{0,1} = 0_{d_x \times 1}$.

Lemma 3 (Expected optimal flow utility)

Under Assumptions 1-4,

$$U_t^o(s_t) = p_t x_t' \delta_{1,t} + (1 - p_t) x_t' \delta_{0,t} - p_t \ln(p_t) - (1 - p_t) \ln(1 - p_t).$$

- By eq. (3.8) of Hotz & Miller (1993).
- Weaker than assuming $u_t(0, x_t) = 0$ for all x_t values and all $t \in \mathcal{T}$ (i.e., $\delta_{0,t} = 0$ for all $t \in \mathcal{T}$), typical in the DDC literature.
- Inter-temporal variation in x_t can help identify $\delta_{0,t}$.

Step 2: Transform into a Partially Linear System (cont'd)

- Theorem 1 & Lemma 3 transform the log odds ratios into a "**triangular**" partially linear system:

$$y_{T-1} = x'_{T-1}\Delta_{T-1} + \beta\Delta\mathbb{E}(\bar{V}_T(x_T, z_T)|s_{T-1}), \text{ and} \tag{2a}$$

$$y_t = x'_t\Delta_t + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}_t^{\tau\prime}\delta_{0,\tau} + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}_{1,t}^{\tau\prime}\Delta_\tau + \beta^{T-t}\Delta\mathbb{E}(\bar{V}_T(x_T, z_T)|s_t), \text{ for } t = 1, \dots, T-2, \tag{2b}$$

where
- We let $\Delta_t \equiv \delta_{1,t} - \delta_{0,t}$ and $T_{tr} = T \leq T_{end}$ (sample terminal period).
- $\Delta\bar{x}_{1,t}^\tau \equiv \Delta\mathbb{E}(p_\tau x_\tau|x_t, z_t)$ and $\Delta\bar{x}_t^\tau \equiv \Delta\mathbb{E}(x_\tau|x_t, z_t)$;
- $y_{T-1} \equiv \ln\left(\frac{p_{T-1}}{1-p_{T-1}}\right)$ and $y_t \equiv \ln\left(\frac{p_t}{1-p_t}\right) + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{\eta}_t^\tau$;
- $\Delta\bar{\eta}_t^\tau \equiv \Delta\mathbb{E}(\eta_\tau|x_t, z_t)$, $\eta_\tau \equiv p_\tau\ln(p_\tau) + (1-p_\tau)\ln(1-p_\tau)$ for $\tau > t$.

- Note: $\Delta\bar{x}_{1,t}^\tau$, $\Delta\bar{x}_t^\tau$ and y_t are either observed or estimable.

Step 2: Transform into a Partially Linear System (cont'd)

- Theorem 1 & Lemma 3 transform the log odds ratios into a "**triangular**" partially linear system:

$$y_{T-1} = x'_{T-1}\Delta_{T-1} + \beta\Delta\mathbb{E}(\bar{V}_T(x_T, z_T)|s_{T-1}), \text{ and} \tag{2a}$$

$$y_t = x'_t\Delta_t + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}_t^{\tau\prime}\delta_{0,\tau} + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}_{1,t}^{\tau\prime}\Delta_\tau + \beta^{T-t}\Delta\mathbb{E}(\bar{V}_T(x_T, z_T)|s_t), \text{ for } t = 1, \ldots, T-2, \tag{2b}$$

  where
  - We let $\Delta_t \equiv \delta_{1,t} - \delta_{0,t}$ and $T_{tr} = T \leq T_{end}$ (sample terminal period).
  - $\Delta\bar{x}_{1,t}^\tau \equiv \Delta\mathbb{E}(p_\tau x_\tau|x_t, z_t)$ and $\Delta\bar{x}_t^\tau \equiv \Delta\mathbb{E}(x_\tau|x_t, z_t)$;
  - $y_{T-1} \equiv \ln\left(\frac{p_{T-1}}{1-p_{T-1}}\right)$ and $y_t \equiv \ln\left(\frac{p_t}{1-p_t}\right) + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{\eta}_t^\tau$;
  - $\Delta\bar{\eta}_t^\tau \equiv \Delta\mathbb{E}(\eta_\tau|x_t, z_t)$, $\eta_\tau \equiv p_\tau\ln(p_\tau) + (1-p_\tau)\ln(1-p_\tau)$ for $\tau > t$.
- Note: $\Delta\bar{x}_{1,t}^\tau$, $\Delta\bar{x}_t^\tau$ and y_t are either observed or estimable.

Step 3: Transform into a Linear System

Assumption 5 (Sample-terminal-period integrated value function)

Suppose there exists a $K \times 1$ vector of known functions of (x, z), denoted by $q^K(x, z) \equiv (q^{K,1}(x, z), \ldots, q^{K,K}(x, z))'$, and a $K \times 1$ unknown vector of parameters γ^K, such that $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)'\gamma^K$.

- This further transforms the "**triangular**" system into a linear one:

$$y_{T-1} = x_{T-1}'\Delta_{T-1} + \beta\Delta\bar{q}_{T-1}^K\gamma^K, \text{ and} \tag{3a}$$

$$y_t = x_t'\Delta_t + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}_t^{\tau\prime}\delta_{0,\tau} + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}_{1,t}^{\tau\prime}\Delta_\tau + \beta^{T-t}\Delta\bar{q}_t^K\gamma^K, \text{ for } t = 1, \ldots, T-2, \tag{3b}$$

where $\Delta\bar{q}_t^K \equiv \Delta\mathbb{E}(q^K(x_T, z_T)|x_t, z_t)$.

Step 3: Transform into a Linear System

Assumption 5 (Sample-terminal-period integrated value function)

Suppose there exists a $K \times 1$ *vector of known functions of* (x, z)*, denoted by* $q^K(x,z) \equiv (q^{K,1}(x,z), \dots, q^{K,K}(x,z))'$*, and a* $K \times 1$ *unknown vector of parameters* γ^K*, such that* $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)'\gamma^K$*.*

- This further transforms the "**triangular**" system into a linear one:

$$y_{T-1} = x'_{T-1}\Delta_{T-1} + \beta\Delta\bar{q}^K_{T-1}\gamma^K, \text{ and} \tag{3a}$$

$$y_t = x'_t\Delta_t + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}^{\tau\prime}_t\delta_{0,\tau} + \sum_{\tau=t+1}^{T-1}\beta^{\tau-t}\Delta\bar{x}^{\tau\prime}_{1,t}\Delta_\tau + \beta^{T-t}\Delta\bar{q}^K_t\gamma^K, \text{ for } t = 1, \dots, T-2, \tag{3b}$$

where $\Delta\bar{q}^K_t \equiv \Delta\mathbb{E}(q^K(x_T, z_T)|x_t, z_t)$.

(Over-)Identification of the Linear System

Theorem 2 (Identification of δ and γ^K)

Under Assumptions 1-5, if: (i) $T \geq 2$; (ii) $\tilde{L}_{T-1}$ is invertible; and (iii) second moment matrices of $(x_t', \Delta\bar{x}_t^{t+1\prime})'$ are all invertible ($t = 1, \ldots, T-2$), then $\delta \equiv (\Delta_1', \delta_{0,2}', \Delta_2', \ldots, \delta_{0,T-1}', \Delta_{T-1}')'$ and γ^K are identified.

$$
\begin{pmatrix}
0_{d_x\times(2(T-4)d_x)} & 0_{d_x\times d_x} & 0_{d_x\times d_x} & 0_{d_x\times d_x} & 0_{d_x\times d_x} & \mathbb{E}(x_{T-1}x_{T-1}') & \beta\mathbb{E}(x_{T-1}\Delta\bar{q}_{T-1}^{K\prime}) \\
0_{K\times(2(T-4)d_x)} & 0_{K\times d_x} & 0_{K\times d_x} & 0_{K\times d_x} & 0_{K\times d_x} & \beta\mathbb{E}(\Delta\bar{q}_{T-1}^{K}x_{T-1}') & \beta^2\mathbb{E}(\Delta\bar{q}_{T-1}^{K}\Delta\bar{q}_{T-1}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & 0_{d_x\times d_x} & 0_{d_x\times d_x} & \mathbb{E}(x_{T-2}x_{T-2}') & \beta\mathbb{E}(x_{T-2}\Delta\bar{x}_{T-2}^{T-1\prime}) & \beta\mathbb{E}(x_{T-2}\Delta\bar{x}_{1,T-2}^{T-1\prime}) & \beta^2\mathbb{E}(x_{T-2}\Delta\bar{q}_{T-2}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & 0_{d_x\times d_x} & 0_{d_x\times d_x} & \beta\mathbb{E}(\Delta\bar{x}_{T-2}^{T-1}x_{T-2}') & \beta^2\mathbb{E}(\Delta\bar{x}_{T-2}^{T-1}\Delta\bar{x}_{T-2}^{T-1\prime}) & \beta^2\mathbb{E}(\Delta\bar{x}_{T-2}^{T-1}\Delta\bar{x}_{1,T-2}^{T-1\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{T-2}^{T-1}\Delta\bar{q}_{T-2}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & 0_{d_x\times d_x} & 0_{d_x\times d_x} & \beta\mathbb{E}(\Delta\bar{x}_{1,T-2}^{T-1}x_{T-2}') & \beta^2\mathbb{E}(\Delta\bar{x}_{1,T-2}^{T-1}\Delta\bar{x}_{T-2}^{T-1\prime}) & \beta^2\mathbb{E}(\Delta\bar{x}_{1,T-2}^{T-1}\Delta\bar{x}_{1,T-2}^{T-1\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{1,T-2}^{T-1}\Delta\bar{q}_{T-2}^{K\prime}) \\
0_{K\times(2(T-4)d_x)} & 0_{K\times d_x} & 0_{K\times d_x} & \beta^2\mathbb{E}(\Delta\bar{q}_{T-2}^{K}x_{T-2}') & \beta^3\mathbb{E}(\Delta\bar{q}_{T-2}^{K}\Delta\bar{x}_{T-2}^{T-1\prime}) & \beta^3\mathbb{E}(\Delta\bar{q}_{T-2}^{K}\Delta\bar{x}_{1,T-2}^{T-1\prime}) & \beta^4\mathbb{E}(\Delta\bar{q}_{T-2}^{K}\Delta\bar{q}_{T-2}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & \mathbb{E}(x_{T-3}x_{T-3}') & \beta\mathbb{E}(x_{T-3}\Delta\bar{x}_{T-3}^{T-2\prime}) & \beta\mathbb{E}(x_{T-3}\Delta\bar{x}_{1,T-3}^{T-2\prime}) & \beta^2\mathbb{E}(x_{T-3}\Delta\bar{x}_{T-3}^{T-1\prime}) & \beta^2\mathbb{E}(x_{T-3}\Delta\bar{x}_{1,T-3}^{T-1\prime}) & \beta^3\mathbb{E}(x_{T-3}\Delta\bar{q}_{T-3}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & \beta\mathbb{E}(\Delta\bar{x}_{T-3}^{T-2}x_{T-3}') & \beta^2\mathbb{E}(\Delta\bar{x}_{T-3}^{T-2}\Delta\bar{x}_{T-3}^{T-2\prime}) & \beta^2\mathbb{E}(\Delta\bar{x}_{T-3}^{T-2}\Delta\bar{x}_{1,T-3}^{T-2\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{T-3}^{T-2}\Delta\bar{x}_{T-3}^{T-1\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{T-3}^{T-2}\Delta\bar{x}_{1,T-3}^{T-1\prime}) & \beta^4\mathbb{E}(\Delta\bar{x}_{T-3}^{T-2}\Delta\bar{q}_{T-3}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & \beta\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-2}x_{T-3}') & \beta^2\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-2}\Delta\bar{x}_{T-3}^{T-2\prime}) & \beta^2\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-2}\Delta\bar{x}_{1,T-3}^{T-2\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-2}\Delta\bar{x}_{T-3}^{T-1\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-2}\Delta\bar{x}_{1,T-3}^{T-1\prime}) & \beta^4\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-2}\Delta\bar{q}_{T-3}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & \beta^2\mathbb{E}(\Delta\bar{x}_{T-3}^{T-1}x_{T-3}') & \beta^3\mathbb{E}(\Delta\bar{x}_{T-3}^{T-1}\Delta\bar{x}_{T-3}^{T-2\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{T-3}^{T-1}\Delta\bar{x}_{1,T-3}^{T-2\prime}) & \beta^4\mathbb{E}(\Delta\bar{x}_{T-3}^{T-1}\Delta\bar{x}_{T-3}^{T-1\prime}) & \beta^4\mathbb{E}(\Delta\bar{x}_{T-3}^{T-1}\Delta\bar{x}_{1,T-3}^{T-1\prime}) & \beta^5\mathbb{E}(\Delta\bar{x}_{T-3}^{T-1}\Delta\bar{q}_{T-3}^{K\prime}) \\
0_{d_x\times(2(T-4)d_x)} & \beta^2\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-1}x_{T-3}') & \beta^3\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-1}\Delta\bar{x}_{T-3}^{T-2\prime}) & \beta^3\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-1}\Delta\bar{x}_{1,T-3}^{T-2\prime}) & \beta^4\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-1}\Delta\bar{x}_{T-3}^{T-1\prime}) & \beta^4\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-1}\Delta\bar{x}_{1,T-3}^{T-1\prime}) & \beta^5\mathbb{E}(\Delta\bar{x}_{1,T-3}^{T-1}\Delta\bar{q}_{T-3}^{K\prime}) \\
0_{K\times(2(T-4)d_x)} & \beta^3\mathbb{E}(\Delta\bar{q}_{T-3}^{K}x_{T-3}') & \beta^4\mathbb{E}(\Delta\bar{q}_{T-3}^{K}\Delta\bar{x}_{T-3}^{T-2\prime}) & \beta^4\mathbb{E}(\Delta\bar{q}_{T-3}^{K}\Delta\bar{x}_{1,T-3}^{T-2\prime}) & \beta^5\mathbb{E}(\Delta\bar{q}_{T-3}^{K}\Delta\bar{x}_{T-3}^{T-1\prime}) & \beta^5\mathbb{E}(\Delta\bar{q}_{T-3}^{K}\Delta\bar{x}_{1,T-3}^{T-1\prime}) & \beta^6\mathbb{E}(\Delta\bar{q}_{T-3}^{K}\Delta\bar{q}_{T-3}^{K\prime}) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots
\end{pmatrix}
$$

$\tilde{L}_{T-1}$

$\tilde{L}_{T-2}$

$\tilde{L}_{T-3}$

(Over-)Identification of the Linear System (cont'd)

In-depth discussion in the paper:

- Excluded variable(s) z_t breaks linear correlation between x_t and $\Delta\bar{x}_t^{t+1} \equiv \Delta\mathbb{E}(x_{t+1}|x_t, z_t)$. details
- Triangularity helps deal with time-invariant variables in x_t. details
- Unknown discount factor easily accommodated ($\tilde{L}_{T-2}$ invertible).
- Assumption 5 holds with $q^K(x_T, z_T) = (x_T', p_T x_T' \Delta_T - \eta_T)'$ when $T = T_{end}$.
- When interpreting Assumption 5 as an approximation, bias can be quantified. details
- Over-identification reduces reliance on Assumption 5. details

(Over-)Identification of the Linear System (cont'd)

In-depth discussion in the paper:

- Excluded variable(s) z_t breaks linear correlation between x_t and $\Delta \bar{x}_t^{t+1} \equiv \Delta \mathbb{E}(x_{t+1}|x_t, z_t)$. details
- Triangularity helps deal with time-invariant variables in x_t. details
- Unknown discount factor easily accommodated ($\tilde{L}_{T-2}$ invertible).
- Assumption 5 holds with $q^K(x_T, z_T) = (x_T', p_T x_T' \Delta_T - \eta_T)'$ when $T = T_{end}$.
- When interpreting Assumption 5 as an approximation, bias can be quantified. details
- Over-identification reduces reliance on Assumption 5. details

(Over-)Identification of the Linear System (cont'd)

In-depth discussion in the paper:

- Excluded variable(s) z_t breaks linear correlation between x_t and $\Delta\bar{x}_t^{t+1} \equiv \Delta\mathbb{E}(x_{t+1}|x_t, z_t)$. details
- Triangularity helps deal with time-invariant variables in x_t. details
- Unknown discount factor easily accommodated ($\tilde{L}_{T-2}$ invertible).
- Assumption 5 holds with $q^K(x_T, z_T) = (x_T', p_T x_T' \Delta_T - \eta_T)'$ when $T = T_{end}$.
- When interpreting Assumption 5 as an approximation, bias can be quantified. details
- Over-identification reduces reliance on Assumption 5. details

(Over-)Identification of the Linear System (cont'd)

In-depth discussion in the paper:

- Excluded variable(s) z_t breaks linear correlation between x_t and $\Delta\bar{x}_t^{t+1} \equiv \Delta\mathbb{E}(x_{t+1}|x_t, z_t)$. details
- Triangularity helps deal with time-invariant variables in x_t. details
- Unknown discount factor easily accommodated ($\tilde{L}_{T-2}$ invertible).
- Assumption 5 holds with $q^K(x_T, z_T) = (x_T', p_T x_T' \Delta_T - \eta_T)'$ when $T = T_{end}$.
- When interpreting Assumption 5 as an approximation, bias can be quantified. details
- Over-identification reduces reliance on Assumption 5. details

(Over-)Identification of the Linear System (cont'd)

In-depth discussion in the paper:

- Excluded variable(s) z_t breaks linear correlation between x_t and $\Delta\bar{x}_t^{t+1} \equiv \Delta\mathbb{E}(x_{t+1}|x_t, z_t)$. details
- Triangularity helps deal with time-invariant variables in x_t. details
- Unknown discount factor easily accommodated ($\tilde{L}_{T-2}$ invertible).
- Assumption 5 holds with $q^K(x_T, z_T) = (x_T', p_T x_T' \Delta_T - \eta_T)'$ when $T = T_{end}$.
- When interpreting Assumption 5 as an approximation, bias can be quantified. details
- Over-identification reduces reliance on Assumption 5. details

(Over-)Identification of the Linear System (cont'd)

In-depth discussion in the paper:

- Excluded variable(s) z_t breaks linear correlation between x_t and $\Delta\bar{x}_t^{t+1} \equiv \Delta\mathbb{E}(x_{t+1}|x_t, z_t)$. details
- Triangularity helps deal with time-invariant variables in x_t. details
- Unknown discount factor easily accommodated ($\tilde{L}_{T-2}$ invertible).
- Assumption 5 holds with $q^K(x_T, z_T) = (x'_T, p_T x'_T \Delta_T - \eta_T)'$ when $T = T_{end}$.
- When interpreting Assumption 5 as an approximation, bias can be quantified. details
- Over-identification reduces reliance on Assumption 5. details

3-Step CCP-Based Semiparametric Estimator

1. For each $t \in \{1, \ldots, T\}$, estimate the CCP function $p_t(\cdot)$, and obtain the estimated $\hat{p}_{it}$ for every $i \in \{1, \ldots, N\}$.
2. Plug $\hat{p}_{i\tau}$ in relevant $h_{i\tau}$ and obtain the estimated $\widehat{\Delta\bar{\eta}}$, $\widehat{\Delta\bar{x}}$ and $\widehat{\Delta\bar{q}}$.
3. Let $\bar{m}_N(\delta, \gamma^K) \equiv \frac{1}{N}\sum_{i=1}^{N} m(x_i, z_i, a_i, \delta, \gamma^K, \hat{p}, \widehat{\Delta\bar{\eta}}, \widehat{\Delta\bar{x}}, \widehat{\Delta\bar{q}^K})$ capture the distance between LHS & RHS of the linear system. Then, the minimum-distance estimator $(\hat{\delta}', \hat{\gamma}_K')'$ solves:

$$(\hat{\delta}', \hat{\gamma}^{K\prime})' \equiv \underset{\delta \in \mathbb{R}^{(2T-3)d_x}, \gamma_K \in \mathbb{R}^K}{\arg\min} \bar{m}_N(\delta, \gamma^K)' W_N \bar{m}_N(\delta, \gamma^K).$$

3-Step CCP-Based Semiparametric Estimator

1. For each $t \in \{1, \dots, T\}$, estimate the CCP function $p_t(\cdot)$, and obtain the estimated $\hat{p}_{it}$ for every $i \in \{1, \dots, N\}$.
2. Plug $\hat{p}_{i\tau}$ in relevant $h_{i\tau}$ and obtain the estimated $\widehat{\Delta\bar{\eta}}$, $\widehat{\Delta\bar{x}}$ and $\widehat{\Delta\bar{q}}$.
3. Let $\bar{m}_N(\delta, \gamma^K) \equiv \frac{1}{N}\sum_{i=1}^{N} m(x_i, z_i, a_i, \delta, \gamma^K, \hat{p}, \widehat{\Delta\bar{\eta}}, \widehat{\Delta\bar{x}}, \widehat{\Delta\bar{q}^K})$ capture the distance between LHS & RHS of the linear system. Then, the minimum-distance estimator $(\hat{\delta}', \hat{\gamma}_K')'$ solves:

$$(\hat{\delta}', \hat{\gamma}^{K\prime})' \equiv \underset{\delta \in \mathbb{R}^{(2T-3)d_x}, \gamma_K \in \mathbb{R}^K}{\arg\min} \bar{m}_N(\delta, \gamma^K)' W_N \bar{m}_N(\delta, \gamma^K).$$

3-Step CCP-Based Semiparametric Estimator

1. For each $t \in \{1, \dots, T\}$, estimate the CCP function $p_t(\cdot)$, and obtain the estimated $\hat{p}_{it}$ for every $i \in \{1, \dots, N\}$.
2. Plug $\hat{p}_{i\tau}$ in relevant $h_{i\tau}$ and obtain the estimated $\widehat{\Delta\bar{\eta}}$, $\widehat{\Delta\bar{x}}$ and $\widehat{\Delta\bar{q}}$.
3. Let $\bar{m}_N(\delta, \gamma^K) \equiv \frac{1}{N}\sum_{i=1}^{N} m(x_i, z_i, a_i, \delta, \gamma^K, \hat{p}, \widehat{\Delta\bar{\eta}}, \widehat{\Delta\bar{x}}, \widehat{\Delta\bar{q}^K})$ capture the distance between LHS & RHS of the linear system. Then, the minimum-distance estimator $(\hat{\delta}', \hat{\gamma}_K')'$ solves:

$$(\hat{\delta}', \hat{\gamma}^{K\prime})' \equiv \underset{\delta \in \mathbb{R}^{(2T-3)d_x}, \gamma_K \in \mathbb{R}^K}{\arg\min} \bar{m}_N(\delta, \gamma^K)' W_N \bar{m}_N(\delta, \gamma^K).$$

3-Step CCP-Based Semiparametric Estimation (cont'd)

Proposition 1 (Asymptotic distribution of $\hat{\delta}$)

$$\sqrt{N}\left(\hat{\delta} - \delta\right) \xrightarrow{d.} \mathcal{N}(0, V),$$

where $V \equiv \mathbb{E}[\psi_\delta(x_i, z_i, a_i)\psi'_\delta(x_i, z_i, a_i)]$ *and* $\psi_\delta(x, z, a)$ *is the influence function of* $\hat{\delta}$*, given in the paper.*

Proposition 2 (Consistent estimator of V)

$\widehat{V} \xrightarrow{p.} V$ *for the* $\widehat{V}$ *given in the paper.*

- Consistency follows usual M-estimator argument.
- The impact of Estimation Steps 1 & 2 is accounted for via Newey (1994) method (infl. fn. contains only the "adjustment terms").

Outline

Concluding Remarks

- State transition is essential for counterfactual analysis.
- Future inquiry:
  - Multiple discrete choices.
  - Unobserved heterogeneity as characterized by latent types (Arcidiacono & Miller, 2011).
  - Serially correlated flow utility shocks?
  - Identification of $u_t(a_t, x_t)$ while relaxing Assumption 4 (linear-in-x_t)?

Function of the Excluded Variable(s) z_t

- Condition (iii) of Theorem 2 requires that variables in x_t and $\Delta\bar{x}_t^{t+1} \equiv \Delta\mathbb{E}(x_{t+1}|x_t, z_t)$ are not perfectly linearly correlated.
- **Takeaway**: the variation in z_t and the way it affects $\Delta\mathbb{E}(x_{t+1}|x_t, z_t)$ is crucial for (iii) of Theorem 2, not d_z. back
- Least favorable case: $\Delta\mathbb{E}(x_{t+1}|x_t, z_t)$ contains an additive component that is linear in x_t (denoted as $\rho_1 x_t$).
- If there exist $\ell(\cdot) \equiv (\ell_1(\cdot), \ldots, \ell_{d_x}(\cdot))$ and ρ_2 such that

$$\underbrace{\begin{bmatrix} x_t \\ \Delta\mathbb{E}(x_{t+1}|x_t, z_t) \end{bmatrix}}_{2d_x \times 1} = \underbrace{\begin{bmatrix} I_{d_x} & 0_{d_x \times d_z} \\ \rho_1 & \rho_2 \end{bmatrix}}_{2d_x \times 2d_x} \underbrace{\begin{bmatrix} x_t \\ \ell(z_t) \end{bmatrix}}_{2d_x \times 1}.$$

Then sufficient conditions for (iii) of Theorem 2 are: (a) $(x_t', \ell(z_t)')'$ has invertible second moment matrix; and (b) ρ_2 has full rank (d_x).

Identification with Time-invariant Variables in x_t

- **Takeaway**: corresponding coordinate in $\delta_{0,t}$ is unidentified, but Δ_t is, and the counterfactual* is unaffected.
- Suppose Δ_{T-1} and γ^K are identified.
- Suppose $x_{t,1}$ is time-invariant, so $\Delta \bar{x}^{\tau}_{1,t} = 0$ for $\tau > t$.
- Consider eq. (3b) for $t = T-2$ and $t = T-3$:

$$y_{T-2} - \beta \Delta \bar{x}^{T-1\prime}_{1,T-2} \Delta_{T-1} - \beta^2 \Delta \bar{q}^{K\prime}_{T-2} \gamma^K$$

$$= x'_{T-2} \Delta_{T-2} + \beta \underbrace{\Delta \bar{x}^{T-1\prime}_{T-2}}_{=0} \delta_{0,T-1},$$

$$y_{T-3} - \beta \Delta \bar{x}^{T-2\prime}_{1,T-3} \Delta_{T-2} - \beta^2 \Delta \bar{x}^{T-1\prime}_{1,T-3} \Delta_{T-1} - \beta^2 \underbrace{\Delta \bar{x}^{T-1\prime}_{T-3}}_{=0} \delta_{0,T-1} - \beta^3$$

$$= x'_{T-3} \Delta_{T-3} + \beta \underbrace{\Delta \bar{x}^{T-2\prime}_{T-3}}_{=0} \delta_{0,T-2}.$$

- Continually & intermittently time-varying x_t both work. back

Bias Induced by Assumption 5

- Assumption 5 can be interpreted as using a series basis functions $q^K(x_T, z_T)$ to **approximate** the expected value function $\bar{V}_T(x_T, z_T)$.
- Define the approximation error

$$r^K(x_T, z_T) \equiv \bar{V}_T(x_T, z_T) - q^K(x_T, z_T)'\gamma^K,$$

  and let $\Delta \bar{r}_t^K \equiv \Delta \mathbb{E}(r^K(x_T, z_T)|x_t, z_t)$ for $t = 1, \ldots, T-1$.
- Assume $\bar{V}_T(x_T, z_T)$ is m times continuously differentiable, then the approximation error using **power series** has the order $\mathbb{E}[(\Delta \bar{r}_t^K)^2] = O\left(K^{-\frac{2m}{d_s}}\right)$. This leads to:

Theorem 3 (Asymptotic bias bound of $\hat{\delta}$, power series)

$$\|\delta^K_{pseudo} - \delta\| = O\left(K^{-\frac{m}{d_s}}\right).$$ back

Bias Induced by Assumption 5

- Assumption 5 can be interpreted as using a series basis functions $q^K(x_T, z_T)$ to **approximate** the expected value function $\bar{V}_T(x_T, z_T)$.
- Define the approximation error

$$r^K(x_T, z_T) \equiv \bar{V}_T(x_T, z_T) - q^K(x_T, z_T)'\gamma^K,$$

  and let $\Delta \bar{r}_t^K \equiv \Delta \mathbb{E}(r^K(x_T, z_T)|x_t, z_t)$ for $t = 1, \dots, T-1$.
- Assume $\bar{V}_T(x_T, z_T)$ is m times continuously differentiable, then the approximation error using **power series** has the order $\mathbb{E}[(\Delta \bar{r}_t^K)^2] = O\left(K^{-\frac{2m}{d_s}}\right)$. This leads to:

Theorem 3 (Asymptotic bias bound of $\hat{\delta}$, power series)

$$\|\delta^K_{pseudo} - \delta\| = O\left(K^{-\frac{m}{d_s}}\right).$$

back

Over-identification May Reduce Reliance on Assumption 5

- If T is large, recall eq. (2b) for $t = 1$:

$$y_1 = x_1' \Delta_1 + \sum_{\tau=2}^{T-1} \beta^{\tau-1} \Delta \bar{x}_1^{\tau\prime} \delta_{0,\tau} + \sum_{\tau=2}^{T-1} \beta^{\tau-1} \Delta \bar{x}_{1,1}^{\tau\prime} \Delta_\tau$$
$$+ \beta^{T-1} \Delta \mathbb{E}(\bar{V}_T(x_T, z_T)|s_1),$$

  which contains **all** parameters of interest.
- Also true for eq. (2b) for $t = 2$, which contains **most** parameters of interest.
- So, using eq. (2b) for the first few periods reduces reliance on Assumption 5.
- Similar to the trunction employed by the HM estimators. back