Identification and Estimation of Nonstationary Dynamic Binary Choice Models

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

Econometric Society North America Summer Meeting (UCLA)
June 23, 2023
Estimation of Dynamic Discrete Choice (DDC) Models

- **Widely useful:**
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- **Estimation:**
  - Solve the full dynamic programming problem $\rightarrow$ MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP's) $\rightarrow$ GMM (Hotz & Miller, 1993).

- **Disadvantages:**
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult $\rightarrow$ parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- Widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- Estimation:
  - Solve the full dynamic programming problem → MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP's) → GMM (Hotz & Miller, 1993).

- Disadvantages:
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult → parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- **Widely useful:**
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- **Estimation:**
  - Solve the full dynamic programming problem → MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP's) → GMM (Hotz & Miller, 1993).

- **Disadvantages:**
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult → parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- Widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- Estimation:
  - Solve the full dynamic programming problem → MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP's) → GMM (Hotz & Miller, 1993).

- Disadvantages:
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult → parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- **Widely useful:**
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- **Estimation:**
  - Solve the full dynamic programming problem $\rightarrow$ MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP’s) $\rightarrow$ GMM (Hotz & Miller, 1993).

- **Disadvantages:**
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult $\rightarrow$ parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- Widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- Estimation:
  - Solve the full dynamic programming problem → MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP’s) → GMM (Hotz & Miller, 1993).

- Disadvantages:
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult → parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- **Widely useful:**
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- **Estimation:**
  - Solve the full dynamic programming problem → MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP’s) → GMM (Hotz & Miller, 1993).

- **Disadvantages:**
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult → parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- Widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- Estimation:
  - Solve the full dynamic programming problem → MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP’s) → GMM (Hotz & Miller, 1993).

- Disadvantages:
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult → parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- Widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- Estimation:
  - Solve the full dynamic programming problem $\rightarrow$ MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP’s) $\rightarrow$ GMM (Hotz & Miller, 1993).

- Disadvantages:
  - **Implementation**: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - **Robustness**: NP state transitions are difficult $\rightarrow$ parametric specification is often assumed.
Estimation of Dynamic Discrete Choice (DDC) Models

- Widely useful:
  - Analyze inter-temporal preference & strategic interactions.
  - Counterfactual analysis.
  - Examples: labor force participation, demand for durable goods, new product offering, etc.

- Estimation:
  - Solve the full dynamic programming problem $\rightarrow$ MLE (Rust, 1987).
  - Match conditional choice probabilities (CCP’s) $\rightarrow$ GMM (Hotz & Miller, 1993).

- Disadvantages:
  - Implementation: both require estimating & simulating from state transition distributions, except for special cases (e.g., terminal choice, renewal choice).
  - Robustness: NP state transitions are difficult $\rightarrow$ parametric specification is often assumed.
We transform the key CCP equation that the Hotz & Miller method is based on into a linear system under mild assumptions. The simplified CCP equation allows us to:

- find clear sufficient identification conditions for structural parameters in flow utility;
- develop a simpler CCP-based semiparametric estimation method that avoids estimating & simulating from state transitions; and
- quantify the bias in estimates if the key new assumption is violated.
We transform the key CCP equation that the Hotz & Miller method is based on into a linear system under mild assumptions.

The simplified CCP equation allows us to:

- find clear sufficient identification conditions for structural parameters in flow utility;
- develop a simpler CCP-based semiparametric estimation method that avoids estimating & simulating from state transitions; and
- quantify the bias in estimates if the key new assumption is violated.
We transform the key CCP equation that the Hotz & Miller method is based on into a linear system under mild assumptions.

The simplified CCP equation allows us to:

- find clear sufficient **identification** conditions for structural parameters in flow utility;
- develop a simpler CCP-based semiparametric **estimation** method that avoids estimating & simulating from state transitions; and
- **quantify** the bias in estimates if the key new assumption is violated.
This Paper: A General Nonstationary DDC Model

- We transform the key CCP equation that the Hotz & Miller method is based on into a linear system under mild assumptions.
- The simplified CCP equation allows us to:
  - find clear sufficient **identification** conditions for structural parameters in flow utility;
  - develop a simpler CCP-based semiparametric **estimation** method that avoids estimating & simulating from state transitions; and
  - **quantify** the bias in estimates if the key new assumption is violated.
We transform the key CCP equation that the Hotz & Miller method is based on into a linear system under mild assumptions.

The simplified CCP equation allows us to:

- find clear sufficient **identification** conditions for structural parameters in flow utility;
- develop a simpler CCP-based semiparametric **estimation** method that avoids estimating & simulating from state transitions; and
- quantify the bias in estimates if the key new assumption is violated.
Outline

1. A General Nonstationary Dynamic Binary Choice Model
   - Model Setup
   - Brief Recount of the Hotz & Miller Method

2. Our Approach
   - Transformation into a Linear System
   - Identification
   - Estimation

3. Bias When Interpreting Assumption 5 as Approximation

4. Conclusion
Outline

1. A General Nonstationary Dynamic Binary Choice Model
   - Model Setup
   - Brief Recount of the Hotz & Miller Method

2. Our Approach
   - Transformation into a Linear System
   - Identification
   - Estimation

3. Bias When Interpreting Assumption 5 as Approximation

4. Conclusion
Each agent makes a binary choice $a_t \in \{0, 1\}$ in each period $t$.

- Decision horizon: $t \in T \equiv \{T_{\text{start}}, \ldots, T_{\text{end}}\}$, $T_{\text{end}} = \infty$ allowed.
- $a_t$ is made by comparing expected lifetime payoff:

$$u_t(a_t, x_t) + \varepsilon_{at} + \beta \mathbb{E}(\bar{V}_{t+1}(s_{t+1}|s_t, a_t)$$

- Current flow utility
- Expected discounted future lifetime payoff

let $\Omega_t = (s_t', \varepsilon_t')'$; $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{0t})'$: 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 affect future payoff but not current flow utility.

Nonstationarity: $T_{\text{end}}$ finite; $u_t(a, x)$ and $f(\Omega_{t+1}|\Omega_t)$ vary with $t$. 

A General Nonstationary Dynamic Binary Choice Model

- Each agent makes a binary choice $a_t \in \{0, 1\}$ in each period $t$.
- Decision horizon: $t \in T \equiv \{T_{\text{start}}, \ldots, T_{\text{end}}\}$, $T_{\text{end}} = \infty$ allowed.
- $a_t$ is made by comparing expected lifetime payoff:

$$u_t(a_t, x_t) + \varepsilon_{at} + \beta \mathbb{E}(\bar{V}_{t+1}(s_{t+1})|s_t, a_t)$$

- Current flow utility
- Expected discounted future lifetime payoff

- Let $\Omega_t = (s'_t, \varepsilon'_t)$; $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{0t})'$: unobserved flow utility shocks.
- $s_t = (x'_t, z'_t)'$: $(d_x + d_z) \times 1$ observed state variables.
- $x_t$ affects the current flow utility.
- $z_t$ are excluded variables, which affect future payoff but not current flow utility.

- Nonstationarity: $T_{\text{end}}$ finite; $u_t(a, x)$ and $f(\Omega_{t+1}|\Omega_t)$ vary with $t$. 
A General Nonstationary Dynamic Binary Choice Model

Each agent makes a binary choice $a_t \in \{0, 1\}$ in each period $t$.

Decision horizon: $t \in \mathcal{T} \equiv \{T_{\text{start}}, \ldots, T_{\text{end}}\}$, $T_{\text{end}} = \infty$ allowed.

$a_t$ is made by comparing expected lifetime payoff:

$$u_t(a_t, x_t) + \varepsilon_{a_t t} + \beta \mathbb{E}(\tilde{V}_{t+1}(s_{t+1}) | s_t, a_t)$$

- current flow utility
- expected discounted future lifetime payoff

- let $\Omega_t \equiv (s'_t, \varepsilon'_t)'$; $\varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{0t})'$: 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 affect future payoff but not current flow utility.

Nonstationarity: $T_{\text{end}}$ finite; $u_t(a, x)$ and $f(\Omega_{t+1} | \Omega_t)$ vary with $t$. 
A General Nonstationary Dynamic Binary Choice Model

- Each agent makes a binary choice \( a_t \in \{0, 1\} \) in each period \( t \).
- Decision horizon: \( t \in \mathcal{T} \equiv \{ T_{\text{start}}, \ldots, T_{\text{end}} \} \), \( T_{\text{end}} = \infty \) allowed.
- \( a_t \) is made by comparing expected lifetime payoff:

\[
\begin{align*}
\underbrace{u_t(a_t, x_t)}_{\text{current flow utility}} + \varepsilon_{a_t t} + \underbrace{\beta \mathbb{E}(\bar{V}_{t+1}(s_{t+1}) | s_t, a_t)}_{\text{expected discounted future lifetime payoff}}
\end{align*}
\]

- let \( \Omega_t \equiv (s'_t, \varepsilon'_t)' \); \( \varepsilon_t \equiv (\varepsilon_{1t}, \varepsilon_{0t})' \): 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 affect future payoff but not current flow utility.

- **Nonstationarity**: \( T_{\text{end}} \) finite; \( u_t(a, x) \) and \( f(\Omega_{t+1} | \Omega_t) \) vary with \( t \).
We Maintain These Common Assumptions

<table>
<thead>
<tr>
<th>Assumption 1 (Controlled Markov process)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Omega_{t+1} \perp \perp (\Omega_{t-j}, a_{t-j}) \mid (\Omega_t, a_t) \text{ for } j \in \mathbb{N}^+ ).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assumption 2 (Utility shocks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (i) \varepsilon_t \perp \perp s_t; (ii) \varepsilon_t \perp \perp s_{t-1}; (iii) \varepsilon_t \text{ is serially independent; (iv) } \varepsilon_{0t} \perp \perp \varepsilon_{1t}, \text{ and they both follow zero-mean standard type I extreme value distribution.}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assumption 3 (Conditional independence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_{t+1} \perp \perp \varepsilon_t \mid (s_t, a_t) ).</td>
</tr>
</tbody>
</table>
Under Assumptions 1-3, the log odds ratio becomes:

$$\ln \left( \frac{p_t(s_t)}{1 - p_t(s_t)} \right) = u_t(1, x_t) - u_t(0, x_t) + \beta \Delta \mathbb{E} \left( \tilde{V}_{t+1}(s_{t+1}) \mid s_t \right)$$

where

- $p_t(s_t) \equiv Pr.(a_t = 1 \mid s_t)$ is the CCP;
- $\Delta \mathbb{E}(\cdot \mid s_t) \equiv \mathbb{E}(\cdot \mid s_t, a_t = 1) - \mathbb{E}(\cdot \mid s_t, a_t = 0)$. 
Brief Recount of the Hotz & Miller Method

Hotz & Miller method tackles $\beta \Delta \mathbb{E} \left( \bar{V}_{t+1}(s_{t+1}) \mid s_t \right)$ based on iteration of the following 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

$$
\ln \left( \frac{p_t(s_t)}{1 - p_t(s_t)} \right) = u_t(1, x_t) - u_t(0, x_t) + \beta \Delta \mathbb{E}(U^o_{t+1}(s_{t+1})|s_t)
$$

$$
+ \beta^2 \Delta \mathbb{E}(\mathbb{E}(U^o_{t+2}(s_{t+2})|s_{t+1})|s_t) + \cdots
$$

$$
+ \beta^{T_{tr} - t - 1} \Delta \mathbb{E}(\mathbb{E}(\cdots \mathbb{E}(U^o_{T_{tr}-1}(s_{T_{tr}-1})|s_{T_{tr}-2}) \cdots |s_{t+1})|s_t)
$$

$$
+ \beta^{T_{tr} - t} \Delta \mathbb{E}(\mathbb{E}(\cdots \mathbb{E}(\mathbb{E}(\bar{V}_{T_{tr}}(s_{T_{tr}})|s_{T_{tr}-1})|s_{T_{tr}-2}) \cdots |s_{t+1}|s_t),
$$

(1)

where $T_{tr} \leq T_{end}$ is some truncation period.
Brief Recount of the Hotz & Miller Method (cont’d)

1. To evaluate the RHS under hypothesized parameter values, the Hotz & Miller method estimates, simulates from, and integrates over state transitions distributions $f_{s_{t+1}|s_t,a_t}$ and $f_{\epsilon_{t+1}|s_{t+1},a_{t+1}}$, as well as CCP’s $Pr.(a_{t+1}|s_{t+1})$, for each $t$.

2. It matches the resulting LHS log odds ratios (or CCP’s) with actual ones from the data, solving for the parameter estimates via GMM.

- Computationally demanding when $d_s$ is large.
- Ad hoc parametric specification often assumed in practice.
To evaluate the RHS under hypothesized parameter values, the Hotz & Miller method estimates, simulates from, and integrates 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 CCP’s \( Pr.(a_{t+1}|s_{t+1}) \), for each \( t \).

It matches the resulting LHS log odds ratios (or CCP’s) with actual ones from the data, solving for the parameter estimates via GMM.

- Computationally demanding when \( d_s \) is large.
- Ad hoc parametric specification often assumed in practice.
To evaluate the RHS under hypthesized parameter values, the Hotz & Miller method estimates, simulates from, and integrates over state transitions distributions $f_{s_{t+1}|s_t,a_t}$ and $f_{\epsilon_{t+1}|s_{t+1},a_{t+1}}$, as well as CCP’s $Pr.(a_{t+1}|s_{t+1})$, for each $t$.

It matches the resulting LHS log odds ratios (or CCP’s) with actual ones from the data, solving for the parameter estimates via GMM.

- Computationally demanding when $d_s$ is large.
- Ad hoc parametric specification often assumed in practice.
To evaluate the RHS under hypothesized parameter values, the Hotz & Miller method estimates, simulates from, and integrates 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 CCP’s $Pr.(a_{t+1}|s_{t+1})$, for each $t$.

It matches the resulting LHS log odds ratios (or CCP’s) with actual ones from the data, solving for the parameter estimates via GMM.

- Computationally demanding when $d_s$ is large.
- Ad hoc parametric specification often assumed in practice.
Outline

1. A General Nonstationary Dynamic Binary Choice Model
   - Model Setup
   - Brief Recount of the Hotz & Miller Method

2. Our Approach
   - Transformation into a Linear System
   - Identification
   - Estimation

3. Bias When Interpreting Assumption 5 as Approximation

4. Conclusion
Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterative conditional means in $\Delta E(\bar{V}_{t+1}(s_{t+1}) | s_t)$ 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.

- Then, identification of the resulting linear system uses the usual argument of linear GMM.

- A 3-step CCP-based semiparametric estimation method follows straightforwardly.
Our Approach

We tackle the problem by obtaining a linear system in three steps:

1. Collapse the iterative conditional means in $\Delta \mathbb{E} \left( \tilde{V}_{t+1}(s_{t+1}) \mid s_t \right)$ 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.

Then, identification of the resulting linear system uses the usual argument of linear GMM.

A 3-step CCP-based semiparametric estimation method follows straightforwardly.
Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterative 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.

- Then, identification of the resulting linear system uses the usual argument of linear GMM.

- A 3-step CCP-based semiparametric estimation method follows straightforwardly.
Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterative conditional means in $\Delta \mathbb{E}(\tilde{V}_{t+1}(s_{t+1}) | s_t)$ 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.

- Then, identification of the resulting linear system uses the usual argument of linear GMM.

- A 3-step CCP-based semiparametric estimation method follows straightforwardly.
Our Approach

- We tackle the problem by obtaining a linear system in three steps:
  1. Collapse the iterative conditional means in $\Delta \mathbb{E} (\tilde{V}_{t+1}(s_{t+1}) \mid s_t)$ 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.

- Then, identification of the resulting linear system uses the usual argument of linear GMM.

- A 3-step CCP-based semiparametric estimation method follows straightforwardly.
Step 1: Collapse Iterative Conditional Means

Lemma 1 (Markovian $s_t$)

Under Assumptions 1-3, $s_t$ 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 (Collapsing iterative conditional means)

Under Assumptions 1-3, $\bar{V}_{t+1}(s_{t+1})$ can be simplified to

\[
\bar{V}_{t+1}(s_{t+1}) = U_{t+1}^0(s_{t+1}) + \sum_{\tau=t+2}^{T_{tr}-1} \beta^{\tau-t-1} \mathbb{E}(U^0_{\tau}(s_{\tau}) \mid s_{t+1})
\]

\[
+ \beta^{T_{tr}-t-1} \mathbb{E}(\bar{V}_{T_{tr}}(s_{T_{tr}}) \mid s_{t+1})
\]

for $T_{tr} < T_{end}$ and $t < T_{tr} - 1$. 

Chou, Ridder & Shi (2023)
Our Approach
Transformation into a Linear System

Step 1: Collapse Iterative Conditional Means

Lemma 1 (Markovian $s_t$)

Under Assumptions 1-3, $s_t$ 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 (Collapsing iterative conditional means)

Under Assumptions 1-3, $\tilde{V}_{t+1}(s_{t+1})$ can be simplified to

$$\tilde{V}_{t+1}(s_{t+1}) = U_{t+1}^o(s_{t+1}) + \sum_{\tau=t+2}^{T_{tr} - 1} \beta^{\tau-t-1} \mathbb{E}(U^o_{\tau}(s_{\tau}) \mid s_{t+1})$$

$$+ \beta^{T_{tr}-t-1} \mathbb{E}(\tilde{V}_{T_{tr}}(s_{T_{tr}}) \mid s_{t+1})$$ (2)

for $T_{tr} \leq T_{end}$ and $t \leq T_{tr} - 1$. 

Lemma 3 (Integrating out $s_{t+1}$)

Under Assumptions 1-3,

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

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

- Due to $s_{t+j} \perp (s_t, a_t) \mid s_{t+1}$, implied by Assumptions 1-3.
Step 1: Collapse Iterative Conditional Means (cont’d)

• Lemmas 1-3 lead to the simplification:

\[
\ln \left( \frac{p_t}{1 - p_t} \right) = u_t(1, x_t) - u_t(0, x_t) \\
+ \sum_{\tau = t+1}^{T_{tr}-1} \beta^{\tau-t} \Delta \mathbb{E}(\mathbb{E}(\mathbb{E}(\mathbb{E}(U_T^{o}(s_{\tau}) \mid s_{\tau-1}) \mid s_{\tau-2}) \cdots \mid s_{t+1}) \mid s_t) \\
+ \beta^{T_{tr}-t} \Delta \mathbb{E}(\mathbb{E}(\mathbb{E}(\tilde{V}_{T_{tr}}(s_{T_{tr}}) \mid s_{T_{tr}-1}) \mid s_{T_{tr}-2}) \cdots \mid s_{t+1}) \mid s_t) \\
= u_t(1, x_t) - u_t(0, x_t) \\
+ \sum_{\tau = t+1}^{T_{tr}-1} \beta^{\tau-t} \Delta \mathbb{E}(U_T^{o}(s_{\tau}) \mid s_t) + \beta^{T_{tr}-t} \Delta \mathbb{E}(\tilde{V}_{T_{tr}}(s_{T_{tr}}) \mid s_t). \quad (3)
\]

\[
\beta \Delta \mathbb{E}(\tilde{V}_{t+1}(s_{t+1}) \mid s_t)
\]
Step 2: Transform into a Partially Linear System

Assumption 4 (Linear flow utility)

For all \( t \in T \), \( u_t(0, x_t) = 0 \) and \( u_t(1, x_t) = x'_t \delta_t \) for some \( \delta_t \).

Lemma 4 (Expected optimal flow utility)

Under Assumptions 1-4,

\[
U^o_t = p_t x'_t \delta_t - p_t \ln(p_t) - (1 - p_t) \ln(1 - p_t).
\]

- Due to \( \mathbb{E}(\varepsilon_{a_t} | s_t, a_t) = \gamma - \ln(\Pr.(a_t|s_t)) \) with \( \gamma = \mathbb{E}(\varepsilon_{a_t}) = 0 \) (e.g., Hotz & Miller, 1993).
- Data horizon: \( T_{da} = \{1, \ldots, T\} \), with \( T_{\text{start}} < 1 \) and \( T_{\text{end}} > T \) allowed. We proceed with \( T_{tr} = T \).
Step 2: Transform into a Partially Linear System

Assumption 4 (Linear flow utility)

For all $t \in \mathcal{T}$, $u_t(0, x_t) = 0$ and $u_t(1, x_t) = x_t' \delta_t$ for some $\delta_t$.

Lemma 4 (Expected optimal flow utility)

Under Assumptions 1-4,

$$U_t^o = p_t x_t' \delta_t - p_t \ln(p_t) - (1 - p_t) \ln(1 - p_t).$$

- Due to $\mathbb{E}(\varepsilon_{at}|s_t, a_t) = \gamma - \ln(\text{Pr.}(a_t|s_t))$ with $\gamma = \mathbb{E}(\varepsilon_{at}) = 0$ (e.g., Hotz & Miller, 1993).
- Data horizon: $\mathcal{T}_{da} = \{1, \ldots, T\}$, with $T_{\text{start}} < 1$ and $T_{\text{end}} > T$ allowed. We proceed with $T_{\text{tr}} = T$. 
Step 2: Transform into a Partially Linear System

Assumption 4 (Linear flow utility)

For all \( t \in T \), \( u_t(0, x_t) = 0 \) and \( u_t(1, x_t) = x'_t \delta_t \) for some \( \delta_t \).

Lemma 4 (Expected optimal flow utility)

Under Assumptions 1-4,

\[
U^o_t = p_t x'_t \delta_t - p_t \ln(p_t) - (1 - p_t) \ln(1 - p_t).
\]

- Due to \( \mathbb{E}(\varepsilon_{at}|s_t, a_t) = \gamma - \ln(Pr.(a_t)|s_t) \) with \( \gamma = \mathbb{E}(\varepsilon_{at}) = 0 \) (e.g., Hotz & Miller, 1993).
- Data horizon: \( T_{da} = \{1, \ldots, T\} \), with \( T_{start} < 1 \) and \( T_{end} > T \) allowed. We proceed with \( T_{tr} = T \).
Step 2: Transform into a Partially Linear System (cont’d)

- Steps 1 & 2 together lead to a “triangular” partially linear system:

\[ y_{T-1} = x'_{T-1} \delta_{T-1} + \beta \Delta \mathbb{E}(\tilde{V}_T(x_T, z_T)|x_{T-1}, z_{T-1}), \quad \text{and} \quad (4a) \]

\[ y_t = x'_t \delta_t + \sum_{\tau=t+1}^{T-1} \beta^{T-t} \Delta \bar{x}_t^{\tau'} \delta_{\tau} + \beta^{T-t} \Delta \mathbb{E}(\tilde{V}_T(x_T, z_T)|x_t, z_t) \]

(4b)

for \( t = 1, \ldots, T - 2 \), where

- \( \Delta \bar{x}_t^{\tau} \equiv \Delta \mathbb{E}(p_{\tau} x_{\tau}|x_t, z_t) \);
- \( y_{T-1} \equiv \ln \left( \frac{p_{T-1}}{1-p_{T-1}} \right) ; \)
- \( y_t \equiv \ln \left( \frac{p_t}{1-p_t} \right) + \sum_{\tau=t+1}^{T-1} \beta^{T-t} \Delta \bar{\eta}_t^{\tau} ; \)
- \( \Delta \bar{\eta}_t^{\tau} \equiv \Delta \mathbb{E}(\eta_{\tau}|x_t, z_t), \eta_{\tau} \equiv p_{\tau} \ln(p_{\tau}) + (1 - p_{\tau}) \ln(1 - p_{\tau}) \) for \( \tau > t \).

- **Key**: \( y_t, x_t \) and \( \Delta \bar{x}_t^{\tau} \) are either directly observed, or can be estimated from the data.

- **Role of excluded variable** \( z_t \): might be able to identify \( \delta_t \) without additional assumption.
Step 2: Transform into a Partially Linear System (cont’d)

- Steps 1 & 2 together lead to a “triangular” partially linear system:

\[
y_{T-1} = x_{T-1}' \delta_{T-1} + \beta \Delta \mathbb{E}(\bar{V}_T(x_T, z_T) | x_{T-1}, z_{T-1}), \text{ and } (4a)
\]

\[
y_t = x_t' \delta_t + \sum_{\tau=t+1}^{T-1} \beta^{T-t} \Delta \bar{x}_t^{\tau} \delta_\tau + \beta^{T-t} \Delta \mathbb{E}(\bar{V}_T(x_T, z_T) | x_t, z_t)
\]

(4b)

for \( t = 1, \ldots, T - 2 \), where

- \( \Delta \bar{x}_t^{\tau} \equiv \Delta \mathbb{E}(p_\tau x_\tau | x_t, z_t) \);
- \( y_{T-1} \equiv \ln \left( \frac{p_{T-1}}{1-p_{T-1}} \right) \);
- \( y_t \equiv \ln \left( \frac{p_t}{1-p_t} \right) + \sum_{\tau=t+1}^{T-1} \beta^{T-t} \Delta \eta_t^{\tau} \);
- \( \Delta \eta_t^{\tau} \equiv \Delta \mathbb{E}(\eta_\tau | x_t, z_t) \), \( \eta_\tau \equiv p_\tau \ln(p_\tau) + (1 - p_\tau) \ln(1 - p_\tau) \) for \( \tau > t \).

- **Key:** \( y_t, x_t \) and \( \Delta \bar{x}_t^{\tau} \) are either directly observed, or can be estimated from the data.

- **Role of excluded variable** \( z_t \): might be able to identify \( \delta_t \) without additional assumption.
Our Approach

Transformation into a Linear System

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

- Steps 1 & 2 together lead to a “triangular” partially linear system:

\[ y_{T-1} = x'_{T-1} \delta_{T-1} + \beta \Delta \mathbb{E}(\bar{V}_T(x_T, z_T) | x_{T-1}, z_{T-1}), \quad \text{and} \quad (4a) \]

\[ y_t = x'_t \delta_t + \sum_{\tau=t+1}^{T-1} \beta^{T-t} \Delta \bar{x}_t^{\tau} \delta_\tau + \beta^{T-t} \Delta \mathbb{E}(\bar{V}_T(x_T, z_T) | x_t, z_t) \]

\[ (4b) \]

for \( t = 1, \ldots, T - 2 \), where

- \( \Delta \bar{x}_t^{\tau} \equiv \Delta \mathbb{E}(p_{\tau} x_{\tau} | x_t, z_t) \);
- \( y_{T-1} \equiv \ln \left( \frac{p_{T-1}}{1-p_{T-1}} \right) \);
- \( y_t \equiv \ln \left( \frac{p_t}{1-p_t} \right) + \sum_{\tau=t+1}^{T-1} \beta^{T-t} \Delta \bar{\eta}_t^{\tau} \);
- \( \Delta \bar{\eta}_t^{\tau} \equiv \Delta \mathbb{E}(\eta_{\tau} | x_t, z_t), \quad \eta_{\tau} \equiv p_{\tau} \ln(p_{\tau}) + (1 - p_{\tau}) \ln(1 - p_{\tau}) \) for \( \tau > t \).

**Key:** \( y_t, x_t \) and \( \Delta \bar{x}_t^{\tau} \) are either directly observed, or can be estimated from the data.

**Role of excluded variable** \( z_t \): might be able to identify \( \delta_t \) without additional assumption.
Step 3: Transform into a Linear System

Assumption 5 (Last-data-period expected value function)

There exists a $K \times 1$ vector of parameters $\gamma^K$, such that $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)\gamma^K$, with $q^K(x, z)$ being a $K \times 1$ vector-valued known functions.

- Under Assumptions 1-5, Steps 1-3 lead to a linear system:

\begin{align*}
  y_{T-1} &= x'_{T-1}\delta_{T-1} + \beta\Delta\bar{q}'_{T-1}\gamma^K, \quad (5a) \\
  y_t &= x'_t\delta_t + \sum_{\tau=t+1}^{T-1} \beta^{\tau-t}\Delta\bar{x}'_{\tau}\delta_{\tau} + \beta^{T-t}\Delta\bar{q}'_{t}\gamma^K, \quad (5b)
\end{align*}

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

- Special case: $\bar{V}_T(x_T, z_T) = x'_T\delta_T$ and $\gamma^K = \delta_T$ if $T = T_{end}$.

- We can quantify the bias in estimates when Assumption 6 is interpreted as an approximation only.
Step 3: Transform into a Linear System

Assumption 5 (Last-data-period expected value function)

There exists a $K \times 1$ vector of parameters $\gamma^K$, such that $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)' \gamma^K$, with $q^K(x, z)$ being a $K \times 1$ vector-valued known functions.

- Under Assumptions 1-5, Steps 1-3 lead to a linear system:

$$y_{T-1} = x'_{T-1} \delta_{T-1} + \beta \Delta \bar{q}'_{T-1} \gamma^K,$$

$$y_t = x'_t \delta_t + \sum_{\tau=t+1}^{T-1} \beta^{T-t} \Delta x'^{\tau} \delta_{\tau} + \beta^{T-t} \Delta \bar{q}'_t \gamma^K,$$

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

- Special case: $\bar{V}_T(x_T, z_T) = x'_T \delta_T$ and $\gamma^K = \delta_T$ if $T = T_{end}$.

- We can quantify the bias in estimates when Assumption 6 is interpreted as an approximation only.
Step 3: Transform into a Linear System

Assumption 5 (Last-data-period expected value function)

There exists a $K \times 1$ vector of parameters $\gamma^K$, such that $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)' \gamma^K$, with $q^K(x, z)$ being a $K \times 1$ vector-valued known functions.

- Under Assumptions 1-5, Steps 1-3 lead to a linear system:

  \begin{align*}
  y_{T-1} &= x_{T-1}' \delta_{T-1} + \beta \Delta \bar{q}_t^{K'} \gamma^K, \quad (5a) \\
  y_t &= x_t' \delta_t + \sum_{\tau=t+1}^{T-1} \beta^{\tau-t} \Delta \bar{x}_t^{\tau'} \delta_{\tau} + \beta^{T-t} \Delta \bar{q}_t^{K'} \gamma^K, \quad (5b)
  \end{align*}

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

  - Special case: $\bar{V}_T(x_T, z_T) = x_T' \delta_T$ and $\gamma^K = \delta_T$ if $T = T_{end}$.

  - We can quantify the bias in estimates when Assumption 6 is interpreted as an approximation only.
Step 3: Transform into a Linear System

Assumption 5 (Last-data-period expected value function)

*There exists a $K \times 1$ vector of parameters $\gamma^K$, such that $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)' \gamma^K$, with $q^K(x, z)$ being a $K \times 1$ vector-valued known functions.*

- Under Assumptions 1-5, Steps 1-3 lead to a linear system:

  
  \[
  y_{T-1} = x'_{T-1} \delta_{T-1} + \beta \Delta \bar{q}'_{T-1} \gamma^K, \tag{5a}
  \]

  \[
  y_t = x'_t \delta_t + \sum_{\tau = t+1}^{T-1} \beta^{\tau-t} \Delta \bar{x}'_\tau \delta_\tau + \beta^{T-t} \Delta \bar{q}'_t \gamma^K, \tag{5b}
  \]

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

- **Special case:** $\bar{V}_T(x_T, z_T) = x'_T \delta_T$ and $\gamma^K = \delta_T$ if $T = T_{end}$.

- We can quantify the bias in estimates when Assumption 6 is interpreted as an approximation only.
Identification of Linear System

**Proposition 1 (Identification of $\delta$ and $\gamma^K$)**

If $T \geq 2$, and $L^*_{T-1}$ and $\mathbb{E}(x_t x'_t)$ for $t = 1, \ldots, T - 2$ are invertible, then $(\delta'_1, \ldots, \delta'_{T-1}, \gamma^{K'})'$ is identified.

\[
L = \begin{pmatrix}
0_{d_x \times (T-4)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'}) \\
0_{d_x \times (T-4)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'}) \\
0_{d_x \times (T-4)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'}) & \beta^2 \mathbb{E}(x_{T-2} \Delta \bar{q}_{T-2}^{K'}) \\
0_{d_x \times (T-4)d_x} & 0_{d_x \times d_x} & \beta \mathbb{E}(\Delta \bar{x}_{T-2}^{T-1} x'_T) & \beta^2 \mathbb{E}(\Delta \bar{x}_{T-2}^{T-1} \Delta \bar{x}_{T-2}^{T-1'}) & \beta^3 \mathbb{E}(\Delta \bar{x}_{T-2}^{T-1} \Delta \bar{q}_{T-2}^{K'}) \\
0_{d_x \times (T-4)d_x} & 0_{K \times d_x} & \beta \mathbb{E}(\Delta \bar{q}_{T-2}^{K} x'_{T-2}) & \beta^2 \mathbb{E}(\Delta \bar{q}_{T-2}^{K} \Delta \bar{x}_{T-2}^{T-1'}) & \beta^3 \mathbb{E}(\Delta \bar{q}_{T-2}^{K} \Delta \bar{q}_{T-2}^{K'}) \\
0_{d_x \times (T-4)d_x} & \mathbb{E}(x_{T-3} x'_{T-3}) & \beta x_{T-3} \mathbb{E}(\Delta \bar{x}_{T-3}^{T-2'}) & \beta^2 \mathbb{E}(x_{T-3} \Delta \bar{x}_{T-3}^{T-1'}) & \beta^3 \mathbb{E}(x_{T-3} \Delta \bar{q}_{T-3}^{K'}) \\
0_{d_x \times (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'}) & \beta^3 \mathbb{E}(\Delta \bar{x}_{T-3}^{T-2} \Delta \bar{x}_{T-3}^{T-1'}) & \beta^4 \mathbb{E}(\Delta \bar{x}_{T-3}^{T-2} \Delta \bar{q}_{T-3}^{K'}) \\
0_{d_x \times (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'}) & \beta^4 \mathbb{E}(\Delta \bar{x}_{T-3}^{T-1} \Delta \bar{x}_{T-3}^{T-1'}) & \beta^5 \mathbb{E}(\Delta \bar{x}_{T-3}^{T-1} \Delta \bar{q}_{T-3}^{K'}) \\
0_{d_x \times (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'}) & \beta^5 \mathbb{E}(\Delta \bar{q}_{T-3}^{K} \Delta \bar{x}_{T-3}^{T-1'}) & \beta^6 \mathbb{E}(\Delta \bar{q}_{T-3}^{K} \Delta \bar{q}_{T-3}^{K'}) \\
0_{K \times (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'}) & \beta^5 \mathbb{E}(\Delta \bar{q}_{T-3}^{K} \Delta \bar{x}_{T-3}^{T-1'}) & \beta^6 \mathbb{E}(\Delta \bar{q}_{T-3}^{K} \Delta \bar{q}_{T-3}^{K'}) \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\end{pmatrix}
\]
3-Step CCP-Based Semiparametric Estimation

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 $\hat{\Delta} \eta$, $\hat{\Delta} \bar{X}$ and $\hat{\Delta} \bar{q}$.

3. Let $\bar{m}_{N}(\delta, \gamma^K) \equiv \frac{1}{N} \sum_{i=1}^{N} m(x_i, z_i, a_i, \delta, \gamma^K, \hat{p}, \hat{\Delta} \eta, \hat{\Delta} \bar{X}, \hat{\Delta} \bar{q}^K)$ capture the distance between LHS & RHS of the linear system. Then, the minimum-distance estimator $(\hat{\delta}', \hat{\gamma}'^K)$ solves:

$$
(\hat{\delta}', \hat{\gamma}'^K) \equiv \arg \min_{\delta \in \mathbb{R}^{(T-1)d_x}, \gamma^K \in \mathbb{R}^K} \bar{m}_{N}(\delta, \gamma^K)' W_N \bar{m}_{N}(\delta, \gamma^K).
$$
Our Approach

3-Step CCP-Based Semiparametric Estimation

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 \( \hat{\Delta}_{\eta}, \hat{\Delta}_{x} \) and \( \hat{\Delta}_{q} \).

3. Let \( \bar{m}_N(\delta, \gamma^K) \equiv \frac{1}{N} \sum_{i=1}^{N} m(x_i, z_i, a_i, \delta, \gamma^K, \hat{p}, \hat{\Delta}_{\eta}, \hat{\Delta}_{x}, \hat{\Delta}_{q^K}) \) capture the distance between LHS & RHS of the linear system. Then, the minimum-distance estimator \( (\hat{\delta}', \hat{\gamma}'_K) \) solves:

\[
(\hat{\delta}', \hat{\gamma}'_K) \equiv \arg \min_{\delta \in \mathbb{R}^{(T-1)d_x}, \gamma_K \in \mathbb{R}^K} \bar{m}_N(\delta, \gamma^K)' W_N \bar{m}_N(\delta, \gamma^K).
\]
3-Step CCP-Based Semiparametric Estimation

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 $\hat{\Delta \bar{\eta}}$, $\hat{\Delta \bar{x}}$ and $\hat{\Delta \bar{q}}$.

3. Let $\bar{m}_N(\delta, \gamma^K) \equiv \frac{1}{N} \sum_{i=1}^{N} m(x_i, z_i, a_i, \delta, \gamma^K, \hat{p}, \hat{\Delta \bar{\eta}}, \hat{\Delta \bar{x}}, \hat{\Delta \bar{q}}^K)$ capture the distance between LHS & RHS of the linear system. Then, the minimum-distance estimator $(\hat{\delta}', \hat{\gamma}'_K)'$ solves:

\[
(\hat{\delta}', \hat{\gamma}'_K)' \equiv \arg \min_{\delta \in \mathbb{R}^{(T-1)dx}, \gamma_K \in \mathbb{R}^K} \bar{m}_N(\delta, \gamma^K)' W_N \bar{m}_N(\delta, \gamma^K).
\]
Proposition 2 (Asymptotic distribution of $\hat{\delta}$)

$$\sqrt{N} \left( \hat{\delta} - \delta \right) \xrightarrow{d} \mathcal{N} \left( 0, \mathbb{E}[\psi_\delta(x_i, z_i, a_i)\psi'_\delta(x_i, z_i, a_i)] \right),$$

where $\psi_\delta$ takes a tedious form given in the paper.

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

1. A General Nonstationary Dynamic Binary Choice Model
   - Model Setup
   - Brief Recount of the Hotz & Miller Method

2. Our Approach
   - Transformation into a Linear System
   - Identification
   - Estimation

3. Bias When Interpreting Assumption 5 as Approximation

4. Conclusion
Bias Induced by Assumption $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)' \gamma^K$

- 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:

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

$$\|\delta^K_{\text{pseudo}} - \delta\| = O \left( K^{-\frac{m}{d_s}} \right).$$
Bias Induced by Assumption $\bar{V}_T(x_T, z_T) = q^K(x_T, z_T)'\gamma^K$

- 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 $\Delta \tilde{r}_t^K \equiv \Delta \mathbb{E}(r^K(x_T, z_T)|x_t, z_t)$ 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

$$\mathbb{E}[(\Delta \tilde{r}_t^K)^2] = O \left( K^{-\frac{2m}{ds}} \right).$$

This leads to:

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

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

1. A General Nonstationary Dynamic Binary Choice Model
   - Model Setup
   - Brief Recount of the Hotz & Miller Method

2. Our Approach
   - Transformation into a Linear System
   - Identification
   - Estimation

3. Bias When Interpreting Assumption 5 as Approximation

4. Conclusion
We transform the CCP equation into a linear system under common and mild new assumptions.

We provide testable sufficient conditions for identification.

The estimation method avoids estimating & simulating from state transition distributions.

**Remark:** state transition distributions are essential in *counterfactual* analysis, and our estimation method may help choosing parametric specification there.