Dynamic Discrete Choice Model and Near Finite Dependence CCP Estimator

Yu Hao [1] Hiroyuki Kasahara [2]

[1]The University of Hong Kong

[2]The University of British Columbia

ASSA 2024

Bellman equation and the Conditional Choice Probabilities

$$
\begin{aligned}
\mathrm{V}(x) &= \int \max_{d\in\{0,1\}} \{\mathrm{v}_\theta(x,d) + \epsilon(d)\}\, g(\epsilon|x) d\epsilon \\
&:= [\Gamma(\theta, V)](x) \qquad \text{(Bellman operator)},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{P}(d=1|x;\theta) &= \frac{1}{1+\exp\left(\mathrm{v}_\theta(x,0) - \mathrm{v}_\theta(x,1)\right)} \qquad \text{(Type I EV)} \\
&:= [\Lambda(\theta, V)](a|x).
\end{aligned}
$$

where

$$
\mathrm{v}_\theta(x,d) := u_\theta(x,d) + \beta \sum_{x^\dagger} \mathrm{V}(x^\dagger) f(x^\dagger|x,d).
$$

Maximum Likelihood Estimator (e.g., Rust (1987))

$$\max_{\theta} \sum_{i=1}^{n} \ln[\Lambda(\theta, V_\theta)](d_i|x_i)$$

where

$$V_\theta = \Gamma(\theta, V_\theta).$$

Hotz and Miller's CCP estimator

- The Bellman equation: a fixed point problem in the space of value functions.

$$V = \Gamma(\theta, V).$$

- The model's restriction can be also characterized by a fixed point problem in the space of conditional choice probabilities:

$$P = \Psi(\theta, P).$$

Hotz and Miller's CCP estimator

$$V(x) = \sum_{d \in \{0,1\}} \mathrm{P}(d|x) \left\{ u_\theta(x,d) + \underbrace{E[\epsilon(d)|\mathrm{P}(d|x)]}_{-\ln \mathrm{P}(d|x)} + \beta \sum_{x^\dagger \in X} V(x') f(x'|x,d) \right\},$$

- The Policy iteration mapping:

$$\Psi_{HM}(\theta, P) := \Lambda(\theta, \varphi(\theta, P)).$$

where

$$V = (I - \beta E_P)^{-1} u_{\theta,P} := \varphi(\theta, P).$$

- CCP estimator:

$$\max_\theta \sum_{i=1}^{n} \ln[\Psi_{HM}(\theta, \hat{P})](d_i|x_i)$$

Alternative policy function mappings under finite dependence

- Finite dependence: there exists a sequence of future choices such that the subsequent continuation values do not depend on the current choice.
- Engine replacement at $t+1$ (i.e., $a_{t+1} = 1$) $\Rightarrow$ continuation value at $t+2$ does not depend on the current choice a_t.

Example of finite dependence (Bus engine replacement model)

$$\max_{d_1,d_2,\ldots} E\left[\sum_{t=1}^{\infty} \beta^t \left\{u_\theta(x_t, d_t) + \epsilon_t(d_t)\right\}\right],$$

where

$$\mathrm{u}_\theta(x_t, d_t) = \begin{cases} 0, & \text{if } d_t = 1 \\ \theta_0 + \theta_1 x_t, & \text{if } d_t = 0, \end{cases}$$

- x_t: mileage
- $f(x_{t+1}|x_t, d_t)$: transition probability with $f(0|x_t, 1) = 1$
- $\epsilon_t = (\epsilon_t(0), \epsilon_t(1))' \overset{iid}{\sim}$ Type I EV

An estimator by Arcidiacono and Miller (2011)

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \ln[\Psi_{FD}(\theta, \hat{P}, \hat{f})](d_i|x_i).$$

$$[\Psi_{FD}(\theta, P, f)](d = 1|x)$$
$$= \frac{1}{1 + \exp\left(\theta_0 + \theta_1 x + \beta \sum_{x^\dagger} \log \mathrm{P}(1|x^\dagger)\left(f(x^\dagger|x, 1) - f(x^\dagger|x, 0)\right)\right)}$$

$\rightarrow$ Given $\hat{P}$ and $\hat{f}$, **estimating** $(\theta_0, \theta_1, \beta)$ **is as easy as estimating a logit model!**

CCP under finite dependence

$$P_\theta(1|x) = \frac{1}{1 + \exp(\mathrm{v}_\theta(x,0) - \mathrm{v}_\theta(x,1))},$$

where

$$\mathrm{v}_\theta(x,d) := \mathrm{u}_\theta(x,d) + \beta \sum_{x^\dagger} \mathrm{V}(x^\dagger) f(x^\dagger|x,d).$$

We can show that

$$\begin{aligned}\mathrm{V}(x) &= \mathrm{v}_\theta(x,1) - \log \mathrm{P}(1|x) = \mathrm{v}_\theta(x,0) - \log \mathrm{P}(0|x) \\ &= w(\mathrm{v}_\theta(x,1) - \log \mathrm{P}(1|x)) + (1-w)(\mathrm{v}_\theta(x,0) - \log \mathrm{P}(0|x)).\end{aligned}$$

CCP under finite dependence

$$
\begin{aligned}
\mathrm{v}_\theta(x,d) &= \mathrm{u}_\theta(x,d) + \beta \sum_{x^\dagger} \mathrm{V}(x^\dagger) f(x^\dagger|x,d) \\
\mathrm{V}(x^\dagger) &= \mathrm{v}_\theta(x^\dagger,1) - \log \mathrm{P}(1|x^\dagger) \\
&= u_\theta(x^\dagger,1) - \log \mathrm{P}(1|x^\dagger) + \underbrace{\beta \sum_{x^{\dagger\dagger}} \mathrm{V}(x^{\dagger\dagger}) f(x^{\dagger\dagger}|x^\dagger,1)}_{=\mathrm{V}(0)}.
\end{aligned}
$$

$\Rightarrow$

$$
\begin{aligned}
\mathrm{v}_\theta(x,d) = \mathrm{u}_\theta(x,d) &+ \beta \sum_{x^\dagger} \left\{ u_\theta(x^\dagger,1) - \log \mathrm{P}(1|x^\dagger) \right\} f(x^\dagger|x,d) \\
&+ \beta^2 \mathrm{V}(0) \sum_{x'} f(x^\dagger|x,d).
\end{aligned}
$$

CCP under decision-specific finite dependence Arcidiacono and Miller (2011)

$$
\begin{aligned}
\mathrm{v}_\theta(x,0) - \mathrm{v}_\theta(x,1) = {} & u_\theta(x,0) - u_\theta(x,1) \\
& + \beta \sum_{x^\dagger} \left\{ u_\theta(x^\dagger,1) - \log \mathrm{P}(1|x^\dagger) \right\} \tilde{f}(x^\dagger|x) \\
& + \beta^2 \mathrm{V}(0) \underbrace{\sum_{x^\dagger} \tilde{f}(x^\dagger|x)}_{=0}
\end{aligned}
$$

where

$$\tilde{f}(x^\dagger|x) := f(x^\dagger|x,0) - f(x^\dagger|x,1).$$

$$
\begin{aligned}
P_\theta(1|x) &= \frac{1}{1+\exp(\mathrm{v}_\theta(x,0) - \mathrm{v}_\theta(x,1))} \\
&= \frac{1}{1+\exp\left(\theta_0 + \theta_1 x_t - \beta \sum_{x^\dagger} \log \mathrm{P}(1|x^\dagger) \tilde{f}(x^\dagger|x)\right)}.
\end{aligned}
$$

CCP without finite dependence

$$[\Psi(\theta, P, f)](d = 1|x) = \frac{1}{1 + \exp\left(\tilde{\mathrm{v}}_\theta(x)\right)}$$

$$\tilde{\mathrm{v}}_\theta(x) = \tilde{\mathrm{u}}_\theta(x) + \beta \sum_{x^\dagger} \log \mathrm{P}(1|x^\dagger)\tilde{f}(x^\dagger|x) + \beta^2 \sum_{x^{\dagger\dagger}} \mathrm{V}(x^{\dagger\dagger}) \sum_{x^\dagger} f(x^{\dagger\dagger}|x^\dagger, 1)\tilde{f}(x^\dagger|x)$$

But:

$$\tilde{\mathrm{v}}_\theta(x, d) = \tilde{\mathrm{u}}_\theta(x, d) + \beta \sum_{x^\dagger} \mathrm{V}(x^\dagger)\tilde{f}(x^\dagger|x, d)$$

$$\begin{aligned} \text{where } \mathrm{V}(x^\dagger) &= w(\mathrm{v}_\theta(x^\dagger, 1) - \log \mathrm{P}(1|x^\dagger)) \\ &\quad + (1 - w)(\mathrm{v}_\theta(x^\dagger, 0) - \log \mathrm{P}(0|x^\dagger)) \\ &= u_\theta^w(x^\dagger) - p^w(x^\dagger) + \sum_{x^{\dagger\dagger}} \mathrm{V}(x^{\dagger\dagger}) f^w(x^{\dagger\dagger}|x^\dagger), \end{aligned}$$

where $p^w(x) = w \log \mathrm{P}(1|x) + (1 - w) \log \mathrm{P}(0|x)$ etc.

Near Finite Dependence

- **Goal:** Select weights $\{w_{x,d}(x^\dagger, d^\dagger)\}$ to satisfy for each (x, d):

$$\sum_{x^\dagger} \left(\sum_{d^\dagger \in \{0,1\}} w_{x,d}(x^\dagger, d^\dagger) f(x^{\dagger\dagger} | x^\dagger, d^\dagger) \right) \tilde{f}(x^\dagger | x, d) \approx 0.$$

- **Advancement:** In Arcidiacono and Miller (2019), weights $\{w(x^\dagger, d^\dagger)\}$ are not (x, d)-specific.
- **Impact:** This choice grants $(|XD|)^2$ degrees of freedom—versus $(|XD|)$—for finer norm minimization.

Our proposed estimator

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \ln[\Psi_{AFD}(\theta, \hat{P}, \hat{f})](d_i|x_i).$$

$$[\Psi_{AFD}(\theta, P, f)](d = 1|x) = \frac{1}{1 + \exp\left(\mathrm{v}_\theta(x, 0) - \mathrm{v}_\theta(x, 1)\right)}$$

where,

$$\begin{aligned}\mathrm{v}_\theta(x, 0) - \mathrm{v}_\theta(x, 1) = {} & u_\theta(x, 0) - u_\theta(x, 1) \\ & + \beta \sum_{x^\dagger} \left(u_\theta^{w_{x,x^\dagger}}(x^\dagger) - p^{w_{x,x^\dagger}}(x^\dagger) \right) \tilde{f}(x^\dagger|x) \\ & + \beta^2 \sum_{x^{\dagger\dagger}} \mathrm{V}(x^{\dagger\dagger}) \underbrace{\sum_{x'} f^{w_{x,x^\dagger}}(x^{\dagger\dagger}|x^\dagger)\tilde{f}(x^\dagger|x)}_{\approx 0}\end{aligned}$$

Computing $\mathbf{w}$

The objective is to minimize the weights, which can be formulated as:

$$\mathbf{w} = \arg\min_{\mathbf{w}} \tilde{\mathbf{F}}\mathbf{F}(\mathbf{w})$$

Here, $\tilde{\mathbf{F}}$ represents the difference between two Markov transition matrices:

$$\tilde{\mathbf{F}} = \mathbf{F}_1 - \mathbf{F}_0$$

We derive a closed-form solution for $\mathbf{w}$, as shown below:

$$\mathbf{w} = -\tilde{\mathbf{F}}\mathbf{F}_0(\tilde{\mathbf{F}})^+ \tag{1}$$

where $(\tilde{\mathbf{F}})^+$ denotes the Moore-Penrose pseudo-inverse of $\tilde{\mathbf{F}}$. The relationship between $\tilde{\mathbf{F}}$, $\mathbf{F}(\mathbf{w})$, and $\mathbf{w}$ is given by:

$$\tilde{\mathbf{F}}\mathbf{F}(\mathbf{w}) = \mathbf{w}\tilde{\mathbf{F}} + \tilde{\mathbf{F}}\mathbf{F}_0 \tag{2}$$

Extend to ρ-period

Simplifying the two-period weight minimization, we approach the problem sequentially:

1. Obtain initial weights $\mathbf{w}^{(1)}$ by solving $\mathbf{w}^{(1)} = \arg\min_{\mathbf{w}^{(1)}} \tilde{\mathbf{F}}\mathbf{F}(\mathbf{w}^{(1)})$.
2. Minimize the influence of future decisions by deriving $\mathbf{w}^{(2)}$ as $\mathbf{w}^{(2)} = \arg\min_{\mathbf{w}} \tilde{\mathbf{F}}\mathbf{F}(\mathbf{w}^{(1)})\mathbf{F}(\mathbf{w}^{(2)})$.

The second-period weights $\mathbf{w}^{(2)}$ are functionally dependent on $\mathbf{w}^{(1)}$, $\mathbf{w}^{(2)}\big(\mathbf{w}^{(1)}\big)$, enabling us to optimize $\mathbf{w}^{(1)}$ for two-period finite dependence effectively.

Dimensionality Reduction

Objective: Preventing the curse of dimensionality in sequential mapping by avoiding recursive multiplication of large matrices.
Strategy: Partition the state into two parts:

- ω - states affected by the decision
- z - exogenous state unaffected by the decision

By defining $\grave{\mathbf{w}}^{(1)}$:

$$\grave{\mathbf{w}}^{(1)} = -\tilde{\mathbf{F}}\mathbf{F}_0(\tilde{\mathbf{F}})^+ = \underbrace{-(\tilde{\mathbf{F}}_\omega \mathbf{F}_{\omega,0} \tilde{\mathbf{F}}_\omega^+)}_{\grave{\mathbf{w}}_\omega^{(1)}} \otimes \mathbf{F}_z,$$

$$\tilde{\mathbf{F}}\mathbf{F}(\mathbf{w}) = \tilde{\mathbf{F}}_\omega \mathbf{F}_{\omega,0}(\mathbf{I} - (\tilde{\mathbf{F}}_\omega^+ \tilde{\mathbf{F}}_\omega)) \otimes \mathbf{F}_z.$$

Simulation Overview

- Dynamic entry/exit problem based on Aguirregabiria and Magesan (2016), incorporating history in firm profits with vector θ.
- State variables $x = (z_1, z_2, z_3, z_4, \omega, y)$ represent market conditions and firm-specific factors; firms decide to operate or exit based on (x, ϵ).
- Flow payoff $u(d_t, x_t; \theta)$:

$$\begin{aligned} u(d_t, x_t; \theta) &= d_t(VP_t - EC_t - FC_t) \\ \text{where } VP_t &= \exp(\omega)[\theta_0^{VP} + \theta_1^{VP} z_{1t} + \theta_2^{VP} z_{2t}] \\ FC_t &= [\theta_0^{FC} + \theta_1^{FC} z_{3t}] \\ EC_t &= (1 - y_t)[\theta_0^{EC} + \theta_1^{EC} z_{4t}]. \end{aligned}$$

- Shocks are AR(1) processes, leading to a state space of dimension $X = 2 \cdot K_z^4 \cdot K_o$ with action-independent transitions.

Transition Probability Formulas

1. Entry cost and fixed cost shocks (z_{jt}): Independent of chosen action

$$f(z_j'|z_j)=\begin{cases}\Phi([z_j^{(1)}+(\omega_j^{(1)}/2)-\gamma_0^j-\gamma_1^j z]/\sigma_j) & z'=z_j^{(1)};\\ 1-\Phi([z_j^{(K-1)}+(\omega_j^{(K-1)}/2)-\gamma_0^j-\gamma_1^j z]/\sigma_j) & z'=z_j^{(K)}.\\ \Phi([z_j^{(k)}+(\omega_j^{(k)}/2)-\gamma_0^j-\gamma_1^j z]/\sigma_j)- & \\ \qquad \Phi([z_j^{(k-1)}+(\omega_j^{(k-1)}/2)-\gamma_0^j-\gamma_1^j z]/\sigma_j) & \text{otherwise;}\end{cases}$$

2. Productivity shock (ω_t): Dependent on chosen action

$$f(\omega'|\omega,d)=\begin{cases}\Phi([\omega^{(1)}+(\omega^{(1)}/2)-\gamma_0^\omega-\gamma_1^\omega\omega-\gamma_d d]/\sigma) & \omega'=\omega^{(1)};\\ 1-\Phi([\omega^{(K-1)}+(\omega^{(K-1)}/2)-\gamma_0^\omega-\gamma_1^\omega\omega-\gamma_d d]/\sigma) & \omega'=\omega^{(K)}.\\ \Phi([\omega^{(k)}+(\omega^{(k)}/2)-\gamma_0^\omega-\gamma_1^\omega\omega-\gamma_d d]/\sigma)- & \\ \qquad \Phi([\omega^{(k-1)}+(\omega^{(k-1)}/2)-\gamma_0^\omega-\gamma_1^\omega\omega-\gamma_d d]/\sigma) & \text{otherwise}\end{cases}$$

Weight Solving

N state	γ_a	Time				Singular Value	
		HM Inverse	$\mathbf{w}$	$\mathbf{H}(\cdot)$	$\mathbf{h}(\cdot)$	$\Delta\mathbf{F}^{(1)}$	$\Delta\mathbf{F}^{(2)}$
5184	0	11.250	0.050	0.120	0.100	1.93E-17	4.15e-17
5184	0.8	17.980	0.080	0.140	0.120	0.048	6.35e-17
5184	1.2	12.580	0.110	0.210	0.200	0.111	5.10e-17
7776	0	40.900	0.180	0.450	0.440	1.41e-16	1.88e-17
7776	0.8	55.520	0.140	0.320	0.250	0.164	8.32e-17
7776	1.2	50.890	0.140	0.320	0.280	0.341	5.59e-17
10368	0	123.260	0.210	0.560	0.480	3.68E-16	7.35e-17
10368	0.8	128.190	0.220	0.540	0.410	0.184	6.26e-16
10368	1.2	124.260	0.180	0.430	0.380	0.420	1.58e-15
12960	0	210.980	0.270	0.840	0.620	2.78E-16	5.06e-17
12960	0.8	244.600	0.330	0.800	0.680	0.197	2.64e-13
12960	1.2	245.770	0.300	0.680	0.600	0.411	6.90e-13

Define the sequence $\Delta\mathbf{F}^{(k)}{}_{k=0}^{\rho}$ in $\mathcal{F}$ recursively by

$$\Delta\mathbf{F}^{(k)} \equiv \begin{cases} \tilde{\mathbf{F}}, & \text{if } k = 0, \\ \Delta\mathbf{F}^{(k-1)}\mathbf{F}(\mathbf{w}^{(k)}), & \text{for } k = 1, 2, \ldots, \rho, \end{cases} \tag{3}$$

Parameter Estimates

Table 1: Non-stationary Model: $nM = 500, nT = 4, nMC = 100$

	VP0	VP1	VP2	FC0	FC1	EC0	EC1	Time	ρ_1	ρ_2
True θ	0.5	1.0	-1.0	0.5	1.0	1.0	1.0			
X=2560, γ_a = 0										
AFD	0.501	1.009	-1.013	0.488	1.019	1.017	1.004	0.45	6.83E-17	9.50E-17
	(0.153)	(0.085)	(0.086)	(0.212)	(0.081)	(0.230)	(0.085)	0.40		
AFD2	0.500	1.001	-1.005	0.485	1.016	1.022	1.003	1.03	1.39E-16	3.17E-16
	(0.065)	(0.044)	(0.043)	(0.154)	(0.071)	(0.223)	(0.084)	0.99		
AM	0.495	1.014	-1.017	0.451	0.997	1.029	0.993	27111.56	1.55E-16	2.55E-16
	(0.152)	(0.085)	(0.086)	(0.214)	(0.078)	(0.232)	(0.085)	27111.52		
X=2560, γ_a = 1										
AFD	0.510	1.010	-1.012	0.266	0.943	0.978	0.991	0.42	2.49E-01	2.52E-01
	(0.071)	(0.049)	(0.051)	(0.277)	(0.085)	(0.239)	(0.086)	0.38		
AFD2	0.526	1.007	-1.015	0.523	1.012	0.984	1.003	0.97	5.57E-10	1.50E-09
	(0.178)	(0.117)	(0.116)	(0.310)	(0.126)	(0.254)	(0.111)	0.93		
AM	0.443	1.080	-1.087	0.059	0.955	0.984	1.029	7511.62	4.94E-01	5.92E-01
	(0.174)	(0.126)	(0.128)	(0.500)	(0.098)	(0.245)	(0.105)	7511.57		

Parameter Estimates

Table 2: Non-stationary Model: $nM = 500, nT = 4, nMC = 100$

	VP0	VP1	VP2	FC0	FC1	EC0	EC1	Time	ρ_1	ρ_2
True θ	0.5	1.0	-1.0	0.5	1.0	1.0	1.0			
X=2560, $\gamma_a = 0$										
AFD	0.501	1.009	-1.013	0.488	1.019	1.017	1.004	0.45	6.83E-17	9.50E-17
	(0.153)	(0.085)	(0.086)	(0.212)	(0.081)	(0.230)	(0.085)	0.40		
AFD2	0.500	1.001	-1.005	0.485	1.016	1.022	1.003	1.03	1.39E-16	3.17E-16
	(0.065)	(0.044)	(0.043)	(0.154)	(0.071)	(0.223)	(0.084)	0.99		
AM	0.495	1.014	-1.017	0.451	0.997	1.029	0.993	7111.56	1.55E-16	2.55E-16
	(0.152)	(0.085)	(0.086)	(0.214)	(0.078)	(0.232)	(0.085)	7111.52		
X=2560, $\gamma_a = 1$										
AFD	0.510	1.010	-1.012	0.266	0.943	0.978	0.991	0.42	2.49E-01	2.52E-01
	(0.071)	(0.049)	(0.051)	(0.277)	(0.085)	(0.239)	(0.086)	0.38		
AFD2	0.526	1.007	-1.015	0.523	1.012	0.984	1.003	0.97	5.57E-10	1.50E-09
	(0.178)	(0.117)	(0.116)	(0.310)	(0.126)	(0.254)	(0.111)	0.93		
AM	0.443	1.080	-1.087	0.059	0.955	0.984	1.029	7511.62	4.94E-01	5.92E-01
	(0.174)	(0.126)	(0.128)	(0.500)	(0.098)	(0.245)	(0.105)	7511.57		

Key Contributions

- We develop a closed-form solution for weight characterization and formulate a new estimator based on it.
- We derive a closed-form mathematical solution for the proposed weight characterization.
- We propose a new estimator that leverages this weight characterization that can be used to estimate non-stationary models.
- Demonstrated the estimator's performance via Monte Carlo simulation.

References I

Aguirregabiria, V. and Magesan, A. (2016). Solution and Estimation of Dynamic Discrete Choice Structural Models Using Euler Equations. *SSRN Electronic Journal*.

Arcidiacono, P. and Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, 79(6):1823–1867.

Arcidiacono, P. and Miller, R. A. (2019). Nonstationary dynamic models with finite dependence. *Quantitative Economics*, 10(3):853–890.