Discrete choice with extreme-value taste shocks

Bence Bardóczy[a]

NBER HA Macro Workshop, Spring 2023

[a]Federal Reserve Board: The views expressed are my own and do not necessarily reflect those of the Board of Governors or the Federal Reserve System.

Today

- **So far**: agents are hit by **discrete shocks** and make **continuous choices**.
    - savings, consumption, hours…
- Many interesting economic decisions are **discrete**.
    - labor force participation, occupation choice
    - lumpy adjustment with fixed costs (price, investment, portfolio…)
- **This class**: one approach to discrete choice that's fairly general and fits naturally into the SSJ framework.
    - focus on method, not economic content

Why is discrete choice hard?

- Only discrete choice would be easy.
  - value function iteration works well

- Interaction between discrete and continuous choices is the hard part.

- **Non-convexity**: FOCs are insufficient to obtain policy functions.
  - EGM does not work, more robust backward iteration is needed
  - solution: **EGM + upper envelope** (Fella, 2014; Druedahl, 2020)

- **Discontinuities** in policy functions.
  - fake news algorithm relies on differentiating policies wrt aggregate inputs
  - solution: **logit smoothing** (Iskhakov et al., 2017)

  $\rightarrow$ common in microeconometrics, also useful for Jacobian computation!

Roadmap

1. SIM model with labor force participation
2. Solving the SIM model with participation
3. General HA framework with stages
4. Jacobians for discrete choices

SIM with labor force participation

SIM model with labor force participation

- Work full time or not at all. Disutility of full-time work is

$$V_t(z_{it}, a_{it-1}) = \max_{c_{it}, n_{it}, a_{it}} u(c_{it}) - \varphi n_{it} + \beta \mathbb{E}_t V_{t+1}(z_{it+1}, a_{it})$$

$$\text{s.t. } c_{it} + a_{it} = (1 + r_t) a_{it-1} + w_t n_{it} z_{it} + T_t$$

$$n_{it} \in \{0, 1\}$$

$$a_{it} \geq \underline{a}$$

- Nests SIM for $\varphi = 0$.

- Solving the model means characterizing
  1. policy functions: $a_t(z, a_-), c_t(z, a_-), n_t(z, a_-)$
  2. distribution: $D_t(z, a_-)$
  3. aggregate outputs: $A = \int a_t dD_t, \quad C_t = \int c_t dD_t, \quad N_t = \int n_t dD_t$

Peek at solution

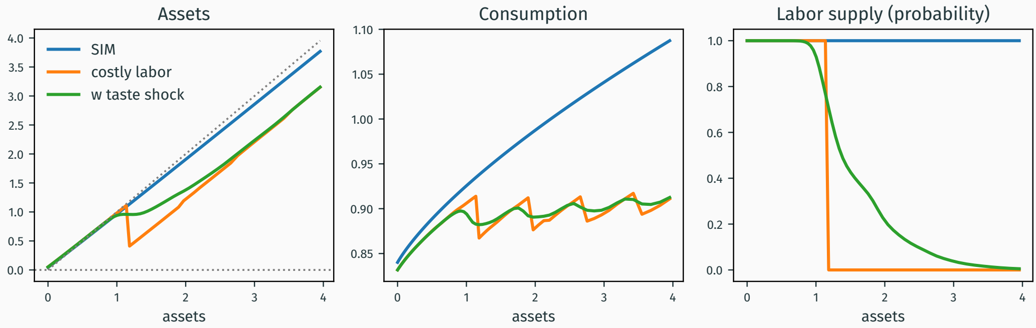


Figure 1: Policy functions conditional on average productivity

Economics of the model

- Rich and unproductive households choose not to work.
- Non-participant households run down assets aggressively to finance consumption.
- Consumption and asset policies are **non-monotonic** in assets and—absent of taste shocks—have **discontinuities**.
  - primary discontinuity from change in participation
  - secondary discontinuities from consumption smoothing in discrete time
  - **intuition:** $a = 1.9$ expects to hit participation threshold in 1 period, while $a = 2$ expects it in 2 periods $\longrightarrow a = 2$ consumes less and saves more today

Solving the SIM model with participation

Stages

- We break up the decision problem into several **stages**.
- Think of each stage as updating a single state variable.
- Stages are a useful concept to describe models with complex timing. They're also the key abstraction behind implementation of discrete choice in SSJ.
  - `StageBlock` in tutorial

Break up problem into stages

- **Stage 0:** enter period t. (z_{it-1}, a_{it-1})
- **Stage 1:** productivity shock $z_{it-1} \to z_{it}$
- **Stage 2:** participation choice $\{\} \to n_{it}$
- **Stage 3:** consumption-saving choice $a_{it-1} \to a_{it}, c_{it}$

Stage 3: continuous choice

- **Stage 3** looks like vanilla SIM model with extra state n

$$V^{(3)}(n, z, a_-) = \max_{c, a \geq \underline{a}} u(c) - \varphi n + \beta \mathbb{E} V^{(1)}(z', a) \quad \text{s.t. } c + a = (1 + r)a_- + wnz + T \tag{1}$$

- Characterizes **discrete choice-specific** policies $a(n, z, a_-), c(n, z, a_-)$.
- Can we get these via endogenous gridpoint method?

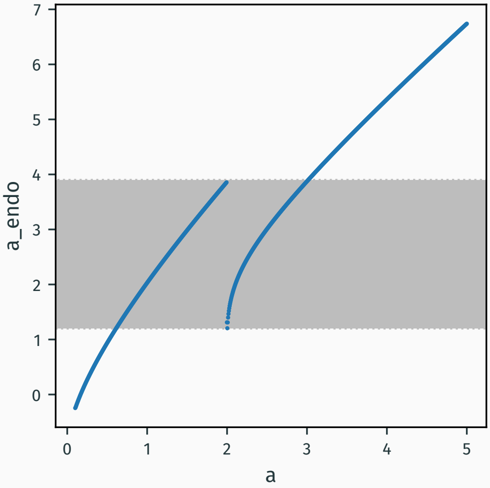
Stage 3: continuous choice

- **Stage 3** looks like vanilla SIM model with extra state n

$$
\begin{aligned}
V^{(3)}(n, z, a_-) = \max_{c, a \geq \underline{a}} & \, u(c) - \varphi n + \beta \mathbb{E} V^{(1)}(z', a) \\
\text{s.t. } c + a &= (1 + r)a_- + wnz + T
\end{aligned}
\tag{1}
$$

- Characterizes **discrete choice-specific** policies $a(n, z, a_-), c(n, z, a_-)$.
- Can we get these via endogenous gridpoint method?
- **No!** Catch is that $V^{(1)}$ is not concave, so $V_a^{(1)}$ is not monotonic.

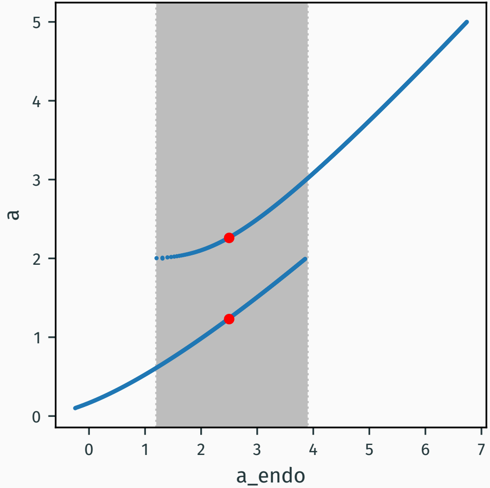
EGM + upper envelope



- Try EGM with non-monotonic $V_a^{(1)}$.
  - $c_{endo}^{-\sigma} = \beta V_a^{(1)}$ (Euler)
  - $c_{endo} + a_{grid} = (1 + r)a_{endo} + y$ (budget)

- $a_{endo}(a)$ may be non-monotonic as well.

EGM + upper envelope



- Try EGM with non-monotonic $V_a^{(1)}$.
  - $c_{endo}^{-\sigma} = \beta V_a^{(1)}$ (Euler)
  - $c_{endo} + a_{grid} = (1+r)a_{endo} + y$ (budget)
- $a_{endo}(a)$ may be non-monotonic as well.
- Can't invert $a_{endo}(a)$ to get policy function. Both ● solve FOCs.
- **Upper envelope**: compute $V^{(3)}$ at both solutions and choose max.
  - implementation details in tutorial

Stage 2: discrete choice

- **Stage 2** is a pure discrete choice problem

$$V^{(2)}(z, a_-) = \max_{n \in \{0,1\}} V^{(3)}(n, z, a_-) + \underbrace{\varepsilon(n)}_{\text{taste shock}} \tag{2}$$

Stage 2: discrete choice

- **Stage 2** is a pure discrete choice problem

$$V^{(2)}(z, a_-) = \max_{n \in \{0,1\}} V^{(3)}(n, z, a_-) + \underbrace{\varepsilon(n)}_{\text{taste shock}} \tag{2}$$

- **Analytical solution** if taste shock is iid EV-1 with scale σ.
  - logit choice probability:

$$P(n|z, a_-) = \exp\left(\frac{V^{(3)}(n, z, a_-)}{\sigma}\right) \Big/ \sum_{n' \in \{0,1\}} \exp\left(\frac{V^{(3)}(n', z, a_-)}{\sigma}\right) \tag{3}$$

  - logsum formula:

$$V^{(2)}(z, a_-) = \sigma \log\left(\sum_{n' \in \{0,1\}} \exp\left(\frac{V^{(3)}(n', z, a_-)}{\sigma}\right)\right) \tag{4}$$

Stage 1: discrete shock

- **Stage 1**: productivity shock follows exogenous Markov process

$$V^{(1)}(z_-, a_-) = \sum_z \Pr(z|z_-) \cdot V^{(2)}(z, a_-) \tag{5}$$

Stage 1: discrete shock

- **Stage 1**: productivity shock follows exogenous Markov process

$$V^{(1)}(z_-, a_-) = \sum_z \Pr(z|z_-) \cdot V^{(2)}(z, a_-) \tag{5}$$

- The circle is complete. Start from an initial guess $V_T^{(1)}$ and iterate backward **stage-by-stage** until convergence

$$V_{t+1}^{(1)} \to V_t^{(3)} \to V_t^{(2)} \to V_t^{(1)} \tag{6}$$

General framework

Warmup

- Consider a 2-state **Markov process** of employment & unemployment.
  - flow utility $\mathbf{u}_t = [u_t^E, u_t^U]$, value function $\mathbf{v}_t = [v_t^E, v_t^U]$
  - distribution is $\mathbf{D}_t = [D_t^E, D_t^U]$
  - transition probabilities are f_t and s_t

Warmup

- Consider a 2-state **Markov process** of employment & unemployment.
  - flow utility $\mathbf{u}_t = [u_t^E, u_t^U]$, value function $\mathbf{v}_t = [v_t^E, v_t^U]$
  - distribution is $\mathbf{D}_t = [D_t^E, D_t^U]$
  - transition probabilities are f_t and s_t

- **Backward iteration** for $\mathbf{v}_t$?

$$\begin{bmatrix} v_t^E \\ v_t^U \end{bmatrix} = \begin{bmatrix} u_t^E \\ u_t^U \end{bmatrix} + \beta \underbrace{\begin{bmatrix} 1-s_t & s_t \\ f_t & 1-f_t \end{bmatrix}}_{\Lambda} \begin{bmatrix} v_{t+1}^E \\ v_{t+1}^U \end{bmatrix} \qquad \text{(Bellman equation)}$$

- **Forward iteration** for $\mathbf{D}_{t+1}$?

Warmup

- Consider a 2-state **Markov process** of employment & unemployment.
  - flow utility $\mathbf{u}_t = [u_t^E, u_t^U]$, value function $\mathbf{v}_t = [v_t^E, v_t^U]$
  - distribution is $\mathbf{D}_t = [D_t^E, D_t^U]$
  - transition probabilities are f_t and s_t

- **Backward iteration** for $\mathbf{v}_t$?

$$\begin{bmatrix} v_t^E \\ v_t^U \end{bmatrix} = \begin{bmatrix} u_t^E \\ u_t^U \end{bmatrix} + \beta \underbrace{\begin{bmatrix} 1 - s_t & s_t \\ f_t & 1 - f_t \end{bmatrix}}_{\Lambda} \begin{bmatrix} v_{t+1}^E \\ v_{t+1}^U \end{bmatrix} \qquad \text{(Bellman equation)}$$

- **Forward iteration** for $\mathbf{D}_{t+1}$?

$$\begin{bmatrix} D_{t+1}^E \\ D_{t+1}^U \end{bmatrix} = \underbrace{\begin{bmatrix} 1 - s_t & f_t \\ s_t & 1 - f_t \end{bmatrix}}_{\Lambda'} \begin{bmatrix} D_t^E \\ D_t^U \end{bmatrix} \qquad \text{(law of motion)}$$

General HA problem without stages

- Consider HA problem with aggregate inputs $\mathbf{X}_t$.

$$\mathbf{v}_t = v(\mathbf{v}_{t+1}, \mathbf{X}_t) \qquad \text{(Bellman equation)}$$

$$\mathbf{D}_{t+1} = D(\mathbf{v}_{t+1}, \mathbf{D}_t, \mathbf{X}_t) \qquad \text{(law of motion)}$$

General HA problem without stages

- Consider HA problem with aggregate inputs $\mathbf{X}_t$.

$$\mathbf{v}_t = v(\mathbf{v}_{t+1}, \mathbf{X}_t) \qquad \text{(Bellman equation)}$$
$$\mathbf{D}_{t+1} = D(\mathbf{v}_{t+1}, \mathbf{D}_t, \mathbf{X}_t) \qquad \text{(law of motion)}$$

- This is a Markov process, just more complex.

- Suppose the state space is discretized on N gridpoints.
  - flow utility, value function, distribution are vectors: $\mathbf{u}, \mathbf{v}, \mathbf{D} \in \mathbb{R}^N$
  - Markov matrix of joint state: $\Lambda \in \mathbb{R}^{N \times N}$

$$\mathbf{v}_t = \mathbf{u}_t + \beta \Lambda(\mathbf{v}_{t+1}, \mathbf{X}_t)\mathbf{v}_{t+1} \qquad \text{(Bellman equation)}$$
$$\mathbf{D}_{t+1} = \Lambda(\mathbf{v}_{t+1}, \mathbf{X}_t)'\mathbf{D}_t \qquad \text{(law of motion)}$$

Solving HA models in discrete vs continuous time

- Achdou et al. (2021) claim 4 advantages for continuous time.
    1. FOCs are sufficient almost everywhere[1] constraints easier, no secondary kinks
    2. FOCs are static no costly root finding
    3. HJB and KFE are adjoint operators "solve policies get distribution for free"
    4. sparsity Markov matrix of joint states is block tridiagonal

- What do you say?

[1]except at boundaries and primary kinks

Solving HA models in discrete vs continuous time

- Achdou et al. (2021) claim 4 advantages for continuous time.
  1. FOCs are sufficient almost everywhere[1] — constraints easier, no secondary kinks
  2. FOCs are static — no costly root finding
  3. HJB and KFE are adjoint operators — "solve policies get distribution for free"
  4. sparsity — Markov matrix of joint states is block tridiagonal

- What do you say?
  1. helps in models with ≥ 2 endogenous states or discrete choices
  2. EGM avoids root finding but requires interpolation
  3. general property of Markov processes
  4. **claim:** we can exploit **sparsity** better in discrete time

[1] except at boundaries and primary kinks

General HA problem with stages

- Let there be $j = 1, \ldots, J$ stages.
  - discrete shock, discrete choice, continuous choice (1-2 states)

- Same logic of backward and forward iteration applies between stages.

$$\mathbf{v}_{t,j} = \Lambda_j(\mathbf{v}_{t,j+1}, \mathbf{X}_{t,j})\mathbf{v}_{t,j+1} \quad \text{(Bellman equation)}$$

$$\mathbf{D}_{t,j+1} = \Lambda_j(\mathbf{v}_{t,j+1}, \mathbf{X}_{t,j})'\mathbf{D}_{t,j} \quad \text{(law of motion)}$$

General HA problem with stages

- Let there be $j = 1, \ldots, J$ stages.
    - discrete shock, discrete choice, continuous choice (1-2 states)

- Same logic of backward and forward iteration applies between stages.

$$\mathbf{v}_t = (\Lambda_1 \cdot \Lambda_2 \cdots \Lambda_J)\, \mathbf{v}_{t+1} \qquad \text{(Bellman equation)}$$

$$\mathbf{D}_{t+1} = \left(\Lambda_J' \cdot \Lambda_{J-1}' \cdots \Lambda_1'\right) \mathbf{D}_t \qquad \text{(law of motion)}$$

- **Key insight:** Stage-specific Markov matrices are sparser than their product.

- Optimize sparse “matrix multiplication” operation for each type of stage.

- Today = tomorrow property of ctime has costs as well as benefits.
    - static FOCs that are sufficient almost everywhere, but can’t divide problem into stages

Taking stock

- **Stage** is a useful abstraction for both intuition & computation.
- If you can write backward iteration for a stage…
    - chain arbitrary many stages together to elegantly represent complex models
    - forward iteration is just the transpose operation
- You can solve cutting edge models in discrete as well as in continuous time.
- Next: last piece of sequence-space **Jacobian** machinery.

Jacobians with discrete choice

Brief overview

- **Fake news algorithm** applies directly to discrete shock and cont choice stages.
- What about discrete choice stage?
- Main reason for working with **EV-1 taste shocks**: Choice probability and value function are **smooth** with **closed-form derivatives**.

Discrete choice derivatives

- Simplified notation: V_i' is vfun in next stage conditional on discrete choice i.
- Recall logsum and logit formulae:

$$V = \sigma \log \left(\sum_i \exp\left(V_i'/\sigma\right) \right) \qquad \text{and} \qquad P_i = \frac{\exp\left(V_i'/\sigma\right)}{\sum_k \exp\left(V_k'/\sigma\right)}$$

Discrete choice derivatives

- Simplified notation: V_i' is vfun in next stage conditional on discrete choice i.
- Recall logsum and logit formulae:

$$V = \sigma \log \left(\sum_i \exp\left(V_i'/\sigma\right) \right) \qquad \text{and} \qquad P_i = \frac{\exp\left(V_i'/\sigma\right)}{\sum_k \exp\left(V_k'/\sigma\right)}$$

- Few lines of algebra yields

$$dV = \sum_i P_i dV_i' \qquad \text{and} \qquad dP_i = \frac{P_i(dV_i' - dV)}{\sigma}$$

- **Takeaway:** propagating small shocks backward $\approx$ expectations with ss probabilities

Conclusion

Conclusion

- Discrete choice itself is easy (choose best of a few alternatives) but causes **non-convexity** that complicates continuous choices.
  - FOCs necessary but not sufficient even in interior
  - “secondary kinks” arise in discrete time
- **EGM + upper envelope**: choose best of few alternatives that satisfy FOCs.
  - have to keep track of vfun & partial vfun
- **EV-1 taste shocks** facilitate differentiation at almost no cost.
- Intuitive concepts that improve computation: **DAG**, **stage**.
  - stage only makes sense in discrete time

Thank you!

References

Achdou, Yves, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach," *The Review of Economic Studies*, 04 2021, *89* (1), 45–86.

Druedahl, Jeppe, "A Guide On Solving Non-Convex Consumption-Saving Models," *Computational Economics*, 2020, pp. 1–29.

Fella, Giulio, "A Generalized Endogenous Grid Method for Non-Smooth and Non-Concave Problems," *Review of Economic Dynamics*, 2014, *17* (2), 329–344.

Iskhakov, Fedor, Thomas H Jørgensen, John Rust, and Bertel Schjerning, "The Endogenous Grid Method for Discrete-Continuous Dynamic Choice Models with (or without) Taste Shocks," *Quantitative Economics*, 2017, *8* (2), 317–365.