

ASSIGNMENT 3 (Arcidiacono and Miller, 2011)

Overview. This assignment concerns CCP-based estimation of single-agent dynamic discrete choice models with *unobserved heterogeneity*, based on Arcidiacono and Miller (2011, *Econometrica*). You will: (i) understand why standard two-step CCP methods fail when agent types are unobserved, (ii) implement the EM algorithm combined with CCP estimation, (iii) compare different CCP updating schemes (data vs. model updating), and (iv) investigate the role of initial conditions in identifying type distributions. The empirical setting is the bus engine replacement model of Rust (1987).

Instructions. Work in groups of about three. Each group should submit a single, self-contained report with well-documented code as an appendix. Hand written work will not be graded. Use the replication code provided in the `tutorial3/` folder. Questions will ask you to explain algorithms, modify parameters, and compare results. Report the specification of your computer (CPU, cores, memory) and computation times. All questions carry equal weight unless otherwise stated.

Model. Consider the bus engine replacement problem of Rust (1987), extended to include unobserved heterogeneity following Arcidiacono and Miller (2011). Time is discrete with periods $t = 1, 2, \dots$, and the decision-maker (Harold Zurcher) discounts the future at rate $\beta \in (0, 1)$.

State variables. The state variables are (x_t, z_i, s_i) where:

- $x_t \in [0, \bar{x}]$: Accumulated mileage (discretized into bins).
- z_i : Bus-specific mileage increment parameter (time-invariant, observed).
- $s_i \in \{0, 1\}$: Unobserved bus type (time-invariant, **not observed** by econometrician).

Choice variables. Each period, the decision-maker chooses $d_t \in \{0, 1\}$ where:

- $d_t = 0$: Replace the engine (mileage resets to zero).
- $d_t = 1$: Keep the engine (mileage continues to accumulate).

Flow payoffs. The systematic component of current utility is:

$$\begin{aligned} u_0(x_t, s_i) &= 0 \quad (\text{replace, normalized}) \\ u_1(x_t, s_i) &= \alpha_1 + \alpha_2 x_t + \alpha_3 s_i \quad (\text{keep}) \end{aligned}$$

where α_1 is the base utility from keeping, $\alpha_2 < 0$ captures the increasing maintenance cost with mileage, and α_3 captures type-specific differences in replacement threshold. Each

period the decision-maker also receives a private shock ε_{jt} that is i.i.d. Type I extreme value across choices and time.

Transition. Mileage increments follow an exponential distribution with bus-specific rate parameter z_i :

$$\Delta x_t \sim \text{Exp}(z_i)$$

After replacement ($d_t = 0$), mileage resets: $x_{t+1} = \Delta x_t$. After keeping ($d_t = 1$), mileage accumulates: $x_{t+1} = x_t + \Delta x_t$.

Unobserved heterogeneity. The type s_i is drawn at the beginning of time from:

$$P(s_i = 1) = 1 - \pi, \quad P(s_i = 0) = \pi$$

The econometrician observes (x_{it}, z_i, d_{it}) for each bus i and period t , but does **not** observe s_i .

Parameters. Use the following baseline values:

Parameter	Symbol	Value
Base utility (keep)	α_1	2.0
Mileage coefficient	α_2	-0.15
Type effect	α_3	1.0
Discount factor	β	0.9
Type 0 probability	π	0.4

Question 1. Define the dynamic optimization problem and explain how finite dependence arises.

- Write down the Bellman equation for a bus of type s . Define the conditional value functions $v_0(x, s)$ and $v_1(x, s)$, and the ex-ante value function $V(x, s)$.
- Explain how the Type I extreme value assumption yields the logit choice probability and the “log-sum” formula for the ex-ante value function.
- This model exhibits one-period finite dependence. Derive the expression for $v_1(x, s) - v_0(x, s)$ that shows the future value terms cancel. Why does finite dependence arise from the replacement (terminal) action?
- How does finite dependence simplify estimation? Compare to models without this property (e.g., where both actions lead to continuation with different transition dynamics).

Question 2. Implement the model solution and simulation.

- Describe the algorithm for computing optimal CCPs via fixed-point iteration. What is the contraction mapping? What determines convergence?

- (b) Using the baseline parameters, solve for the optimal CCPs separately for type $s = 0$ and type $s = 1$. Plot the keep probability as a function of mileage for both types, for several values of z . Interpret the economic patterns.
- (c) Simulate panel data with $N = 1000$ buses and $T = 30$ periods (after a burn-in of 10 periods). Report summary statistics: replacement rates by type, mileage distributions, and the empirical relationship between mileage and replacement.
- (d) Modify the type effect α_3 (e.g., try $\alpha_3 = 0.5$ and $\alpha_3 = 2.0$). How do equilibrium CCPs and simulated outcomes change? Discuss identification: what variation in the data identifies α_3 ?

Question 3. Estimate the model assuming types are **observed**.

- (a) **NFXP (Nested Fixed Point).** Describe the algorithm: what is the inner loop, what is the outer loop? Implement NFXP and report parameter estimates, standard errors, and computation time.
- (b) **Two-Step CCP.** Describe the Hotz-Miller (1993) approach: what is estimated in the first stage? How does the Hotz-Miller inversion recover future values from CCPs? Implement the two-step estimator and report results.
- (c) Compare NFXP and Two-Step: parameter estimates, standard errors, and computation time. When types are observed, which method is preferred and why?

Question 4. Now suppose types are **unobserved**. Explain why standard methods fail.

- (a) If we estimate CCPs by pooling all buses (ignoring types), what do we actually estimate? Write the relationship between the pooled CCP and the type-specific CCPs.
- (b) Explain the problem: to get type-specific CCPs, we need posterior type probabilities; but to compute posteriors, we need parameters (and hence CCPs).
- (c) How does the EM algorithm resolve this problem? Describe the E-step (computing posterior type probabilities) and the M-step (updating parameters given posteriors).

Question 5. Implement EM-based estimation with unobserved heterogeneity using two different CCP updating approaches.

- (a) **CCP-EM with Data Updating.** In the M-step, update CCPs via weighted reduced-form logit:

$$\hat{p}^{k+1}(x, s) = \frac{\sum_{i,t} \tau_{is}^k \mathbf{1}\{x_{it} = x, d_{it} = 1\}}{\sum_{i,t} \tau_{is}^k \mathbf{1}\{x_{it} = x\}}$$

where τ_{is}^k is the posterior probability that bus i is type s . Implement this method and report parameter estimates, type classification accuracy, and computation time.

- (b) **CCP-EM with Model Updating.** In the M-step, update CCPs using the model's best-response operator:

$$p^{k+1} = \Psi(p^k; \theta^{k+1})$$

This ensures CCPs are always model-consistent. Implement this method and report results.

- (c) Compare the two CCP-EM methods. Which converges faster? Which yields more accurate parameter estimates? When might each be preferred?
- (d) Also compare to NFXP+EM (which solves the full DP in the M-step). Report computation times and discuss the trade-off between computational cost and statistical efficiency.

Question 6. Investigate the role of initial conditions.

- (a) Plot the distribution of initial mileage x_{i0} separately for type 0 and type 1 buses in your simulated data. Are they different? Explain why initial conditions depend on unobserved type.
- (b) If we ignore the initial conditions problem (i.e., treat $P(s_i = 1)$ as the same for all buses regardless of x_{i0}), what bias might arise in estimating the type distribution π ?
- (c) Arcidiacono and Miller (2011) propose modeling initial conditions as:

$$P(s_i = 1 | x_{i0}, z_i) = \Lambda(\gamma_0 + \gamma_1 x_{i0} + \gamma_2 z_i)$$

Explain how this addresses the initial conditions problem. Modify your estimator to incorporate this correction and compare results with and without the correction.

Question 7. Conduct a Monte Carlo study to assess finite-sample performance.

- (a) **Design.** Conduct at least 100 Monte Carlo replications. For each replication: (i) simulate a new dataset, (ii) estimate using Two-Step (with observed types), CCP-EM (Data), and CCP-EM (Model), and (iii) record point estimates for $(\alpha_1, \alpha_2, \alpha_3, \pi)$ and computation time.
- (b) **Visualize.** For each parameter, create a figure showing the distribution of estimates across replications. Include vertical lines at the true values. Compare the three methods in the same figure.
- (c) **Summary statistics.** Report a table with mean estimate, bias, standard deviation, and RMSE for each parameter and method. Include mean computation time.
- (d) **Discussion.** Which parameters are hardest to estimate? Does the type effect α_3 exhibit more bias than the mileage coefficient α_2 ? Explain intuitively.

Deliverables. Submit a report that:

- (i) Answers each question with appropriate explanations and derivations.
- (ii) Includes all tables, figures, and summary statistics requested.
- (iii) Reports and interprets estimation results across methods (Two-Step, CCP-EM Data, CCP-EM Model, NFXP+EM).
- (iv) Discusses the practical implications of unobserved heterogeneity for empirical work.

Attach your code (well commented) as an appendix, highlighting any modifications you made to the provided code. Report computation times and computer specifications.

References.

- Aguirregabiria, V. and Mira, P. (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica*, 75(1):1–53.
- Arcidiacono, P. and Miller, R.A. (2011). Conditional Choice Probability Estimation of Dynamic Discrete Choice Models With Unobserved Heterogeneity. *Econometrica*, 79(6):1823–1867.
- Hotz, V.J. and Miller, R.A. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies*, 60(3):497–529.
- Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, 55(5):999–1033.