Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher

Summarized by Wenzhi Wang

M.Phil. Student at the University of Oxford

March 15, 2023

Table of Contents

Notations

$C(x_t)$: Choice set; a finite set of allowable values of the control variable i_t when the state variable is x_t.
$\varepsilon_t = \{\varepsilon_t(i) \mid i \in C(x_t)\}$: A $\dim(C(x_t))$-dimensional vector of state variables observed by agent but not by the econometrician.
$x_t = \{x_t(1), \ldots, x_t(K)\}$: K-dimensional vector of state variables observed by both the agent and the econometrician.

Notations

$u(x_t, i, \theta_1) + \varepsilon_t(i)$: Realized single-period utility of decision i when state variable is (x_t, ε_t).
$p(x_{t+1}, \varepsilon_{t+1} \mid x_t, \varepsilon_t, i_t, \theta_2, \theta_3)$: Markov transition density for state variable (x_t, ε_t) when alternative i_t is selected.
$\theta = (\beta, \theta_1, \theta_2, \theta_3)$: The complete $(1 + K_1 + K_2)$ vector of parameters to be estimated.

Sequential Problem

Given the stochastic evolution of the state variables (x_t, ε_t) embodied by the transition probability p, the agent must choose a sequence of decision rules or controls $f_t(x_t, \varepsilon_t, \theta)$ to maximize

$$V_\theta(x_t, \varepsilon_t) = \sup_{\Pi} E\left\{\sum_{j=t}^{\infty} \beta^{j-t}\left[u\left(x_j, f_j, \theta_1\right) + \varepsilon_j\left(f_j\right)\right] \mid x_t, \varepsilon_t, \theta_2, \theta_3\right\} \tag{1}$$

where $\Pi = \{f_t, f_{t+1}, \ldots\}$, $f_t \in C(x_t)$ for all t.

Sequential Problem

Expectation is taken with respect to the controlled stochastic process (x_t, ε_t) whose probability density is defined from Π and the transition probability p by

$$\begin{aligned} & p\left\{x_{t+1}, \varepsilon_{t+1}, \ldots, x_{t+N}, \varepsilon_{t+N} \mid x_t, \varepsilon_t\right\} \\ & \quad = \prod_{i=t}^{N-1} p\left(x_{i+1}, \varepsilon_{i+1} \mid x_i, \varepsilon_i, f_i\left(x_i, \varepsilon_i\right), \theta_2, \theta_3\right) \end{aligned} \tag{2}$$

Sequential Problem

This is an infinite-horizon, discounted Markovian decision problem. Under certain regularity assumptions, the solution is given by a stationary decision rule

$$i_t = f(x_t, \varepsilon_t, \theta) \tag{3}$$

Functional Problem

The optimal value function V_θ is the unique solution to Bellman's equation given by

$$V_\theta(x_t,\varepsilon_t) = \max_{i \in C(x_t)} \left[u(x_t,i,\theta_1) + \varepsilon_t(i) + \beta EV_\theta(x_t,\varepsilon_t,i)\right] \tag{4}$$

where

$$EV_\theta(x_t,\varepsilon_t,i) \equiv \int_y \int_\eta V_\theta(y,\eta) p\left(dy, d\eta \mid x_t,\varepsilon_t,i,\theta_2,\theta_3\right) \tag{5}$$

and the optimal control f is defined by

$$f(x_t,\varepsilon_t,\theta) \equiv \operatorname*{argmax}_{i \in C(x_t)} \left[u(x_t,i,\theta_1) + \varepsilon_t(i) + \beta EV_\theta(x_t,\varepsilon_t,i)\right] \tag{6}$$

Difficulties Facing the Econometricians

- Unbounded distribution ε_t can raise serious dimensionality problems since the optimal stationary policy f will ordinarily be computed by solving for the fixed point V_θ from Bellman's equation.
- To obtain the choice probabilities $P(i_t \mid x_t, \theta)$, we need to integrate $V_\theta(x_t, \varepsilon_t)$ over the unknown random vectors ε_t.
- Solution: Introduce assumptions to circumvent certain difficulties.

CI Assumption

Conditional Independence Assumption (CI):
The transition density of the controlled process x_t, ε_t factors as

$$p(x_{t+1}, \varepsilon_{t+1} \mid x_t, \varepsilon_t, i_t, \theta_2, \theta_2) = q(\varepsilon_{t+1} \mid x_{t+1}, \theta_2) p(x_{t+1} \mid x_t, i, \theta_3). \tag{7}$$

Restrictions Imposed by the CIA

- x_{t+1} is a sufficient statistic for ε_{t+1}, implying that any statistical dependence between ε_t and ε_{t+1} is transmitted entirely through the vector x_{t+1}.
- The probability density of x_{t+1} depends only on x_t and not ε_t. Intuitively, the ε_t process can e regarded as noise superimposed on the underlying x_t process.

Payoffs of CIA

- CIA implies that EV_θ is not a function of ε_t, so that required choice probabilities will not require integration over the unknown function EV_θ.
- CIA implies EV_θ is a fixed point of a separate contraction mapping on the reduced state space $\Gamma = \{(x, \varepsilon) \mid x \in \mathbb{R}^K, \varepsilon \in \mathbb{R}^{\dim C(x)}\}$ and avoiding the numerical integration required to obtain EV_θ from V_θ.

Theorem 1

Assume that CIA holds. Let $P(i \mid x, \theta)$ denote the conditional probability of choosing action $i \in C(x)$ given state variable x. Denote the social surplus function corresponding to the density $q(\varepsilon \mid x, \theta_2)$ by $G([u(x,\theta_1) + \beta EV_\theta(x)] \mid x, \theta_2)$, which is defined as

$$\begin{aligned} & G\left(\left[u\left(x,\theta_1\right) + \beta EV_\theta(x)\right] \mid x, \theta_2\right) \\ & \quad \equiv \int_\varepsilon \max_{j \in C(x)} \left[u\left(x,j,\theta_1\right) + \beta EV_\theta(x,j)\right] q\left(d\varepsilon \mid x, \theta_2\right). \end{aligned} \tag{8}$$

Theorem 1

Then $P(i \mid x, \theta)$ is given by

$$P(i \mid x, \theta) = G_i\left(\left[u(x, \theta_1) + \beta EV_\theta(x)\right] \mid x, \theta_2\right) \tag{9}$$

where G_i denotes the partial derivative of G with respect to $u(x, i, \theta)$ and the function EV_θ is the unique fixed point to a contraction mapping T_θ, $T_\theta(EV_\theta) = EV_\theta$, defined for each $(x, i) \in \Gamma$ by

$$EV_\theta(x, i) = \int_y G\left(\left[u(y, \theta_1) + \beta EV_\theta(y)\right] \mid y, \theta_2\right) p(dy \mid x, i, \theta_3). \tag{10}$$

Relationship to the Static Case

- The static model of discrete choice appears as a special case of Theorem 1 when $p(\cdot \mid x, i, \theta_3)$ is independent of i. In that case, the expected utilities $EV_\theta(x, i)$ are also independent of i which implies that G is a function of $\{u(x, j, \theta) \mid j \in C(x)\}$ alone.
- When current choices do have future consequences, the term $\beta EV_\theta(x, i)$, provides the appropriate "shadow price" for the future consequences of each action and must be added to the current utility in order to correctly describe the optimal behaviour of the agent.

Theorem 2

- Given time series observations $\{(i_1, x_1), (i_2, x_2), \ldots, (i_T, x_T)\}$for a single individual we form the likelihood function $\ell^f((i_1, x_1, , \ldots, i_T, x_T \mid i_0, x_0, \theta)$ and estimate the unknown parameters θ by the method of maximum likelihood.
- Theorem 2: Under Assumption CI, the likelihood function ℓ^f is given by

$$\ell^f(i_1, x_1, , \ldots, i_T, x_T \mid i_0, x_0, \theta) = \Pi_{t=1}^{T} P(i_t \mid x_t, \theta) p(x_t \mid x_{t-1}, i_{t-1}, \theta_3), \tag{11}$$

where $P(i_t \mid x_t, \theta)$ is given by 9.

Special Case: Extreme Value Distribution

If $q(\varepsilon \mid y, \theta_2)$ is given by a multivariate extreme value distribution,

$$q(\varepsilon \mid x, \theta_2) = \prod_{j \in C(x)} \exp\{-\varepsilon(j) + \theta_2\} \exp\{-\exp\{-\varepsilon(j) + \theta_2\}\} \tag{12}$$

then the social surplus function G is given by

$$G\left(\left[u(x,\theta_1) + \beta EV_\theta(x)\right] \mid x, \theta_2\right) = \ln\left\{\sum_{j \in C(x)} \exp\left[u(x,j,\theta_1) + \beta EV_\theta(x,j)\right]\right\}, \tag{13}$$

$P(i \mid x, \theta)$ is given by the well-known multinomial logit formula

$$P(i \mid x, \theta) = \frac{\exp\left\{u(x,i,\theta_1) + \beta EV_\theta(x,i)\right\}}{\sum_{j \in C(x)} \exp\left\{u(x,j,\theta_1) + \beta EV_\theta(x,j)\right\}} \tag{14}$$

and EV_θ is given by the unique solution to the functional equation

$$EV_\theta(x,i) = \int_y \log\left\{\sum_{j \in C(y)} \exp\left[u(y,j,\theta_1) + \beta EV_\theta(y,j)\right]\right\} p(dy \mid x, i, \theta_3). \tag{15}$$

Model Setup

- Let the state variable x_t, denote the accumulated mileage (since last replacement) on the bus engine at time t and suppose that expected per period operating costs are given by an increasing, differentiable function of x_t, $c(x_t, \theta_1)$.
- Let RC denote the expected cost of a replacement bus engine. The company's implied utility function is

$$u(x_t, i, \theta_1) + \varepsilon_t(i) = \begin{cases} -RC - c(0, \theta_1) + \varepsilon_t(1) & \text{if } i = 1, \\ -c(x_t, \theta_1) + \varepsilon_t(0) & \text{if } i = 0. \end{cases} \tag{16}$$

- The functional form for c which I estimated include (1) polynomial: $c(x, \theta_1) = \theta_{11}x + \theta_{12}x^2 + \theta_{13}x^3$, (2) exponential: $c(x, \theta_1) = \theta_{11}\exp(\theta_{12}x)$, (3) hyperbolic: $c(x, \theta_1) = \frac{\theta_{11}}{91-x}$, and (4) square root $c(x, \theta_1) = \theta_{11}\sqrt{x}$.

Model Setup

- The monthly mileage $x_{t+1} - x_t$ has an arbitrary parametric density function g, which implies a transition density of the form

$$p(x_{t+1} \mid x_t, i_t, \theta_3) = \begin{cases} g(x_{t+1} - x_t, \theta_3) & \text{if } i_t = 0, \\ g(x_{t+1} - 0, \theta_3) & \text{if } i_t = 1. \end{cases} \tag{17}$$

- I assume that the unobserved state variables $\{\varepsilon_t(0), \varepsilon_t(1)\}$ obey an i.i.d. bivariate extreme value process, with mean normalized to $(0, 0)$ and variance normalized to $(\frac{\pi^2}{6}, \frac{\pi^2}{6})$.

Decomposing the Likelihood Function

- The data consists of $\{i_t^m, x_t^m\}$ $(t = 1, ..., T_m; m = 1, ..., M)$ where i_t^m is the engine replacement decision in month t for bus m and x_t^m is the mileage since last replacement of bus m in month t.
- Let's first consider a particular bus. According to Theorem 2, to maximize the log-likelihood function, we can define the following two partial likelihood function:

$$\ell^1(x_1, \ldots, x_T, i_1, \ldots, i_T \mid x_0, i_0, \theta) = \prod_{t=1}^{T} p(x_t \mid x_{t-1}, i_{t-1}, \theta_3) \tag{18}$$

$$\ell^2(x_1, \ldots, x_T, i_1, \ldots, i_T \mid \theta) = \prod_{t=1}^{T} P(i_t \mid x_t, \theta). \tag{19}$$

Estimation Procedure

- Stage 1: Maximize ℓ^1 to estimate θ_3, which is a data-driven process, especially we discretize the state variable x_t.
- Stage 2: Maximize ℓ^2, which requires us to calculate $P(i_t \mid x_t, \theta)$. But to do that, we need first the expression of EV_θ.
- Here, we introduce Nested Fixed Point Algorithm.

Nested Fixed Point Algorithm

- An "inner" fixed point algorithm computes the unknown function EV_θ for each value of θ. In our setting, this step calculates the fixed point to the contraction mapping in Equation 15.

$$EV_\theta(x,i) = \int_y \log\left\{\sum_{j\in C(y)} \exp\left[u(y,j,\theta_1) + \beta EV_\theta(y,j)\right]\right\} p(dy \mid x,i,\theta_3).$$

- And an "outer" hill climbing algorithm searches for the value of θ which maximizes the likelihood function ℓ^2, using Equation 14

$$P(i \mid x,\theta) = \frac{\exp\{u(x,i,\theta_1) + \beta EV_\theta(x,i)\}}{\sum_{j\in C(x)} \exp\{u(x,j,\theta_1) + \beta EV_\theta(x,j)\}}. \tag{20}$$

Summary

- Under CIA and this general dynamic discrete choice framework, we should understand the implications of Theorem 1 and 2. In particular, how to theoretically express the choice probabilities $P(i \mid x, \theta)$.
- Nested Fixed Point Algorithm: an “inner” fixed point algorithm computes the unknown function EV_θ for each value of θ, and an “outer” hill climbing algorithm searches for the value of θ which maximizes the likelihood function.
- Programming matters!