John Rust (1987, Econometrica) “Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher”

结构估计动态离散选择模型

李博

北京大学经济学院

2026 年 4 月 27 日

Motivation of Rust (1987)

Research Question:

- For how long should one continue to operate and maintain a bus before it is optimal to replace or rebuild the engine?
- Did the decision maker (the superintendent of maintenance, Harold Zurcher) behave according to the optimal replacement rule implied by the dynamic discrete choice model?
- Can the model accurately predict the mileage thresholds at which replacements occur?
- Critique of top-down replacement investment models (Jorgenson, 1973; Feldstein & Rothschild, 1974). Limitations of aggregate capital stock measures.
- Proposal of a bottom-up approach: aggregating individual replacement decisions to understand aggregate investment.
- Focus on micro-theoretic model of individual optimizing behavior.

What Rust Models?

Optimal replacement decision: Solution to a dynamic optimization problem that formalizes the trade-off between two conflicting objectives:

- *minimizing replacement costs*
- *minimizing operating and maintenance cost as well as unexpected engine failures*

(quality of engine declines over time, but replacing it is costly)

The Simple Regenerative Optimal Stopping Model

- State variable: Accumulated mileage (x_t).
- Operating costs: $c(x_t, \theta_1)$.
- Decision: Replace (1) or keep (0) the engine.
- Objective: Minimize expected discounted costs.
- Solution: Threshold mileage (y) for replacement.

Answers using structural estimation

Structural estimation: Using data on *monthly mileage and engine replacements* for a sample of GMC busses, Rust estimate the structural parameters in the engine replacement model using NFXP(Nested Fixed Point Algorithm).

Research Design: Data and Methodology

- **Data Source:** Detailed maintenance records from the Madison Metropolitan Bus Company (MTC). Provided by superintendent Harold Zurcher.
- **Time Span:** 1974 - 1984 (Approximately 10 years).
- **Frequency:** Monthly data.
- **Key Variables:**
  - Monthly odometer readings.
  - Dates of engine replacement/overhaul.
- **Core Methodology:**
  - Construct a Stochastic Dynamic Programming (SDP) model to describe the optimal replacement decision for a single engine.
  - Use Structural Estimation methods, specifically the Nested Fixed Point (NFXP) Algorithm, to estimate model parameters.
  - Utilize Maximum Likelihood Estimation (MLE).

Contributions of Rust (1987)

Main contributions

1. Development and implementation of Nested Fixed Point Algorithm
2. Formulation of assumptions, that makes dynamic discrete choice models tractable.
3. An illustrative application in a simple model of engine replacement.
4. The first ML estimates of discrete choice dynamic programming models

Components of the dynamic model

- **State variables** —vector of variables that describe all relevant information about the modeled decision process, x_t
- **Decision variables** —vector of variables describing the choices, d_t
- **Instantaneous payoff** —utility function, $u(x_t, d_t)$, with time separable discounted utility
- **Motion rules** —agent' s beliefs of how state variable evolve through time, conditional on choices, $x_{t+1} \sim F(x_t, d_t)$

Solution is given by:

- **Value function** —maximum attainable utility $V(x_t)$
- **Policy function** —mapping from state space to action space that returns the optimal choice, $d^\star(x_t)$

Zurcher optimization problem

- Choice set: $\{\text{keep}, \text{replace}\} = \{0, 1\}$

Each bus comes in for repair once a month and Zurcher chooses between ordinary maintenance $(d_t = 0)$ and overhaul/engine replacement $(d_t = 1)$.

- State space: mileage at time t since last engine overhaul x_t

Harold observes many different attributes of the buses which come into the shop, but we focus on the main one for now.

Zurcher's preferences

Instantaneous payoffs are given by the cost function that depends on the choice

$$u(x_t, d_t, \theta_1) = \begin{cases} -RC - c(0, \theta_1) & \text{if } d_t = \text{replace} = 1 \\ -c(x_t, \theta_1) & \text{if } d_t = \text{keep} = 0 \end{cases}$$

- RC = replacement cost
- $c(x, \theta_1)$ = cost of maintenance with preference parameters θ_1

Motion rules

- First of all, mileage is continuous. How to deal with the *continuous* state space?
- Rust discretized the range of travelled miles into $n = 175$ bins, indexed with i, $\hat{X} = \{\hat{x}_1, ..., \hat{x}_n\}$ with $\hat{x}_1 = 0$

Mileage transition probability: for $j = 0, ..., J$

$$p(x'|\hat{x}_k, d, \theta_2) = \begin{cases} Pr\{x' = \hat{x}_{k+j}|\theta_2\} = \theta_{2j} & \text{if } d = 0 \\ Pr\{x' = \hat{x}_{1+j}|\theta_2\} = \theta_{2j} & \text{if } d = 1 \end{cases}$$

- Mileage in the next period x' can move up at most J grid points
- J is determined from the observed distribution of mileage
- Effectively, the probabilities of increase of mileage from any existing mileage are the same

Transition matrix for mileage

If not replacing $(d = 0)$

$$\Pi(d=0)_{n\times n} = \begin{pmatrix} \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \theta_{20} & \theta_{21} & \theta_{22} & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & \theta_{20} & \theta_{21} & \theta_{22} \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & \theta_{20} & 1-\theta_{20} \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix}$$

Transition matrix for mileage

If replacing $(d = 1)$

$$\Pi(d=1)_{n\times n} = \begin{pmatrix} \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & \cdot & 0 \\ \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & \cdot & 0 \\ \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & \cdot & 0 \\ \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & \cdot & 0 \\ \theta_{20} & \theta_{21} & \theta_{22} & 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix}$$

Dynamic optimization problem

To minimize the discounted expected value of the costs, Zurcher should find such policy function $f(x_t) : X \rightarrow \{\text{keep}, \text{replace}\}$ such that $d_t = f(x_t)$ maximizes

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t u(x_t, d_t) \longrightarrow \max$$

(Note: The problem is often framed as minimizing costs, which is equivalent to maximizing negative costs as shown in the utility function earlier).

Value function

The function $V(x_t)$ denotes the maximum attainable value at period t

$$V(x_t) = \max_{\Pi} \mathbb{E} \sum_{j=t}^{\infty} \beta^{j-t} u(x_t, d_t)$$

where Π is a space of policy functions $f(x_t) : X \to \{\text{keep}, \text{replace}\}$, and $d_t = f(x_t)$.

Nested Fixed Point Algorithm (NFXP)

- **Motivation:** Overcome limitations of traditional estimation methods that require analytical solutions or simplifying assumptions.
- **Core Idea:** Embed the solution of the Bellman equation (finding the value function) within the maximization process of the likelihood function.
- **Two-Loop Structure:**
  - **Inner Loop: Fixed Point Iteration**
    - ★ Given a candidate parameter vector $\theta = (\beta, \theta_1, RC, \theta_3)$.
    - ★ Iteratively solve for the **expected value function** $EV(x; \theta)$ using the Bellman equation operator T_θ.
  - **Outer Loop: Likelihood Maximization**
    - ★ Use a numerical optimization algorithm to search for the optimal parameters $\hat{\theta}$ that maximize the likelihood of observing the data.
    - ★ Each time the likelihood function or its gradient needs to be evaluated for a given θ, the inner loop must be run to compute the corresponding $EV(x; \theta)$.
- **Advantages:**
  - Handles complex DP models without analytical solutions.
  - Allows for consistent estimation of models with unobserved state variables/shocks.
  - High flexibility in specifying model components.
- **Computational Cost:** Can be computationally intensive, especially for the time (1987). Rust's work was a pioneering successful application.

NFXP Algorithm

Explanation:

1. **Outer Loop Start:** Choose initial parameters $\theta^{(0)}$.
2. **Inner Loop (for current $\theta^{(k)}$):**
   - Iterate $EV_{j+1} = T_{\theta^{(k)}}(EV_j)$ starting from some EV_0 until EV_j converges to $EV(x; \theta^{(k)})$.
3. **Likelihood Evaluation:**
   - Use the computed $EV(x; \theta^{(k)})$ and the observed data (x_t, d_t) to calculate the likelihood function value $\mathcal{L}(\theta^{(k)})$. (This typically involves calculating choice probabilities $P(d_t|x_t; \theta^{(k)})$ based on EV).
4. **Outer Loop Update:**
   - Use an optimization algorithm based on $\mathcal{L}(\theta^{(k)})$ and possibly its gradient/Hessian to find the next parameter guess $\theta^{(k+1)}$.
5. **Convergence Check:** Check if the outer loop has converged (e.g., $\theta^{(k+1)} \approx \theta^{(k)}$ or gradient is near zero). If not, return to Step 2 with $\theta^{(k+1)}$.

Recursive form of the maximization problem (Bellman equation)

Using Bellman Principle of Optimality, we can show that the value function $V(x_t)$ constitutes the solution of the following functional equation

$$V(x) = \max_{d \in \{\text{keep},\text{replace}\}} \Big\{ u(x, d) + \beta \mathbb{E}\big[V(x') \big| x, d \big] \Big\}$$

where expectation is taken over the next period values of state x' given the motion rule of the problem.

Bellman equation and Bellman operator

$$V(x) = \max_{d \in \{\text{keep,replace}\}} \left\{ u(x,d) + \beta \mathbb{E}\left[V(x') \middle| x, d \right] \right\}$$

can be written as a fixed point equation of the **Bellman operator** in the functional space

$$T(V)(x) \equiv \max_{d \in \{\text{keep,replace}\}} \left\{ u(x,d) + \beta \mathbb{E}\left[V(x') \middle| x, d \right] \right\}$$

The Bellman equations is then $V(x) = T(V)(x)$, with the solution given by the fixed point $T(V) = V$.

Contraction Mapping Theorem (Banach Fixed Point Theorem)

Theorem

Let (S, ρ) be a complete metric space with a contraction mapping $T : S \to S$.

Then

1. T admits a unique fixed-point $V^\star \in S$, i.e. $T(V^\star) = V^\star$.
2. $V^\star$ can be found by repeated application of the operator T, i.e. $T^n(V) \to V^\star$ as $n \to \infty$.

Value function iterations solution algorithm

Algorithm:

1. Initialize the value function V with arbitrary values
2. Solve Bellman equation to compute $T(V)$
3. Repeat until convergence

- Unique fixed point $\Leftrightarrow$ unique solution to the Bellman equation
- The fixed point algorithm can start from **arbitrary initial guess**! VFI algorithm converges globally
- Direct resemblance with backward induction, only have to stop after unspecified number of iterations

Making the model suitable for empirical work

- Remember that Zurcher observes many different attributes of the busses that come into the shop
- But we as an econometrician do not!
- Yet, these are likely to be the reason for observing different behavior in same states (for the same observed mileage)

Need to include **error terms** into the model, denote ϵ.

Three independence assumptions

1. Error terms are **independent across observations** due to random sampling
2. Error terms come in pairs, one for each decision $d = 0$ and $d = 1$, and are **independent across choices**
3. Conditional on x, there is no serial correlation in error terms **across time**

Modified Bellman equation

ε is a new (vector) state variable

$$V(x,\varepsilon) = \max_{d\in\{0,1\}} \left\{ u(x,\varepsilon_d,d) + \beta \mathbb{E}\left[V(x',\varepsilon') \middle| x,\varepsilon,d \right] \right\}$$

$$V(x,\varepsilon) = \max_{d\in\{0,1\}} \left\{ u(x,\varepsilon_d,d) + \beta \int_X \int_\Omega V(x',\varepsilon') p(x',\varepsilon'|x,\varepsilon,d) dx' d\varepsilon' \right\}$$

where ε_d is the component of vector $\varepsilon \in \mathbb{R}^2$ which corresponds to d.

Rust assumptions

(AS) Additive separability in preferences

$$u(x, \varepsilon_d, d) = u(x, d) + \varepsilon_d,$$

(CI) Conditional independence

$$p(x', \varepsilon' | x, \varepsilon, d) = q(\varepsilon' | x') \cdot \pi(x' | x, d)$$

(EV) Extreme value Type I (EV1) distribution of ε

What Rust assumptions allow:

$$V(x,\varepsilon) = \max_{d\in\{0,1\}} \left\{ u(x,d) + \varepsilon_d + \beta \int_X \int_\Omega V(x',\varepsilon')\pi(x'|x,d)q(\varepsilon'|x')dx'd\varepsilon' \right\}$$

1. Separate out the deterministic part of **choice specific value function** $v(x,d)$ (SA)
2. Compute the expectation by part (CI)
3. Use max-stability of EV1 to compute expectation w.r.t. ε' (EV)

Simplifying the Expectation

$$V(x,\varepsilon) = \max_{d\in\{0,1\}} \Big\{ \underbrace{u(x,d) + \beta \int_X \Big(\int_\Omega V(x',\varepsilon') q(\varepsilon'|x') d\varepsilon' \Big) \pi(x'|x,d) dx'}_{v(x,d)} + \varepsilon_d \Big\}$$

$$V(x',\varepsilon') = \max_{d'\in\{0,1\}} \{ v(x',d') + \varepsilon'_{d'} \}$$

Using properties of EV1 distribution for ε':

$$\mathbb{E}_{\varepsilon'}\left[V(x',\varepsilon') | x' \right] = \int_\Omega V(x',\varepsilon') q(\varepsilon'|x') d\varepsilon' = \log\left(\exp[v(x',0)] + \exp[v(x',1)] \right) + \gamma$$

where γ is the Euler-Mascheroni constant (often normalized away). Assuming $\gamma = 0$:

$$\mathbb{E}\left[V(x',\varepsilon') \middle| x,d \right] = \int_X \Big(\log\left(\exp[v(x',0)] + \exp[v(x',1)] \right) \Big) \pi(x'|x,d) dx'$$

Expected value function

Let $\mathbb{E}\left[V(x', \varepsilon')\middle|x, d\right] = EV(x, d)$ denote the expected value function, then:

$$
\begin{aligned}
EV(x, d) &= \int_X \log\left(\exp[v(x', 0)] + \exp[v(x', 1)]\right) \pi(x'|x, d) dx' \\
v(x, d) &= u(x, d) + \beta EV(x, d)
\end{aligned}
$$

- this is Bellman equation in expected value function space
- when the state space is discrete the integral is, of course, a simple sum over future values

Benefits of Rust's approach

- The Bellman operator expressed in terms of expected values (like $EV(x)$ or $v(x,d)$) is also a contraction mapping $\Rightarrow$ VFI works!
- Dimensionality of this fixed point problem is smaller than the one involving $V(x,\varepsilon)$ (because ε is integrated out).
- It is also numerically easier to work with smooth expected values rather than the potentially non-smooth $V(x,\varepsilon)$.

Three-Stage Estimation Strategy

- **Motivation:** To reduce the computational burden of Full Information Maximum Likelihood (FIML) and to isolate the estimation of different parameter subsets.
- **Stage 1: Estimate Mileage Process Parameters θ_3**
  - Uses only the mileage data $\{x_{it}\}$ to estimate the parameters of the transition probability $g(x_{t+1}|x_{it}, d_{it}; \theta_3)$ via partial likelihood maximization.
  - $\mathcal{L}_1(\theta_3) = \sum_{i,t} \ln g(x_{it+1}|x_{it}, d_{it}; \theta_3)$ (Ignoring choice probabilities).
  - Used to test homogeneity assumptions about the mileage process across and within bus groups (Tables V, VI in the paper).
  - Used to test the random walk hypothesis for mileage increments (Table VII).

Three-Stage Estimation Strategy

- **Stage 2: Estimate Structural Parameters** (β, RC, θ_1)
  - Fixes $\hat{\theta}_3$ at the estimate from Stage 1.
  - Estimates the core structural parameters (β, RC, θ_1) by maximizing the conditional likelihood based only on the choice probabilities $P(d_{it}|x_{it}; \beta, RC, \theta_1, \hat{\theta}_3)$.
  - $\mathcal{L}_2 = \sum_{i,t} \ln P(d_{it}|x_{it}; \beta, RC, \theta_1, \hat{\theta}_3)$ (Ignoring transition probabilities).
  - **This stage heavily utilizes the NFXP algorithm** to compute the EV needed for the choice probabilities.
  - Used to compare different functional forms for the cost function $c(x, \theta_1)$ (Table VIII).
- **Stage 3: Full Information Maximum Likelihood (FIML)**
  - Uses the estimates from Stages 1 and 2 as **starting values**.
  - Simultaneously estimates all parameters $\theta = (\beta, RC, \theta_1, \theta_3)$ by maximizing the full log-likelihood function $\mathcal{L}_f(\theta) = \mathcal{L}_1(\theta_3) + \mathcal{L}_2(\beta, RC, \theta_1|\theta_3)$.
  - Provides the asymptotically most efficient estimates. (Results in Tables IX, X).

Choice probabilities

Once the fixed point (e.g., $v(x, d)$) is found, the *optimal* choice probability $P(d|x)$ is given by the Logit structure (assumption EV):

$$P(d|x) = \frac{\exp[v(x, d)]}{\sum_{d' \in \{0,1\}} \exp[v(x, d')]}$$

The choice probability serves as the basis for forming the likelihood function.

Maximum likelihood estimator (MLE)

Maximum likelihood estimator is applicable when the model yields a probability distribution for the observable data

- Let $L(x, \theta)$ denote the distribution (pdf) of the observables x implied by the model with parameter vector θ
- Let $Z_n = (z_i, \ldots, z_n)$ denote the data, consisting of n independent observations (random sample)

Likelihood function

- If $L(x, \theta)$ is a discrete distribution then $L(x, \theta)$ computed at the data point gives the probability of observing the given data point exactly
- If $L(x, \theta)$ is a continuous distribution then $L(x, \theta)$ computed at the data point is analogous to the probability of observing the given data point

Likelihood function is

$$\mathcal{L}_n(\theta) = L(Z_n, \theta) = \prod_{i=1}^{n} L(z_i, \theta)$$

Definition of MLE estimator

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} \sum_{i=1}^{n} \underbrace{\log L(z_i, \theta)}_{\ell(z_i,\theta)} = \arg\max_{\theta \in \Theta} \ell_n(\theta)$$

- $\theta \in \Theta$ is parameter space
- $Z_n = (z_1, \ldots, z_n)$ is observed data

Where we are so far

- for every value of structural parameters $\theta = (RC, \theta_1, \theta_{20}, \ldots, \theta_{2J}, \beta)$
- we can solve the model using VFI
- based on the solution $EV_\theta(x)$ (or related $v_\theta(x, d)$) can form choice probabilities $P(\text{keep}|x, \theta)$ and $P(\text{replace}|x, \theta)$

Next: given the data on mileage (x) and choices (d), can form likelihood function $L(x, d, \theta)$, and proceed with maximum likelihood estimation

Data

- Harold Zurcher’s Maintenance records of 162 busses in 8 groups
- Monthly observations of mileage on each bus (odometer reading)
- Data on maintenance operations:
  1. Routine periodic maintenance (i.e. brake adjustment)
  2. Replacement or repair at time of failure
  3. Major engine overhaul and/or replacement (*the focus of the paper*)

Data $(x_{i,t}, d_{i,t})$ where $x_{i,t}$ is discretized mileage (bin indexes), and $d_{i,t}$ is the observed choice at this mileage for each bus i in each month t.

Likelihood function (Zurcher specific)

Assuming the full likelihood includes both choice probability and state transition probability:

$$\mathcal{L}_n(\theta, EV_\theta) = \prod_{i=1}^{162} \prod_{t=2}^{T_i} P(d_{i,t}|x_{i,t}, \theta)\pi(x_{i,t}|x_{i,t-1}, d_{i,t-1}, \theta)$$

Log-likelihood:

$$\ell_n(\theta, EV_\theta) = \log \mathcal{L}_n(\theta, EV_\theta) = \sum_{i=1}^{162} \sum_{t=2}^{T_i} \big(\log P(d_{i,t}|x_{i,t}, \theta) + \log \pi(x_{i,t}|x_{i,t-1}, d_{i,t-1}, \theta) \big)$$

MLE estimator (Zurcher specific)

$$\hat{\theta} = \arg\max_{\theta} \ell_{\boldsymbol{n}}(\theta, EV_{\theta})$$

Unconstrained optimization, but requires the computation of EV_{θ} for each value of parameter θ. (EV_{θ} represents the solution to the Bellman fixed point for a given θ).

Nested loop

Outer loop = Hill-climbing algorithm

- Log-likelihood function $\ell_n(\theta, EV_\theta)$ is maximized with respect to θ
- Quasi-Newton algorithm: Usually BHHH, BFGS or a combination
- Each evaluation of $\ell_n(\theta, EV_\theta)$ requires solution for the fixed point EV_θ

Inner loop = Fixed point algorithm

- Solver for the fixed point of the Bellman operator (e.g., $EV_\theta = T^*(EV_\theta)$)
- Successive approximations (VFI)

Key Estimation Results

- **Mileage Process (Stage 1, Tables V-VII):**
  - Rejected the hypothesis that all bus groups share the same mileage process $\implies$ heterogeneity between groups.
  - Could generally not reject within-group homogeneity for the mileage process.
  - Monthly mileage increment ξ found to be approximately independent of the current mileage level x_t (supporting the "random walk" like assumption).
- **Cost Function Form (Stage 2, Table VIII):**
  - Marginal operating costs increase significantly with accumulated mileage.
  - Square root, quadratic, and cubic models provided significantly better fits than the linear model.
- **Structural Parameter Estimates (Stage 3, Tables IX, X):**
  - Estimated Replacement Cost (RC) and cost function parameters (θ_1) were statistically significant and economically plausible (costs increase with mileage).
  - The discount factor β was estimated to be very close to 1 (e.g., 0.999), implying either a very low discount rate (high patience) or that the monthly decision interval is very short relative to the agent's time preference.
  - Parameter estimates showed relative robustness to the number of grid points used for state space discretization (e.g., 90 vs 175 points).
  - Using external data on average overhaul costs (Table III), the estimated RC and θ_1 parameters were scaled to provide cost estimates in dollar terms.

Model Fit and Evaluation

- **Estimated Value and Hazard Functions (Figures 2 & 3):**
  - The estimated expected value function $EV(x)$ decreases smoothly with mileage x (reflecting increasing future costs).
  - The implied **replacement probability function** $P(1|x)$ **(the hazard rate)** increases smoothly with mileage x.
  - This smooth hazard rate contrasts sharply with the step-function predicted by the simple model without unobservables and matches the data better.
  - Comparisons between the estimated dynamic model $(\beta < 1)$ and a myopic model $(\beta = 0)$ showed differences in predicted behavior, highlighting the importance of forward-looking behavior.
- **Test of Conditional Independence (CI) Assumption (Table XI):**
  - Rust performed specification tests (Lagrange Multiplier tests) to check the validity of the crucial CI assumption (i.e., are the ϵ_t shocks truly independent of past information conditional on x_t?).
  - The test results **rejected the Conditional Independence assumption**.
  - **Important Implication:** This suggests that the model might be misspecified. There could be unobserved state variables that are persistent over time (e.g., the engine's true underlying "health" state beyond just mileage) or serial correlation in the cost shocks ϵ_t. This points towards avenues for future research with richer models.

Data Limitations & Modeling Considerations I

- **Mileage Data Issues:**
  - **Missing monthly odometer readings** exist. Rust used imputation methods to handle this.
  - Potential **measurement error** in odometers, though not explicitly modeled.
- **Definition of "Replacement":**
  - Data records "engine overhaul". The model assumes this is equivalent to resetting the engine to a near-new state.
  - Does not distinguish between potential variations in the quality or extent of the overhaul.
- **Modeling Implications:**
  - Simple models based only on mileage fail to explain observed replacement patterns (e.g., why buses with similar mileage aren't always treated the same).
  - **Necessitates introducing unobserved state variables or shocks (ϵ_t)**. This is a key reason for choosing a stochastic DP model with shocks and using the NFXP algorithm.
  - Model parameters (like cost function parameters θ_1) will capture the average effects of some unobservables.
- **Crucial Unobserved Variables:** ***This is the core challenge!***
  - **Individual Engine Heterogeneity:** Intrinsic "health" or condition beyond mileage is unobserved.

Data Limitations & Modeling Considerations II

- **Maintenance Practice Variation:** Quality of routine maintenance, specific repair events not captured.
- **Operating Conditions:** Route severity, weather patterns, traffic conditions.
- **Driver Behavior:** Driving habits affecting wear and tear.

Methods for solving Dynamic Discrete Choice Models

- Rust (1987): Nested-Fixed Point Algorithm (NFXP)
- Hotz and Miller (1993): CCP estimator (two step estimator)
- Keane and Wolpin (1994): Simulation and interpolation
- Rust (1997): Randomization algorithm
- Aguirregabiria and Mira (2002): Nested Pseudo Likelihood (NPL)
- Bajari, Benkard and Levin (2007): Two step min. distance
- Arcidiacono Miller (2002): CCP with unobserved heterogeneity (EM Algorithm)
- Norets (2009): Bayesian Estimation
- Su and Judd (2012): MLE using constrained optimization (MPEC)
- Any estimator method or solution algorithm of DDC models must confront Harold Zurcher