Numerical Methods-Lecture XIII: Dynamic Discrete Choice Estimation

(See Rust 1987)

Trevor Gallen

Introduction

- Rust (1987) wrote a paper about Harold Zurcher's decisions as the superintendent of maintenance at the Madison (Wisconsin) Metropolitan Bus Company.
  - 10 years of monthly data
  - bus mileage and engine replacement
  - 104 buses
  - 1 Harold

Introduction

- Rust (1987) wrote a paper about Harold Zurcher's decisions as the superintendent of maintenance at the Madison (Wisconsin) Metropolitan Bus Company.
  - 10 years of monthly data
  - bus mileage and engine replacement
  - 104 buses
  - 1 Harold
- Not interested in Harold per se

Introduction

- Rust (1987) wrote a paper about Harold Zurcher's decisions as the superintendent of maintenance at the Madison (Wisconsin) Metropolitan Bus Company.
  - 10 years of monthly data
  - bus mileage and engine replacement
  - 104 buses
  - 1 Harold
- Not interested in Harold per se
- Interested in application of Dynamic Discrete Choice framework

Description

- Observe monthly bus mileage
- Observe maintenance diary with date, milage, and list of components repaired or replaced
- Three types of maintenance operations
  - Routine adjustments (brake adjustments, tire rotation)
  - Replacement of individual components when failed
  - Major engine overhauls
- Model Zurcher's decision to replace bus engines based on observables *and unobservables*

Description

- Observe monthly bus mileage
- Observe maintenance diary with date, milage, and list of components repaired or replaced
- Three types of maintenance operations
  - Routine adjustments (brake adjustments, tire rotation)
  - Replacement of individual components when failed
  - Major engine overhauls
- Model Zurcher's decision to replace bus engines based on observables *and unobservables*

Model

- Agent is forward looking
- Maximizes expected intertemporal payoff
- Estimate parameters of the models
- Test whether the agent's behavior is consistent with the model

Replacement data-I

TABLE IIa

SUMMARY OF REPLACEMENT DATA

(Subsample of buses for which at least 1 replacement occurred)

	Mileage at Replacement				Elapsed Time (Months)				
Bus Group	Max	Min	Mean	Standard Deviation	Max	Min	Mean	Standard Deviation	Number of Observations
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	273,400	124,800	199,733	37,459	74	38	59.1	10.9	27
4	387,300	121,300	257,336	65,477	116	28	73.7	23.3	33
5	322,500	118,000	245,291	60,258	127	31	85.4	29.7	11
6	237,200	82,400	150,786	61,007	127	49	74.7	35.2	7
7	331,800	121,000	208,963	48,981	104	41	68.3	16.9	27
8	297,500	132,000	186,700	43,956	104	36	58.4	22.2	19
Full Sample	387,400	83,400	216,354	60,475	127	28	68.1	22.4	124

Replacement data-II

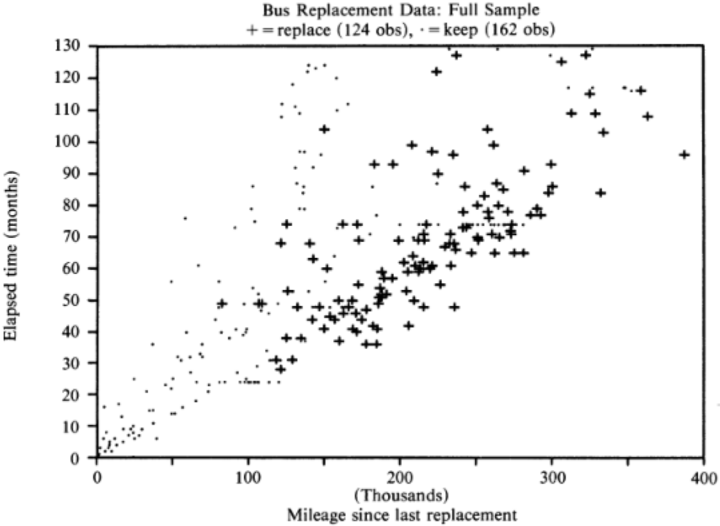
TABLE IIb

Censored Data

(Subsample of buses for which no replacements occurred)

	Mileage at May 1, 1985				Elapsed Time (months)				
Bus Group	Max	Min	Mean	Standard Deviation	Max	Min	Mean	Standard Deviation	Number of Observations
1	120,151	65,643	100,117	12,929	25	25	25	0	15
2	161,748	142,009	151,183	8,530	49	49	49	0	4
3	280,802	199,626	250,766	21,325	75	75	75	0	21
4	352,450	310,910	337,222	17,802	118	117	117.8	0.45	5
5	326,843	326,843	326,843	0	130	130	130	0	1
6	299,040	232,395	265,264	33,332	130	128	129.3	1.15	3
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
Full Sample	352,450	65,643	207,782	85,208	130	25	66.4	34.6	49

Replacement data-III



Sample

- Focus on bus groups 1-4
- Most recent acquisitions
- Have data on replacement costs only for this group
- Utilization fairly constant within each group (necessary for model)

Model - I

- Harold chooses $\{i_1, i_2, ..., i_t, ...\}$ to maximize his expected utility

$$\max_{\{i_1, i_2, \ldots, i_t, \ldots\}} E_t \sum_{t=1}^{\infty} \beta^{t-1} u(x_t, \varepsilon_t, i_t; \theta)$$

- x_t is the total mileage on an engine since last replacement
- ε_t are unobservable (to the econometrician) shocks
- i_t is an indicator of engine replacement
- θ is vector of parameters (to be discussed below)

Model - II

- Cost function

$$c(x,\theta_1) = m(x,\theta_{11}) + \mu(x,\theta_{12})b(x,\theta_{13})$$

- $m(x,\theta_{11})$ is the conditional expectation of normal maintenance and operation expenditure
- $\mu(x,\theta_{12})$ is the conditional probability of an unexpected engine failure
- $b(x,\theta_{13})$ is the conditional expectation of towing costs, repair costs, and loss of customer goodwill costs resulting from engine failure
- No data on maintenance and operating costs, so only estimating the sum of costs $c(x,\theta_1)$

Model - III

- Stochastic process governing $\{i_t, x_t\}$ is the solution to

$$V_\theta(x_t) = \sup_{\Pi} E \left\{ \sum_{j=t}^{\infty} \beta^{j-t} u(x_j, f_j, \theta_1) \,|x_t \right\}$$

- where utility u is given by

$$u(x_t, i_t, \theta_1) = \begin{cases} -c(x_t, \theta_1) & \text{if} \quad i_t = 0 \\ -\left[\, \overline{P} - \underline{P} + c(0, \theta_1)\right] & \text{if} \quad i_t = 1 \end{cases}$$

- $\Pi = \{f_t, f_{t+1}, ...\}$ is a sequence of decision rules where
- each f indicates the optimal choice (replace or not) at time t given the entire history of investment $i_{t-1}, ..., i_1$ and mileage $x_t, x_{t-1}, ..., x_1$ observed to date
- $\underline{P}$ is the scrap value of an engine
- $\overline{P}$ is the cost of a new engine

Model - IV

- Evolution of x_t is given by stochastic process:

$$p(x_{t+1}|x_t, i_t, \theta_2)$$

$$= \begin{cases} \theta_2 \exp[\theta_2(x_{t+1} - x_t)] & \textit{if} \quad i_t = 0 \quad \& \quad x_{t+1} \geq x_t \\ \theta_2 \exp[\theta_2(x_{t+1})] & \textit{if} \quad i_t = 1 \quad \& \quad x_{t+1} \geq 0 \\ 0 & o/w \end{cases}$$

- Without investment, next period mileage is drawn from exponential CDF $1 - \exp[\theta_2(x_{t+1} - x_t)]$

- With investment, next period mileage is drawn from exponential CDF $1 - \exp[\theta_2(x_{t+1} - 0)]$

Model - V

Bellman:

$$V_\theta(x_t) = \max_{i_t \in \{0,1\}} u(x_t, i_t, \theta_1) + \beta EV_\theta(x_t, i_t)$$

$$EV_\theta(x_t, i_t) = \int_o^\infty V_\theta(y) P(dy|x_t, i_t, \theta_2)$$

$$i_t = f(x_t, \theta) = \begin{cases} 1 & if \quad x_t > \gamma(\theta_1, \theta_2) \\ 0 & if \quad x_t \leq \gamma(\theta_1, \theta_2) \end{cases}$$

- Where $\gamma(\theta_1, \theta_2)$ is the investment cut-off ("optimal stopping barrier") given by the unique solution to

$$(\ \overline{P} - \underline{P})(1-\beta) = \int_0^{\gamma(\theta_1,\theta_2)} [1 - \beta \exp\{-\theta_2(1-\beta)y\}] \frac{\partial c(y, \theta_1)}{\partial y dy}$$

We have a closed-form solution!?

Model - VI

- Problem: all engines are replaced at same point.
- How do we solve this? (In standard MLE)?

Model - VI

- Problem: all engines are replaced at same point.
- How do we solve this? (In standard MLE)?
- Add an error term:

$$i_t = f(x_t, \theta) + \varepsilon_t$$

Model - VI

- Problem: all engines are replaced at same point.
- How do we solve this? (In standard MLE)?
- Add an error term:

$$i_t = f(x_t, \theta) + \varepsilon_t$$

- But that's dumb! Why?

Model - VI

- Problem: all engines are replaced at same point.
- How do we solve this? (In standard MLE)?
- Add an error term:

$$i_t = f(x_t, \theta) + \varepsilon_t$$

- But that's dumb! Why?
- Whoops! Accidentally replaced! (Assuming agent is making decisions without thinking about the error term before this period)
- Reinterpret: ε_t is an unobservable state variable known to the agent but not the econometrician

Model - VI

- Problem: all engines are replaced at same point.
- How do we solve this? (In standard MLE)?
- Add an error term:

$$i_t = f(x_t, \theta) + \varepsilon_t$$

- But that's dumb! Why?
- Whoops! Accidentally replaced! (Assuming agent is making decisions without thinking about the error term before this period)
- Reinterpret: ε_t is an unobservable state variable known to the agent but not the econometrician
- Add structure so model is consistent: no "mistakes" just unobserved to econometrician but not agent data

Model - VII

$C(x_t)$	Choice set. A finite set of allowable values of the control variable i_t when state variable is x_t
$\varepsilon_t = \{e_t(i) \mid i \in C(x_t)\}$	A $\#C(x_t)$-dimensional vector of state variables observed by agent by not by the econometrician.
$x_t = \{x_t(1), ..., x_t(K)\}$	$K-$dimensional vector of state variables observed by both the agent and the econometrician.
$u(x_t, i_t, \theta_1) + \varepsilon_t(i)$	Realized single period utility of decision i when state variable is $x_t, \varepsilon_t)$. θ_1 is a vector of unknown parameters to be estimated
$p(x_{t+1}, \varepsilon_{t+1} \mid x_t, \varepsilon_t, i_t, \theta_2, \theta_3)$	Markov transitional denisty for state variable (x_t, ε_t) when alternative i_t is selected. θ_1 and θ_2 are vectors of unknown parameters to be estimated.
$\theta = (\beta, \theta_1, \theta_2, \theta_3)$	The complete $1 + K_1 + K_2 + K_3$ vector of parameters to be estimated

Model - VIII

Now we have our dynamic discrete choice model. Infinite horizon, discounted Harold decision problem:

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

The optimal value function V_θ is the unique solution to

$$V_\theta(x_t, \varepsilon_t) = \max_{i_t \in C(x_t)} [u(x_t, i_t, \theta_1) + \varepsilon_t(i) + \beta E V_\theta(x_t, \varepsilon_t, i_t)]$$

with the decision rule

$$i_t = f(x_t, \varepsilon_t, \theta) \equiv arg \max_{i_t \in C(x_t)} [u(x_t, i_t, \theta_1) + \varepsilon_t(i_t) + \beta E V_\theta(x_t, \varepsilon_t, i_t)]$$

Model - IX

Problems:

- ε_t appears nonlinearly. Have to integrate out over ε_t to obtain choice probabilities
- Dimensionality: grid approach to estimating ε would still be too large to be computationally tractable (especially in 1987).

Model - X

- Don't want to integrate over all possible x_t and ϵ_t combinations

- Fortunately, this problem is separable (the two are unrelated):

- Conditional Independence Assumption:

  $$P(x_{t+1}, \varepsilon_{t+1} | x_t, \varepsilon_t, i, \theta_2, \theta_3) = q(\varepsilon_{t+1} | x_{t+1}, \theta_2) P(x_{t+1} | x_t, i, \theta_3)$$

- Estimate the transition matrix of miles

- Then, given the miles today, have disturbance term

Estimation-I

- Three likelihood functions. Two partial likelihoods, and one full.
- Mileage transition probability (first step)

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

- Choice given mileage (and shock) (second step)

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

And the full likelihood function:

$$\ell^f(x_1, ..., x_T, i_1, ..., i_T | x_0, i_0, \theta) = \prod_{t=1}^{T} P(i_t | x_t, \theta) p(x_t | x_{t-1}, i_{t-1}, \theta_3)$$

Three stages of estimation.

Estimation-II

Recall, have value function which is given by the functional equation

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

Goal is to estimate θ using a nested fixed point algorithm:

- For each θ, compute EV_θ using a fixed point algorithm
- Outer hill climbing algorithm searches for the value of θ which maximizes the likelihood function.

Estimation-III

First estimate the parameters θ_3 of the transition probability

$$p(x_{t+1}|x_t, i_t, \theta_3) = \begin{cases} g(x_{t+1} - x_t, \theta_3) & if \quad i_t = 1 \\ g(x_{t+1} - 0, \theta_3) & if \quad i_t = 0 \end{cases}$$

Using

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

- g is a multinomial distribution on the set $\{0, 1, 2\}$ corresponding to monthly mileage intervals $[0, 5000), [5000, 10000), [10000, \infty)$
- so $\theta_{3j} = Pr\{x_{t+1} = x_t + j|x_t, i_t = 0\}$, $j = 0, 1$
- This first stage doesn't require estimation of EV_θ!

Idea of Estimation

- Need to estimate mileage cost parameters θ_1, replacement cost RC, and transition parameters θ_3
- First, estimate transition probabilities θ_3
- Then, use these transitions and guess at θ_1 and RC to solve for V and simulate out. Choose θ_1 and RC parameters to best fit likelihood of replacement
- Then estimate all together using consistent first stage estimates
- Important part is nested fixed point estimation:
    1. Guess parameters
    2. Given parameters, solve for V
    3. Simulate outcomes
    4. Calculate error between outcomes and data: go back to step 1

Estimation-IV

TABLE V

Within Group Estimates of Mileage Process
Within Group Heterogeneity Tests

(Standard errors in parentheses)

	Group 1 1983 Grumman	Group 2 1981 Chance	Group 3 1979 GMC	Group 4 1975 GMC	Group 5 1974 GMC (8V)	Group 6 1974 GMC (6V)	Group 7 1972 GMC (8V)	Group 8 1972 GMC (6V)
θ_{31}	.197	.391	.307	.392	.489	.618	.600	.722
	(.021)	(.035)	(.008)	(.007)	(.013)	(.014)	(.010)	(.009)
θ_{32}	.789	.599	.683	.595	.507	.382	.397	.278
	(.021)	(.035)	(.008)	(.007)	(.013)	(.014)	(.010)	(.009)
θ_{33}	.014	.010	.010	.013	.005	.000	.003	.000
	(.006)	(.007)	(.002)	(.002)	(.002)	(0)	(.001)	(0)
Restricted Log, Likelihood	−203.99	−138.57	−2219.58	−3140.57	−1079.18	−831.05	−1550.32	−1330.35
Unrestricted Log Likelihood	−187.71	−136.77	−2167.04	−3094.38	−1068.45	−826.32	−1523.49	−1317.69
Likelihood ratio test statistic	32.56	3.62	105.08	92.39	21.46	9.46	53.67	25.31
Degrees of Freedom	42	9	141	108	33	18	51	34
Marginal Significance Level	.852	.935	.990	.858	.939	.948	.372	.859

Estimation-V

TABLE VI

Between Group Estimates of Mileage Process
Between Group Heterogeneity Tests

(Standard errors in parentheses)

	1, 2, 3	1, 2, 3, 4	4, 5	6, 7	6, 7, 8	5, 6, 7, 8	Full Sample
θ_{31}	.301	.348	.417	.607	.652	.618	.475
	(.007)	(.005)	(.006)	(.008)	(.006)	(.006)	(.004)
θ_{32}	.688	.639	.572	.392	.347	.380	.517
	(.007)	(.005)	(.007)	(.008)	(.006)	(.006)	(.004)
θ_{33}	.011	.012	.011	.002	.001	.002	.007
	(.002)	(.001)	(.001)	(.001)	(.004)	(.001)	(.000)
Restricted Log Likelihood	−2575.98	−5755.00	−4243.73	−2384.50	−3757.76	−4904.41	−11,237.68
Unrestricted Log Likelihood	−2491.51	−5585.89	−4162.83	−2349.81	−3668.50	−4735.95	−10,321.84
Likelihood ratio test statistic	168.93	338.21	161.80	69.39	180.52	336.93	1,831.67
Degrees of Freedom	198	309	144	81	135	171	483
Marginal Significance Level	.934	.121	.147	.818	.005	1.5E−17	7·7E−10

Estimation-VI

Given θ_3, and using ℓ^2 we can obtain estimates for β, θ_1, and RC ($RC = \overline{P} - \underline{P}$):

$$\ell^2(x_1, ..., x_T, i_1, ...i, T)|\theta) = \prod_{t=1}^{T} P(i_t|x_t, \theta)$$

where

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

with

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

and

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

Estimation-VII

- θ_1 are the parameters of the cost function
- Compare different (parsimonious) specifications
- Linear and square root forms do well

Estimation-VIII

TABLE VIII

Summary of Specification Search[a]

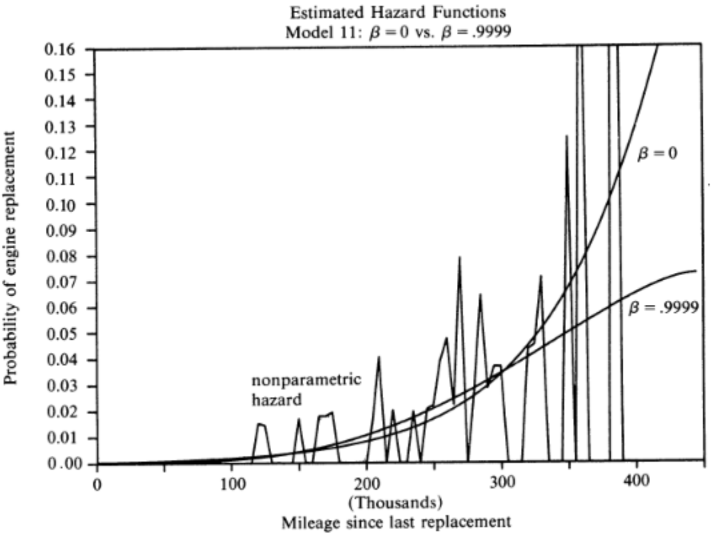
Cost Function	Bus Group 1, 2, 3	Bus Group 4	Bus Group 1, 2, 3, 4
Cubic $c(x, \theta_1) = \theta_{11}x + \theta_{12}x^2 + \theta_{13}x^3$	Model 1 −131.063 −131.177	Model 9 −162.885 −162.988	Model 17 −296.515 −296.411
quadratic $c(x, \theta_1) = \theta_{11}x + \theta_{12}x^2$	Model 2 −131.326 −131.534	Model 10 −163.402 −163.771	Model 18 −297.939 −299.328
linear $c(x, \theta_1) = \theta_{11}x$	Model 3 −132.389 −134.747	Model 11 −163.584 −165.458	Model 19 −300.250 −306.641
square root $c(x, \theta_1) = \theta_{11}\sqrt{x}$	Model 4 −132.104 −133.472	Model 12 −163.395 −164.143	Model 20 −299.314 −302.703
power $c(x, \theta_1) = \theta_{11}x^{\theta_{12}}$	Model 5[b] N.C. N.C.	Model 13[b] N.C. N.C.	Model 21[b] N.C. N.C.
hyperbolic $c(x, \theta_1) = \theta_{11}/(91 - x)$	Model 6 −133.408 −138.894	Model 14 −165.423 −174.023	Model 22 −305.605 −325.700
mixed $c(x, \theta_1) = \theta_{11}/(91 - x) + \theta_{12}\sqrt{x}$	Model 7 −131.418 −131.612	Model 15 −163.375 −164.048	Model 23 −298.866 −301.064
nonparametric $c(x, \theta_1)$ any function	Model 8 −110.832 −110.832	Model 16 −138.556 −138.556	Model 24 −261.641 −261.641

[a] First entry in each box is (partial) log likelihood value ℓ^2 in equation (5.2)) at $\beta = .9999$. Second entry is partial log likelihood value at $\beta = 0$.

Estimation-IX

- Next look at “myopic” replacement rule
- replace only when operating costs $c(x_t, \theta_1)$ exceed current cost of replacement $RC + c(0, \theta_1)$
- This model is rejected
- $\beta = .999$ produces a statistically significantly better fit of the model to the data

Estimation-VIII



Estimation-IX

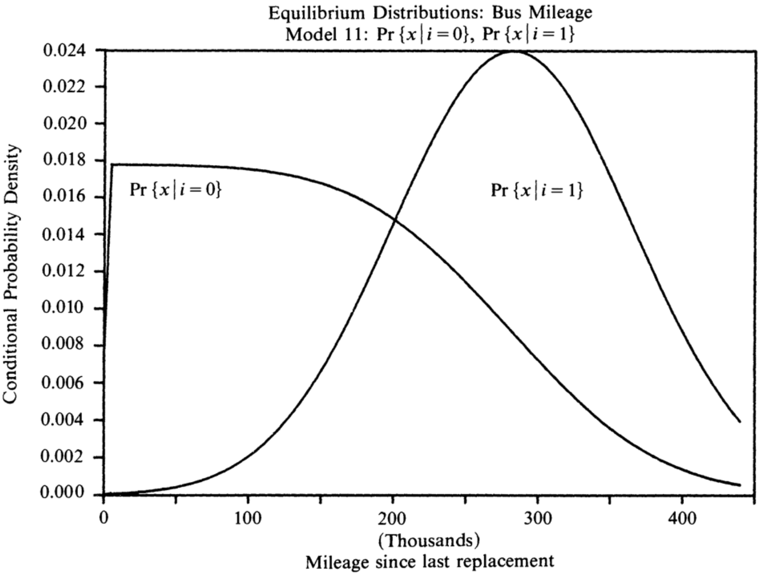


FIGURE 6

Takeaways

- So what? Why not just do a logit?

TAKEAWAYS

- So what? Why not just do a logit?
- Rust estimates Harold's problem at a micro level

Takeaways

- So what? Why not just do a logit?
- Rust estimates Harold's problem at a micro level
- Changes in interest rates, costs, etc. he can still deal with *even if no variation in data!*

Takeaways

- So what? Why not just do a logit?
- Rust estimates Harold's problem at a micro level
- Changes in interest rates, costs, etc. he can still deal with *even if no variation in data!*
- Not myopic

Takeaways

- So what? Why not just do a logit?
- Rust estimates Harold's problem at a micro level
- Changes in interest rates, costs, etc. he can still deal with *even if no variation in data!*
- Not myopic
- Good example of indirectly inferring underlying parameters from agent behavior

Counterfactuals

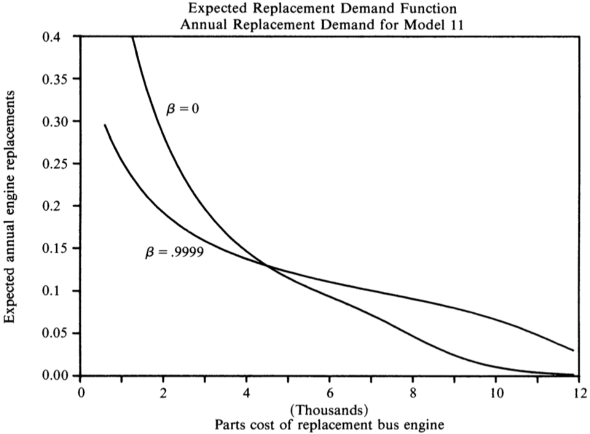


Figure 7

Because parts costs haven't changed, all **reduced form** estimates would be clustered around the intersection: unable to predict counterfactual reliably (not true of structural method)

A better way? MPEC

- Ordinarily, we have error function: $g(\theta; X; V)$ and value function: $Vnew(\theta) = \max\{R + \beta Vold\}$.
- Fixed point algorithm:
  1. Given θ, solve for V (inner loop, many iterations)
  2. Evaluate $g(\theta; X; V)$, find new θ to minimize g (outer loop, needs precise inner loop)
- MPEC: put the inner loop in as a *constraint*

$$\min_{\theta, V} \{g(\theta; X; V)\} \quad s.t. Vnew(\theta) = \max\{R + \beta Vold\}$$

- Idea: dont' need solve precisely for inner loop on bad θ's.
- Much faster...3-800 times faster
- Constraints could instead be equilibrium
  - BLP: inner loop: (shares(demand shocks(θ)), outer loop: choose θ to match moments

Takeaways

- Rust's method is a dynamic form of what we've been saying all semester
- Most problems we write down have a natural fixed point in their estimation
- Guess estimated parameters, can solve your model, simulate your model, calculate likelihoods from your model
- With the feedback (errors, likelihood) can change parameters to minimize error, maximize likelihood, etc.
- Rust gives a now-canonical example of how to do that in a dynamic discrete choice framework