

# Structural Estimation of Dynamic Discrete Choice Models

The Nested Fixed Point Algorithm (NFXP)

Bertel Schjerning, University of Copenhagen

Dynamic Programming and Structural Econometrics #6

# The Nested Fixed Point Algorithm (NFXP)

Rust (ECTA, 1987):

OPTIMAL REPLACEMENT OF GMC BUS ENGINES:  
AN EMPIRICAL MODEL OF HAROLD ZURCHER



Harold Alois Zuercher June 16, 1926 - June 21, 2020 (age 94)

# Overview of Rust (1987)

**This is a path-breaking paper** that introduces a methodology to estimate a single-agent dynamic discrete choice (DDC) models.

## Main contributions

1. An illustrative application in a simple model of engine replacement.
2. Development and implementation of [Nested Fixed Point Algorithm](#)
3. Formulation of assumptions that makes DDC models tractable
4. The first researcher to obtain ML estimates of DDC models
5. Bottom-up approach: Micro-aggregated demand for durable assets

## Policy experiments:

- ▶ How does changes in replacement cost affect the demand for engines and the equilibrium distribution of mileage?

# Who cares about Harold Zurcher?

- ▶ Occupational Choice (Keane and Wolpin, JPE 1997)
- ▶ Retirement (Rust and Phelan, ECMA 1997)
- ▶ Brand choice and advertising (Erdem and Keane, MaScience 1996)
- ▶ Choice of college major (Arcidiacono, JoE 2004)
- ▶ Individual migration decisions (Kennan and Walker, ECMA 2011)
- ▶ High school attendance and work decisions (Eckstein and Wolpin, ECMA 1999)
- ▶ Sales and dynamics of consumer inventory behavior (Hendel and Nevo, ECMA 2006)
- ▶ Advertising, learning, and consumer choice in experience good markets (Akerberg, IER 2003)
- ▶ Route choice models (Fosgerau et al, Transp. Res. B)
- ▶ Fertility and labor supply decisions (Francesconi, JoLE 2002)
- ▶ Residential and Work-location choice (Buchinsky et al, ECMA 2015)
- ▶ Equilibrium Allocations Under Alternative Waitlist Designs: Evidence From Deceased Donor Kidneys (Argarwal et al, ECMA 2021)
- ▶ Equilibrium Trade in Automobiles (Gillingham et al, JPE 2022)
- ▶ ...and many more

# Who cares about Harold Zurcher?

- ▶ Occupational Choice (Keane and Wolpin, JPE 1997)
- ▶ Retirement (Rust and Phelan, ECMA 1997)
- ▶ Brand choice and advertising (Erdem and Keane, MaScience 1996)
- ▶ Choice of college major (Arcidiacono, JoE 2004)
- ▶ Individual migration decisions (Kennan and Walker, ECMA 2011)
- ▶ High school attendance and work decisions (Eckstein and Wolpin, ECMA 1999)
- ▶ Sales and dynamics of consumer inventory behavior (Hendel and Nevo, ECMA 2006)
- ▶ Advertising, learning, and consumer choice in experience good markets (Akerberg, IER 2003)
- ▶ Route choice models (Fosgerau et al, Transp. Res. B)
- ▶ Fertility and labor supply decisions (Francesconi, JoLE 2002)
- ▶ Residential and Work-location choice (Buchinsky et al, ECMA 2015)
- ▶ **Equilibrium Allocations Under Alternative Waitlist Designs: Evidence From Deceased Donor Kidneys** (Argarwal et al, ECMA 2021)
- ▶ **Equilibrium Trade in Automobiles** (Gillingham et al, JPE 2022)
- ▶ ...and many more

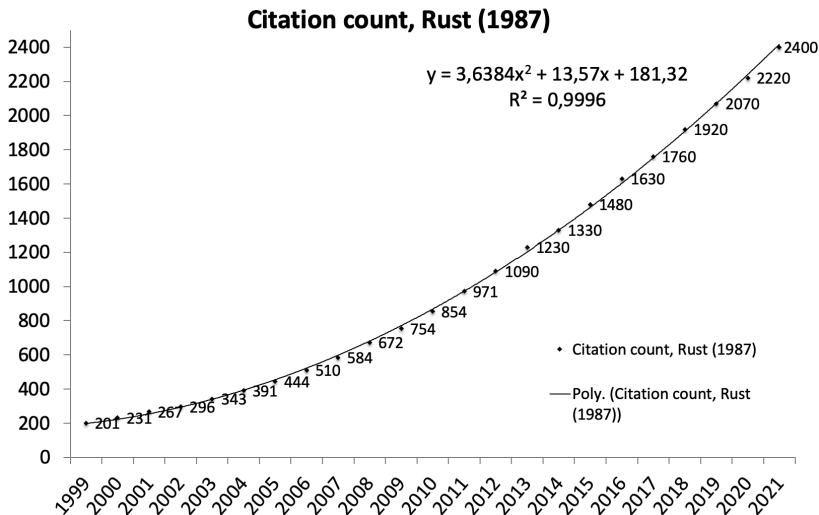
# Who cares about Harold Zurcher?

- ▶ Occupational Choice (Keane and Wolpin, JPE 1997)
- ▶ Retirement (Rust and Phelan, ECMA 1997)
- ▶ Brand choice and advertising (Erdem and Keane, MaScience 1996)
- ▶ Choice of college major (Arcidiacono, JoE 2004)
- ▶ Individual migration decisions (Kennan and Walker, ECMA 2011)
- ▶ High school attendance and work decisions (Eckstein and Wolpin, ECMA 1999)
- ▶ Sales and dynamics of consumer inventory behavior (Hendel and Nevo, ECMA 2006)
- ▶ Advertising, learning, and consumer choice in experience good markets (Akerberg, IER 2003)
- ▶ Route choice models (Fosgerau et al, Transp. Res. B)
- ▶ Fertility and labor supply decisions (Francesconi, JoLE 2002)
- ▶ Residential and Work-location choice (Buchinsky et al, ECMA 2015)
- ▶ **Equilibrium Allocations Under Alternative Waitlist Designs: Evidence From Deceased Donor Kidneys** (Argarwal et al, ECMA 2021)
- ▶ **Equilibrium Trade in Automobiles** (Gillingham et al, JPE 2022)
- ▶ ...and many more

# Who cares about Harold Zurcher?

- ▶ Occupational Choice (Keane and Wolpin, JPE 1997)
- ▶ Retirement (Rust and Phelan, ECMA 1997)
- ▶ Brand choice and advertising (Erdem and Keane, MaScience 1996)
- ▶ Choice of college major (Arcidiacono, JoE 2004)
- ▶ Individual migration decisions (Kennan and Walker, ECMA 2011)
- ▶ High school attendance and work decisions (Eckstein and Wolpin, ECMA 1999)
- ▶ Sales and dynamics of consumer inventory behavior (Hendel and Nevo, ECMA 2006)
- ▶ Advertising, learning, and consumer choice in experience good markets (Akerberg, IER 2003)
- ▶ Route choice models (Fosgerau et al, Transp. Res. B)
- ▶ Fertility and labor supply decisions (Francesconi, JoLE 2002)
- ▶ Residential and Work-location choice (Buchinsky et al, ECMA 2015)
- ▶ **Equilibrium Allocations Under Alternative Waitlist Designs: Evidence From Deceased Donor Kidneys** (Argarwal et al, ECMA 2021)
- ▶ **Equilibrium Trade in Automobiles** (Gillingham et al, JPE 2022)
- ▶ ...and many more

# Big Mac Index of Dynamic Structural Econometrics





# Methods for estimating Dynamic Discrete Choice Models

- ▶ Rust (1987): MLE using 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 (breaks curse of dimensionality)
- ▶ Aguirregabiria and Mira (2002): Nested Pseudo Likelihood (NPL).
- ▶ Bajari, Benkard and Levin (2007): Two step-minimum distance (equilibrium inequalities).
- ▶ Arcidiacono Miller (2002): CCP with unobserved heterogeneity (EM Algorithm).
- ▶ Norets (2009): Bayesian Estimation (allows for serial correlation in  $\epsilon$ )
- ▶ Su and Judd (2012): MLE using constrained optimization (MPEC)
- ▶ and MUCH more
- ▶ Any estimator method or solution algorithm of DDC models must confront NFXP and Harold Zurcher

# Methods for estimating Dynamic Discrete Choice Models

- ▶ Rust (1987): MLE using 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 (breaks curse of dimensionality)
- ▶ Aguirregabiria and Mira (2002): Nested Pseudo Likelihood (NPL).
- ▶ Bajari, Benkard and Levin (2007): Two step-minimum distance (equilibrium inequalities).
- ▶ Arcidiacono Miller (2002): CCP with unobserved heterogeneity (EM Algorithm).
- ▶ Norets (2009): Bayesian Estimation (allows for serial correlation in  $\epsilon$ )
- ▶ Su and Judd (2012): MLE using constrained optimization (MPEC)
- ▶ and MUCH more
- ▶ Any estimator method or solution algorithm of DDC models must confront NFXP and Harold Zurcher

# Formulating, solving and estimating a dynamic model



## Components of the dynamic model

- ▶ **Decision variables:** vector describing the choices,  $d_t \in C(s_t)$
- ▶ **State variables:** vector of variables,  $s_t$ , that describe all relevant information about the modeled decision process
- ▶ **Instantaneous payoff:** utility function,  $u(s_t, d_t)$ , with time separable discounted utility
- ▶ **Motion rules:** agent's beliefs of how state variable evolve through time, conditional on states and choices. Here formalized by a Markov transition density  $p(s_{t+1} | s_t, d_t)$

Solution is given by:

- ▶ **Value function:** maximum attainable utility  $V(s_t)$
- ▶ **Policy function:** mapping from **state space** to **action space** that returns the optimal choice,  $d^*(s_t)$

## Structural Estimation

- ▶ **Parametrize model:** utility function  $u(s_t, d_t; \theta_u)$ , motion rules for states  $p(s_{t+1} | s_t, d_t; \theta_p)$ , choice sets  $C(s_t; \theta_c)$ , etc.
- ▶ Search for (policy invariant) parameters  $\theta$  so that model **fits** targeted aspects of data on (a subset of) **decisions, states, payoff's, etc.**

# Zurcher's Bus Engine Replacement Problem

- ▶ **Choice set:** Binary choice set,  $C(x_t) = \{0, 1\}$ .
  - ▶ Engine replacement ( $d_t = 1$ ) or ordinary maintenance ( $d_t = 0$ )
- ▶ **State variables:** Harold Zurcher observes  $s_t = (x_t, \varepsilon_t)$ :
  - ▶  $x_t$ : mileage at time  $t$  since last engine overhaul/replacement
  - ▶  $\varepsilon_t = [\varepsilon_t(d_t = 0), \varepsilon_t(d_t = 1)]$ : decision specific state variable
- ▶ **Utility function:**  $U(x_t, \varepsilon_t, d_t; \theta_1) =$

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

- ▶ **State variables process**

- ▶  $\varepsilon_t$  is iid with conditional density  $q(\varepsilon_t | x_t, \theta_2)$
- ▶  $x_t$  (mileage since last replacement)

$$p(x_{t+1} | x_t, d_t, \theta_2) = \begin{cases} g(x_{t+1} - 0, \theta_3) & \text{if } d_t = 1 \\ g(x_{t+1} - x_t, \theta_3) & \text{if } d_t = 0 \end{cases} \quad (2)$$

If engine is replaced, state of bus regenerates to  $x_t = 0$ .

- ▶ **Parameters to be estimated**  $\theta = (RC, \theta_1, \theta_3)$   
(Fixed parameters:  $(\beta, \theta_2)$ )

# General Behavioral Framework

## The decision problem

- ▶ The decision maker chooses a sequence of actions to maximize expected discounted utility over a (in)finite horizon

$$V_{\theta}(s_t) = \sup_{\Pi} E \left[ \sum_{j=0}^T \beta^j U(s_{t+j}, d_{t+j}; \theta_1) | s_t, d_t \right]$$

where

- ▶  $\Pi = (f_t, f_{t+1}, \dots), d_t = f_t(s_t, \theta) \in C(x_t) = \{1, 2, \dots, J\}$
- ▶  $\beta \in (0, 1)$  is the discount factor
- ▶  $U(s_t, d_t; \theta_1)$  is a choice and state specific utility function
- ▶ We may consider an infinite horizon, i.e.  $T = \infty$
- ▶  $E$  summarizes expectations of future states given  $s_t$  and  $d_t$

# Recursive form of the maximization problem

- ▶ By Bellman Principle of Optimality, the value function  $V(s)$  constitutes the solution of the following functional (Bellman) equation

$$V(x, \varepsilon) \equiv T(V)(x, \varepsilon) = \max_{d \in C(x)} \{u(x, \varepsilon, d) + \beta E[V(x', \varepsilon') | x, \varepsilon, d]\}$$

- ▶ Expectations are taken over the next period values of state  $s' = (x', \varepsilon')$  given it's controlled motion rule,  $p(s' | s, d)$

$$E[V(x', \varepsilon') | x, \varepsilon, d] = \int_X \int_{\Omega} V(x', \varepsilon') p(x', \varepsilon' | x, \varepsilon, d) dx' d\varepsilon'$$

where  $\varepsilon = (\varepsilon(1), \dots, \varepsilon(J)) \in \mathbb{R}^J$

Hard to compute fixed point  $V$  such that  $T(V) = V$

- ▶  $x$  is continuous and  $\varepsilon$  is continuous and  $J$ -dimensional
- ▶  $V(x, \varepsilon)$  is high dimensional
- ▶ Evaluating  $E$  may require high dimensional integration
- ▶ Evaluating  $V(x', \varepsilon')$  may require high dimensional interpolation/approximation
- ▶  $V(x, \varepsilon)$  is non-differentiable

# Rust's Assumptions

1. Additive separability in preferences (**AS**):

$$U(s_t, d) = u(x_t, d; \theta_1) + \varepsilon_t(d)$$

2. Conditional independence (**CI**):

State variables,  $s_t = (x_t, \varepsilon_t)$  obeys a (conditional independent) controlled Markov process with probability density

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

3. Extreme value Type I (EV1) distribution of  $\varepsilon$  (**EV**)

Each of the choice specific state variables,  $\varepsilon_t(d)$  are assumed to be iid. extreme value distributed with CDF

$$F(\varepsilon_t(d); \mu, \lambda) = \exp(-\exp(-(\varepsilon_t(d) - \mu)/\lambda)) \text{ for } \varepsilon_t(d) \in \mathbb{R}$$

with  $\mu = 0$  and  $\lambda = 1$

## Rust's Assumptions simplifies DP problem

$$V(x, \varepsilon) = \max_{d \in C(x)} \{u(x, d) + \varepsilon(d) + \beta \int_X \int_{\Omega} V(x', \varepsilon') p(x'|x, d) q(\varepsilon'|x') dx' d\varepsilon'\}$$

1. Separate out the deterministic part of choice specific value  $v(x, d)$  (assumptions SA and CI)
2. Reformulate Bellman equation on reduced state space (assumption CI)
3. Compute the expectation of maximum using properties of EV1 (assumption EV)



# 1. DP problem under AS and CI

Separate out the deterministic part of choice specific value  $v(x, d)$

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

So that

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



where

$$v(x, d) = u(x, d) + \beta E[V(x', \varepsilon') | x, d]$$

## 2a. Bellman equation in expected value function space

Let  $EV(x, d) = E[V(x', \varepsilon') | x, d]$  denote the expected value function.

Because of CI we can now express the Bellman equation in expected value function space

$$EV(x, d) = \Gamma(EV)(x, d) \equiv \int_X \int_{\Omega} [V(x', \varepsilon') q(\varepsilon' | x') d\varepsilon'] p(x' | x, d) dx'$$

where

$$V(x', \varepsilon') = \max_{d' \in C(x')} [u(x', d') + \beta EV(x', d') + \varepsilon'(d')]$$

- ▶  $\Gamma$  is a contraction mapping with unique fixed point  $EV$ , i.e.  
 $\|\Gamma(EV) - \Gamma(W)\| \leq \beta \|EV - W\|$
- ▶ Global convergence of VFI
- ▶  $EV(x, d)$  is lower dimensional: does not depend on  $\varepsilon$

## 2b. Bellman equation in integrated value function space

Let  $\bar{V}(x) = E[V(x, \varepsilon)|x]$  denote the integrated value function

Because of CI we can express Bellman equation in integrated value function space

$$\bar{V}(x) = \bar{\Gamma}(\bar{V})(x) \equiv \int_{\Omega} V(x, \varepsilon) q(\varepsilon|x) d\varepsilon$$

where

$$V(x, \varepsilon) = \max_{d \in C(x)} [u(x, d) + \varepsilon(d) + \beta \int_X \bar{V}(x') p(x'|x, d) dx']$$

- ▶  $\bar{\Gamma}$  is a contraction mapping with unique fixed point  $\bar{V}$ , i.e.  
 $\|\bar{\Gamma}(\bar{V}) - \bar{\Gamma}(W)\| \leq \beta \|\bar{V} - W\|$
- ▶ Global convergence of VFI
- ▶  $\bar{V}(x)$  is lower dimensional: does not depend on  $\varepsilon$  and  $d$

### 3. Compute the expectation of maximum under EV

We can express expectation of maximum using properties of EV1 distribution (assumption EV)

Expectation of maximum,  $\bar{V}(x)$ , can be expressed as "the log-sum"

$$\bar{V}(x) = E \left[ \max_{d \in \{1, \dots, J\}} \{v(x, d) + \lambda \varepsilon(d)\} \mid x \right] = \lambda \log \sum_{j=1}^J \exp(v(x, j)/\lambda)$$



Conditional choice probability,  $P(x, d)$  has closed form logit expression

$$\begin{aligned} P(d \mid x) &= E \left[ \mathbb{1} \left\{ d = \arg \max_{j \in \{1, \dots, J\}} \{v(x, j) + \lambda \varepsilon(j)\} \right\} \mid x \right] \\ &= \frac{\exp(v(x, d)/\lambda)}{\sum_{j=1}^J \exp(v(x, j)/\lambda)} \end{aligned}$$

HUGE benefits

- ▶ Avoids  $J$  dimensional numerical integration over  $\varepsilon$
- ▶  $P(d \mid x)$ ,  $\bar{V}(x)$  and  $EV(x, d)$  are smooth functions.



# The DP problem under AS, CI and EV

Putting all this together

- ▶ Conditional Choice Probabilities (CCPs) are given by

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

- ▶ The expected value function can be found as the unique fixed point to the contraction mapping  $\Gamma_\theta$ , defined by

$$\begin{aligned} EV_\theta(x, d) &= \Gamma_\theta(EV_\theta)(x, d) \\ &= \int_y \ln \left[ \sum_{d' \in C(y)} \exp [u(y, d'; \theta_1) + \beta EV_\theta(y, d')] \right] \\ &\quad p(dy|x, d, \theta_2) \end{aligned}$$

- ▶ We have used the subscript  $\theta$  to emphasize that the Bellman operator,  $\Gamma_\theta$  depends on the parameters.
- ▶ In turn, the fixed point,  $EV_\theta$ , and the resulting CCPs,  $P(d|x, \theta)$  are implicit functions of the parameters we wish to estimate.

# How to deal with continuous mileage state?

Rust discretize the mileage state space  $x$  into  $n$  grid points

$$X = \{x_1, \dots, x_n\} \text{ with } x_1 = 0$$


Mileage transition probability: for  $l = 0, \dots, L$

$$p(x' | \hat{x}_k, d, \theta_2) = \begin{cases} Pr\{x' = x_{k+l} | \theta_2\} = \pi_l & \text{if } d = 0 \\ Pr\{x' = x_{1+l} | \theta_2\} = \pi_l & \text{if } d = 1 \end{cases}$$

- ▶ where  $\theta_2 = [\pi_1, \dots, \pi_L]$ ,  $\pi_0 = 1 - \sum_{l=1}^L \pi_l$ , and  $\pi_l \geq 0$
- ▶ Mileage in the next period  $x'$  can move up at most  $L$  grid points.
- ▶  $L$  is determined by the empirical distribution of mileage.

# Transition matrix for mileage is sparse

Transition matrix conditional on keeping engine


$$\Pi(d = \text{keep})_{n \times n} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & & & & \pi_0 & \pi_1 & \pi_2 & 0 \\ 0 & & & & & \pi_0 & \pi_1 & \pi_2 \\ 0 & & & & & & \pi_0 & 1 - \pi_0 \\ 0 & 0 & & & & & & 1 \end{pmatrix}$$

# Transition matrix for mileage is sparse

Transition matrix conditional on replacing engine

$$\Pi(d = \text{replace})_{n \times n} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix}$$



# Bellman equation in matrix form

Bellman equation in **expected value** function space

$$EV(d) = \Gamma(EV) = \Pi(d) \ln \left[ \sum_{d'} \exp[u(d') + \beta EV(d')] \right]$$

Bellman equation in **integrated value** function space

$$\bar{V} = \bar{\Gamma}(\bar{V}) = \ln \left[ \sum_{d'} \exp[u(d') + \beta \Pi(d') \bar{V}] \right]$$

where

- ▶  $u(d) = [u(x_1, d), \dots, u(x_n, d)]$
- ▶  $EV(d) = [EV(x_1, d), \dots, EV(x_n, d)]$
- ▶  $\bar{V} = [\bar{V}(x_1), \dots, \bar{V}(x_n)]$
- ▶  $\Pi(d)$  is a  $n \times n$  state transition matrix conditional on decision  $d$

# Structural Estimation

Data:  $(d_{i,t}, x_{i,t})$ ,  $t = 1, \dots, T_i$  and  $i = 1, \dots, N$

Log likelihood function

$$L(\theta, EV_\theta) = \sum_{i=1}^N \ell_i^f(\theta, EV_\theta)$$

$$\ell_i^f(\theta, EV_\theta) = \sum_{t=2}^{T_i} \log(P(d_{i,t}|x_{i,t}, \theta)) + \sum_{t=2}^{T_i} \log(p(x_{i,t}|x_{i,t-1}, d_{i,t-1}, \theta_3))$$

where

$$P(d|x, \theta) = \frac{\exp\{u(x, d, \theta_1) + \beta EV_\theta(x, d)\}}{\sum_{d' \in \{0,1\}} \{u(x, d', \theta_1) + \beta EV_\theta(x, d')\}}$$

and

$$\begin{aligned} EV_\theta(x, d) &= \Gamma_\theta(EV_\theta)(x, d) \\ &= \int_y \ln \left[ \sum_{d' \in \{0,1\}} \exp[u(y, d'; \theta_1) + \beta EV_\theta(y, d')] \right] p(dy|x, d, \theta_3) \end{aligned}$$

# The Nested Fixed Point Algorithm

Since the contraction mapping  $\Gamma_\theta$  always has a unique fixed point, the constraint  $EV_\theta = \Gamma(EV_\theta)$  implies that the fixed point  $EV_\theta$  is an implicit function of  $\theta$ .

Hence, NFXP solves the unconstrained optimization problem

$$\max_{\theta} L(\theta, EV_\theta)$$

Outer loop (Hill-climbing algorithm):

- ▶ Likelihood function  $L(\theta, EV_\theta)$  is maximized w.r.t.  $\theta$
- ▶ Quasi-Newton algorithm: Usually BHHH, BFGS or a combination.
- ▶ Each evaluation of  $L(\theta, EV_\theta)$  requires solution of  $EV_\theta$

Inner loop (fixed point algorithm):

The implicit function  $EV_\theta$  defined by  $EV_\theta = \Gamma(EV_\theta)$  is solved by:

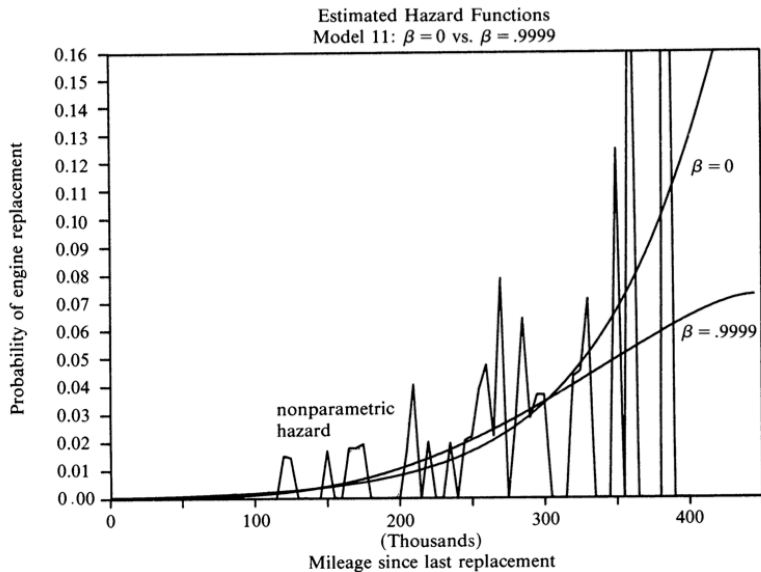
- ▶ Successive Approximations (SA)
- ▶ Newton-Kantorovich (NK) Iterat



# Data

- ▶ Harold Zurcher's Maintenance records of 162 busses
- ▶ Monthly observations of mileage on each bus (odometer reading)
- ▶ Data on maintenance replacement decisions

# Estimated Hazard Functions



# Structural Estimates, n=90

TABLE IX  
STRUCTURAL ESTIMATES FOR COST FUNCTION  $c(x, \theta_1) = .001\theta_{11}x$   
FIXED POINT DIMENSION = 90  
(Standard errors in parentheses)

Parameter		Data Sample			Heterogeneity Test	
Discount Factor	Estimates/ Log-Likelihood	Groups 1, 2, 3 3864 Observations	Group 4 4292 Observations	Groups 1, 2, 3, 4 8156 Observations	LR Statistic ( $df = 4$ )	Marginal Significance Level
$\beta = .9999$	RC	11.7270 (2.602)	10.0750 (1.582)	9.7558 (1.227)	85.46	1.2E-17
	$\theta_{11}$	4.8259 (1.792)	2.2930 (0.639)	2.6275 (0.618)		
	$\theta_{30}$	.3010 (.0074)	.3919 (.0075)	.3489 (.0052)		
	$\theta_{31}$	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	LL	-2708.366	-3304.155	-6055.250		
$\beta = 0$	RC	8.2985 (1.0417)	7.6358 (0.7197)	7.3055 (0.5067)	89.73	1.5E-18
	$\theta_{11}$	109.9031 (26.163)	71.5133 (13.778)	70.2769 (10.750)		
	$\theta_{30}$	.3010 (.0074)	.3919 (.0075)	.3488 (.0052)		
	$\theta_{31}$	.6884 (.0075)	.5953 (.0075)	.6394 (.0053)		
	LL	-2710.746	-3306.028	-6061.641		
Myopia test:	LR Statistic ( $df = 1$ )	4.760	3.746	12.782		
$\beta = 0$ vs. $\beta = .9999$	Marginal Significance Level	0.0292	0.0529	0.0035		

# Structural Estimates, n=175

TABLE X  
STRUCTURAL ESTIMATES FOR COST FUNCTION  $c(x, \theta_1) = .001\theta_{11}x$   
FIXED POINT DIMENSION = 175  
(Standard errors in parentheses)

Parameter		Data Sample			Heterogeneity Test	
Discount Factor	Estimates Log-Likelihood	Groups 1, 2, 3 3864 Observations	Group 4 4292 Observations	Groups 1, 2, 3, 4 8156 Observations	LR Statistic (df = 6)	Marginal Significance Level
$\beta = .9999$	RC	11.7257 (2.597)	10.896 (1.581)	9.7687 (1.226)	237.53	1.89E - 48
	$\theta_{11}$	2.4569 (.9122)	1.1732 (0.327)	1.3428 (0.315)		
	$\theta_{30}$	.0937 (.0047)	.1191 (.0050)	.1071 (.0034)		
	$\theta_{31}$	.4475 (.0080)	.5762 (.0075)	.5152 (.0055)		
	$\theta_{32}$	.4459 (.0080)	.2868 (.0069)	.3621 (.0053)		
	$\theta_{33}$	.0127 (.0018)	.0158 (.0019)	.0143 (.0013)		
	LL	-3993.991	-4495.135	-8607.889		
$\beta = 0$	RC	8.2969 (1.0477)	7.6423 (.7204)	7.3113 (0.5073)	241.78	2.34E - 49
	$\theta_{11}$	56.1656 (13.4205)	36.6692 (7.0675)	36.0175 (5.5145)		
	$\theta_{30}$	.0937 (.0047)	.1191 (.0050)	.1070 (.0034)		
	$\theta_{31}$	.4475 (.0080)	.5762 (.0075)	.5152 (.0055)		
	$\theta_{32}$	.4459 (.0080)	.2868 (.0069)	.3622 (.0053)		
	$\theta_{33}$	.0127 (.0018)	.0158 (.0019)	.0143 (.0143)		
	LL	-3996.353	-4496.997	-8614.238		
Myopia tests:	LR Statistic (df = 1)	4.724	3.724	12.698		
$\beta = 0$ vs. $\beta = .9999$	Marginal Significance Level	0.0297	0.0536	.00037		

# MATLAB implementation, n=175, (replication of Table X)

Output from run\_busdata.m:

```
fzp386 — cefhelper (Renderer) • MATLAB_maci64 -nodesktop — 89x35

>> run_busdata
Structural Estimation using busdata from Rust(1987)
Bustypes      = [ 1  2  3  4 ]
Beta          = 0.99990
n             = 175.00000
Sample size   = 8156.00000

Method nfxp (pmle)
Param.        Estimates      s.e.      t-stat
-----
RC            9.7910         1.2684     7.7190
c             1.3486         0.3458     3.8996

log-likelihood = -300.57017
runtime (seconds) = 0.07795
g'*inv(h)*g    = 2.65552e-09

Method nfxp (mle)
Param.        Estimates      s.e.      t-stat
-----
RC            9.7915         1.2689     7.7168
c             1.3488         0.3460     3.8982
p             0.1070         0.0034     31.2111
p             0.5152         0.0055     93.0533
p             0.3622         0.0053     68.0413
p             0.0143         0.0013     10.8947
p             0.0009         0.0003     2.6469

log-likelihood = -8607.88844
runtime (seconds) = 0.07484
g'*inv(h)*g    = 7.26854e-09
>>
```



# Equilibrium bus mileage and demand for engines

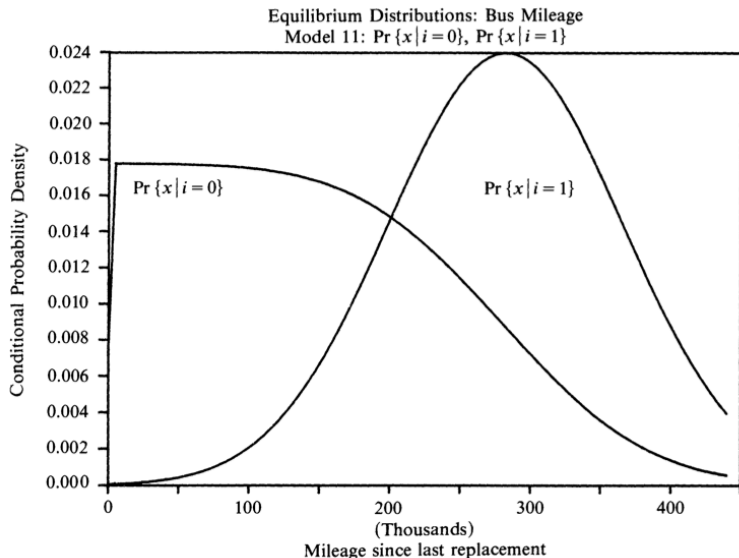
- ▶ Let  $\pi$  be the long run stationary (or equilibrium) distribution of the controlled process  $\{i_t, x_t\}$
- ▶  $\pi$  is then given by the unique solution to the functional equation

$$\pi(x, i) = \int_y \int_j P(i|x, \theta) p(x|y, j, \theta_3) \pi(dy, dj)$$

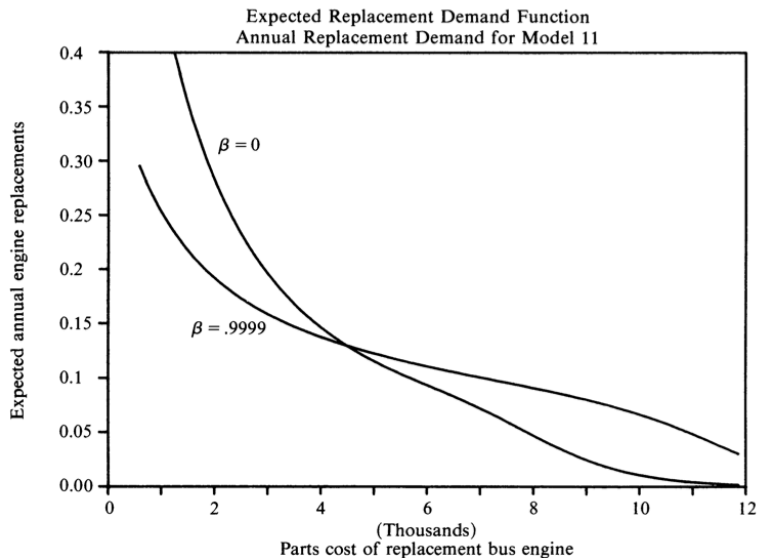
- ▶  $\pi$  is the ergodic distribution of the controlled state transition matrix
- ▶ Clearly the equilibrium distribution of  $\pi$  is an implicit function of the structural parameters  $\theta$ , which we emphasize by the notation  $\pi_\theta$
- ▶ Given  $\pi_\theta$ , we can also obtain the following simple formula for annual equilibrium demand for engines as a function of  $RC$

$$d(RC) = 12M \int_0^\infty \pi_\theta(dx, 1)$$

# Equilibrium Bus mileage, bus group 4



## Demand Function, bus group 4



# Why not a reduced form for demand?

## Reduced form

- ▶ Regress engine replacements on replacement costs

Problem: Lack of variation in replacement costs

- ▶ Data would be clustered around the intersection of the demand curves for  $\beta = 0$  and  $\beta = 0.9999$   
(both models predict that  $RC$  is around the actual  $RC$  of \$4343)
- ▶ Demand also depends on how operating costs varies with mileage
- ▶ Need exogenous variation in  $RC$   
.... that doesn't vary with operating costs
- ▶ Even if we had exogenous variation, this does not help us to understand the underlying economic incentives

# Structural Approach

## Attractive features

- ▶ Structural parameters have a transparent interpretation
- ▶ Evaluation of (new) policy proposals by counterfactual simulations.
- ▶ Economic theories can be tested directly against each other.
- ▶ Economic assumptions are more transparent and explicit (compared to statistical assumptions)

## Less attractive features

- ▶ We impose more structure and make more assumptions
- ▶ Truly “structural” (policy invariant) parameters may not exist
- ▶ The curse of dimensionality
- ▶ The identification problem
- ▶ The problem of multiplicity and indeterminacy of equilibria
- ▶ Intellectually demanding and a huge amount of work