The Nested Fixed Point Algorithm

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

Dynamic Programming and Structural Econometrics #6

Bertel Schjerning, University of Copenhagen

Structural Estimation in Microeconomics

Some methods for solving 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 ϵ)
- Su and Judd (2012): MLE using constrained optimization (MPEC)
- ..... and MUCH more

- NFXP is still the Swiss-army knife for structural estimation of dynamic structural choice models
- Any estimator method or solution algorithm of DDC models must confront *Harold Zurcher*

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)

Rust (1987)

This is a path-breaking paper that introduces a methodology to estimate a single-agent dynamic discrete choice 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 dynamic discrete choice models tractable.
4. The first researcher to obtain ML estimates of discrete choice dynamic programming models
5. Bottom-up approach: Micro-aggregated demand for durable assets

Policy experiments:

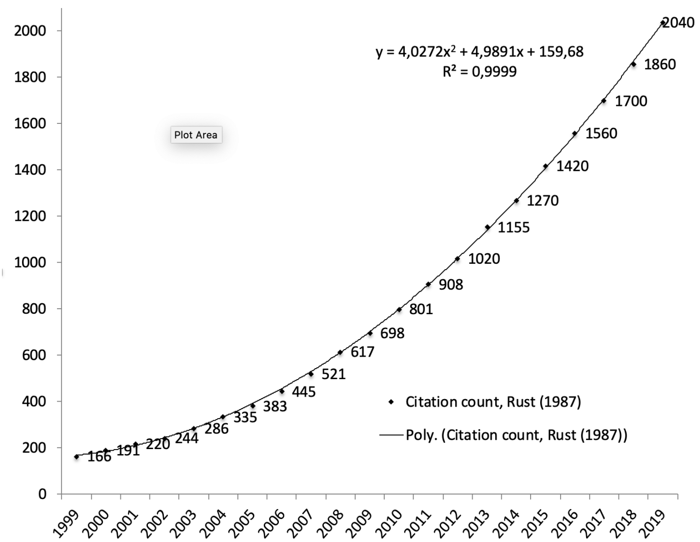
- How does changes in replacement cost affect?
  - the distribution of mileage
  - the demand for engines

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 (Ackerberg, IER 2003)
- Route choice models (Fosgerau et al, Transp. Res. B)
- Fertility and labor supply decisions (Francesconi, JoLE 2002)
- Car ownership, type choice and use (Gillingham et al, WP)
- Residential and Work-location choice (Carstensen et al, WP)
- ...and many more (2355 cites, Nov 2021)

Big Mac Index for Dynamic Structural Econometrics

Citation count for Rust (1987)



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 $p(s_{t+1} \mid s_t, d_t)$

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, $d^\star(s_t)$

Structural Estimation

- Parametrize model: utility function $u(s_t, d_t; \theta_u)$, motion rules for states $p(s_{t+1} \mid s_t, d_t; \theta_p)$, choice sets $C(s_t; \theta_c)$, etc.
- Search for (*policy invariant*) parameters θ 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, $C(x_t) = \{0, 1\}$. 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 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: $U(x_t, \varepsilon_t, d_t; \theta_1) =$

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

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

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

  If engine is replaced, state of bus regenerates to $x_t = 0$.
- Parameters to be estimated $\theta = (RC, \theta_1, \theta_3)$
  (Fixed parameters: (β, θ_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

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

where

- $\Pi = \left(f_t, f_{t+1}, .., \right), d_t = f_t\left(s_t, \theta\right) \in C\left(x_t\right) = \{1, 2, .., J\}$
- $\beta \in (0, 1)$ is the discount factor
- $U\left(s_t, d_t; \theta_1\right)$ 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

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

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

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

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

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

- x is continuous and ε 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 ε (**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

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

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)

DP problem under AS and CI

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

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

So that

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

where

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

Bellman equation in expected value function space

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

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

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

where

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

- Γ 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 ε

Bellman equation in integrated value function space

Let $\bar{V}(x) = E\big[V(x,\varepsilon)\big|x\big]$ 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. $\left\|\bar{\Gamma}\left(\bar{V}\right) - \bar{\Gamma}\left(W\right)\right\| \leq \beta \left\|\bar{V} - W\right\|$
- Global convergence of VFI
- $\bar{V}(x)$ is lower dimensional: does not depend on ε and d

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"

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

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 ε
- $P(d \mid x)$, $\bar{V}(x)$ 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

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

- The expected value function can be found as the unique fixed point to the contraction mapping Γ_θ, defined by

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

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

Mileage is continuous. How to deal with continuous state?

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 $j = 1, ..., J$

$$p(x'|\hat{x}_k, d, \theta_2) = \begin{cases} Pr\{x' = \hat{x}_{k+j}|\theta_3\} = \theta_{3j} \text{ if } d = 0 \\ Pr\{x' = \hat{x}_{1+j}|\theta_3\} = \theta_{3j} \text{ if } d = 1 \end{cases}$$

- Mileage in the next period x' can move up at most J grid points
- J is determined by the distribution of mileage

Choice-specific expected value function for $\hat{x} \in \hat{X}$

$$EV_\theta(\hat{x}, d) = \hat{\Gamma}_\theta(EV_\theta)(\hat{x}, d)$$

$$= \sum_j^J \ln \left[\sum_{d' \in D(y)} \exp[u(x', d'; \theta_1) + \beta EV_\theta(x', d')] \right] p(x'|\hat{x}, d, \theta_2)$$

Bellman equation in matrix form (expected value)

The choice specific expected value function can be found as fixed point on the Bellman operator

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

where

$EV(d) = [EV(1, d), .., EV(n, d)]$ and $u(d) = [u(1, d), .., u(n, d)]$

$\Pi(d)$ is a $n \times n$ state transition matrix conditional on decision d

Bellman equation in matrix form (integrated value)

The choice integrated value function can be found as fixed point on the Bellman operator

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

where

$\bar{V} = [\bar{V}(1), .., \bar{V}(n)]$ and $u(d) = [u(1, d), .., u(n, d)]$

$EV(d) = \Pi(d)\bar{V}$

$\Pi(d)$ is the $n \times n$ state transition matrix conditional on decision d

Transition matrix for mileage, $d = 0$

If not replacing ($d = 0$)

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

Transition matrix for mileage, $d = 1$

If replacing ($d = 1$)

$$\Pi(d=1)_{nxn} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \\ \pi_0 & \pi_1 & \pi_2 & 0 & \cdot & \cdot & \cdot & 0 \end{pmatrix}$$

Likelihood Function

Likelihood

- Under assumption (CI) the likelihood function ℓ^f has the particular simple form

$$\ell^f\left(x_1,\dots x_T,d_1,\dots d_r|x_0,d_0,\theta\right)=\prod_{t=1}^{T}P\left(d_t|x_t,\theta\right)p\left(x_t|x_{t-1},d_{t-1},\theta_3\right)$$

where $P\left(d_t|x_t,\theta\right)$ is the choice probability given the observable state variable, x_t.

How to compute the choice probability, $P\left(d_t|x_t,\theta\right)$?

- Need to solve dynamic program Formula given on slide 7: calc EV from solving Bellman

How to estimate the transition probability, $p\left(x_t|x_{t-1},d_{t-1},\theta_3\right)$?

- Can be estimated in a first step without solving DP problem (non-parametrically or parametrically)
  ...or jointly with DP problem if $p\left(x_t|x_{t-1},d_{t-1},\theta_3\right)$ is fully specified.

Structural Estimation

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

Log likelihood function

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

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

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 Γ always has a unique fixed point, the constraint $EV = \Gamma_\theta(EV)$ implies that the fixed point EV_θ is an *implicit function* of θ.

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. θ
- Quasi-Newton algorithm: Usually BHHH, BFGS or a combination.
- Each evaluation of $L(\theta, EV_\theta)$ requires solution of EV_θ

Inner loop (fixed point algorithm):
The implicit function EV_θ defined by $EV_\theta = \Gamma(EV_\theta)$ is solved by:

- Successive Approximations (SA)
- Newton-Kantorovich (NK) Iterations

MATLAB implementation of the likelihood function

Abbreviated version of zucher.ll in zurcher.m

```
function [f,g,h]=ll(data, mp, theta)
  global V0;

  % Update model primitives
  n_u = numel(struct2vec(mp,mp.pnames_u));  % # of utility parameters
  n_P = numel(struct2vec(mp,mp.pnames_P));  % # of transition parameters
  mp=vec2struct(theta, [mp.pnames_u mp.pnames_P], mp);  % update model params.
  P = zurcher.statetransition(mp);  % update transition matrices
  u = zurcher.u(mp);  % Update pay-off

  % Solve model
  bellman= @(V) zurcher.bellman(V, mp, u, P);
  [V0, pk, dBellman_dV]=dpsolver.poly(bellman, V0, mp.ap, mp.beta);

  % Evaluate likelihood function
  y_j=[(1-data.d) data.d];  % choice dummies [keep replace]
  px_j=[pk(data.x) 1-pk(data.x)];  %  choice probabilities evaluated at observed
  logl=log(sum(y_j.*px_j,2));  % 1) log likelihood regarding replacement choice
  if n_P>0;  % 2) add on log like for mileage process
    p=[mp.p; 1-sum(mp.p)];
    logl=logl + log(p(1+ data.dx1));
  end

  f=mean(-logl);  % Objective function (negative mean log likelihood)

end
```

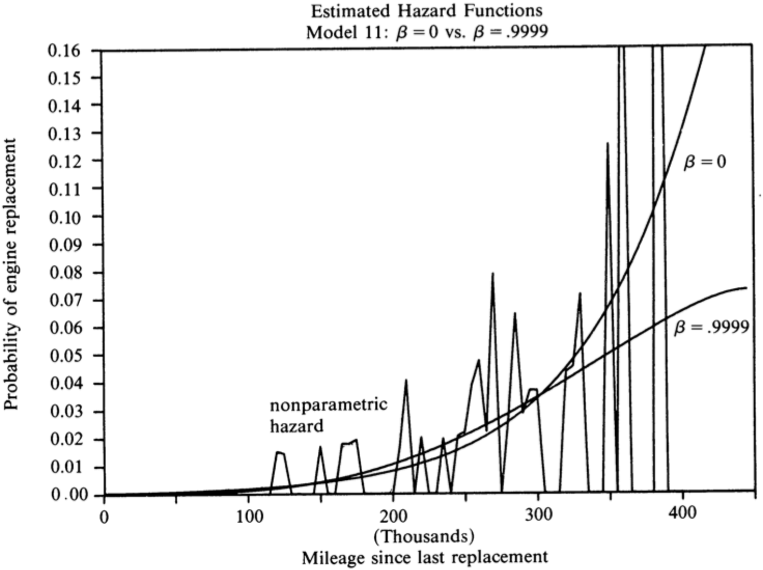
Python implementation of the likelihood function

```python
def ll(theta, data, mp, pnames, ap,x_int):
    global ev; ev0 = ev

    #Update model parameters, payoffs and transition matrix
    mp = updatepar(mp,pnames,theta)
    c = mp.c*0.001*mp.grid
    P = zucher.state_transition(mp.p,mp.n)

    # Solve the model
    bellman_sa = lambda ev: zucher.bellman_equ(ev,P,c,mp,1)
    bellman_nk = lambda ev: zucher.bellman_equ(ev,P,c,mp,2)
    ev, pk, dev, _ , _ = Solve.poly(bellman_sa,bellman_nk,
                                    ev0,ap,mp.beta)

    # Evaluate likelihood function
    log_lik = pk[x_int]*(1-data.d)+(1-pk[x_int])*data.d
    if theta.size>2:      # add on log like for mileage process
        p = np.append(mp.p,1-np.sum(mp.p))
        dx1_int = data.dx1.astype(int) # recast type to int
        log_lik += np.log(p[dx1_int])

    return log_lik
```

Data

- Harold Zurcher's Maintenance records of 162 busses
- Monthly observations of mileage on each bus (odometer reading)
- Data on maintenance operations
  1. Routine, periodic maintenance (e.g. brake adjustments)
  2. Replacement or repair at time of failure
  3. Major engine overhaul and/or replacement
- Rust focus on 3)

Estimated Hazard Functions



Specification Search

TABLE VIII

SUMMARY OF SPECIFICATION SEARCH[a]

	Bus Group		
Cost Function	1, 2, 3	4	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

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
	θ_{11}	4.8259 (1.792)	2.2930 (0.639)	2.6275 (0.618)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3489 (.0052)		
	θ_{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
	θ_{11}	109.9031 (26.163)	71.5133 (13.778)	70.2769 (10.750)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3488 (.0052)		
	θ_{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
	θ_{11}	2.4569 (.9122)	1.1732 (0.327)	1.3428 (0.315)		
	θ_{30}	.0937 (.0047)	.1191 (.0050)	.1071 (.0034)		
	θ_{31}	.4475 (.0080)	.5762 (.0075)	.5152 (.0055)		
	θ_{32}	.4459 (.0080)	.2868 (.0069)	.3621 (.0053)		
	θ_{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
	θ_{11}	56.1656 (13.4205)	36.6692 (7.0675)	36.0175 (5.5145)		
	θ_{30}	.0937 (.0047)	.1191 (.0050)	.1070 (.0034)		
	θ_{31}	.4475 (.0080)	.5762 (.0075)	.5152 (.0055)		
	θ_{32}	.4459 (.0080)	.2868 (.0069)	.3622 (.0053)		
	θ_{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:

Estimating parameters for bus types 1,2,3,4 (model 19)

Output from `run_busdata.m`:

```
bertelschjerning — MATLAB_maci64 -nodesktop ▸ matlab_helper — 94×20
>> run_busdata
Structural Estimation using busdata from Rust(1987)
Beta           =    0.99990
n              =  175.00000
Sample size    = 8156.00000


    Param.                      Estimates          s.e.        t-stat
-------------------------------------------------------------------------------------------
    RC                              9.7712        1.2127        8.0572
    c                               1.3439        0.3236        4.1529
    p                (1)            0.1070        0.0034       31.2090
    p                (2)            0.5152        0.0055       93.0605
    p                (3)            0.3622        0.0053       68.0442
    p                (4)            0.0143        0.0013       10.8946
    p                (5)            0.0009        0.0003        2.6469
-------------------------------------------------------------------------------------------
log-likelihood    = -8607.88684
runtime (seconds) =     0.41346
>>
```

You will do a similar implementation for python in the exercise classes

Identification - scale of cost function

Identification problem?

- We only identify RC/σ and $c(x, \theta_1)/\sigma = 0.001 * \theta_1/\sigma * x$, (where σ is parameter that index the scale of the cost function).
- σ is unidentified form mileage and replacement data

How to deal with identification problem related to scale of utility?

- Using replacement cost data and structural estimates we can obtain a scale estimate
- Scale the estimates with observed average replacement costs

Average Engine Replacement Costs

TABLE III

AVERAGE ENGINE REPLACEMENT COSTS[a]

	Bus Group		
Operation	1, 2, 3	4	1, 2, 3, 4
Labor time[b] to drop engine	\$ 150	\$ 150	\$ 150
Labor time[b] to overhaul engine	3373	2870	3032
Parts required to overhaul engine	5826	4343	4730
Labor time[b] to re-install engine	150	150	150
Total cost of replacement	\$9499	\$7513	\$8062

- Replacement costs are *higher* for group 1,2,3 although engine replacements for this group occur 57.000 miles and 15 month *earlier*
- Presumably operating and maintenance costs for these busses increase much faster

Identification - scale of cost function

- Using replacement cost data (prev. slide) and structural estimates from Table IX (next slide) we can obtain a scale estimate

$$\begin{aligned}\sigma_{bus\ 1,2,3} &= \frac{RC}{RC/\sigma} \\ &= \$9499/11.7257 \\ \sigma_{bus\ 4} &= \$7513/10.0750\end{aligned}$$

- We can the obtain a dollar estimate of $c(x, \theta_1)$ (i.e monthly maintenance costs per accumulated 5000 miles)

$$\begin{aligned}c(x, \theta_1)_{bus\ 1,2,3} &= \sigma * 0.001\theta_{11}/\sigma * x \\ &= \$9499/11.725 * 0.001 * 4.82 * x = \$3.9 * x \\ c(x, \theta_1)_{bus\ 4} &= \$7513/10.0750 * 0.001 * 2.2930 * x = \$1.7 * x\end{aligned}$$

- Hence, a bus with mileage of 300.000 (i.e. $x = 300.000/5.000$) is $(3.9 - 1.7) * 300000/5000 = \$132$ more expensive to operate per month

Structural Estimates

TABLE IX

STRUCTURAL ESTIMATES FOR COST FUNCTION $c(x, \theta_1) = .001\theta_{11}x$

FIXED POINT DIMENSION = 90

(Standard errors in parentheses)

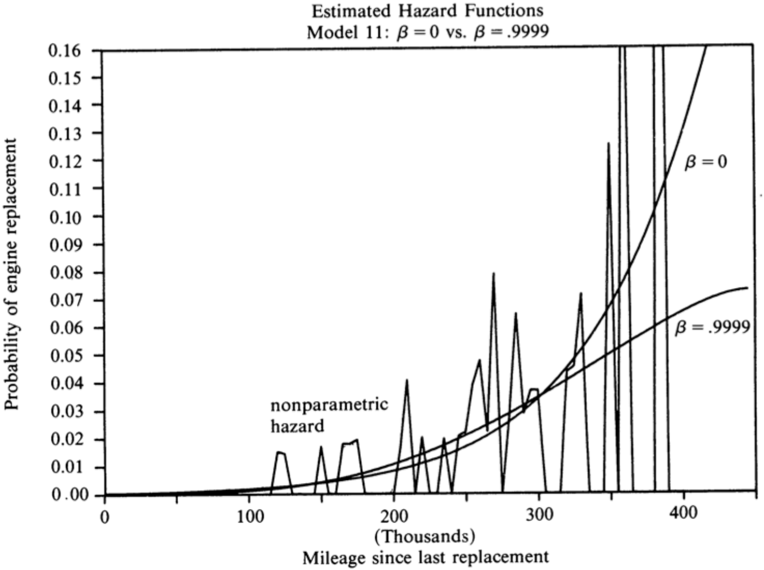
Parameter		Data Sample			Heterogeneity Test	
Discount Factor	Estimates/ Log-Likelihood	Groups 1, 2, 3 3864 Observations	Group 4 4292 Observations	Groups 1, 2, 3, 4 8156 Observations	LR Statistic ($df = 4$)	Marginal Significance Level
$\beta = .9999$	RC	11.7270 (2.602)	10.0750 (1.582)	9.7558 (1.227)	85.46	1.2E−17
	θ_{11}	4.8259 (1.792)	2.2930 (0.639)	2.6275 (0.618)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3489 (.0052)		
	θ_{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
	θ_{11}	109.9031 (26.163)	71.5133 (13.778)	70.2769 (10.750)		
	θ_{30}	.3010 (.0074)	.3919 (.0075)	.3488 (.0052)		
	θ_{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		

Why a dynamic model?

Suppose the "true" β is > 0, but we estimate the model with $\beta = 0$

- Our estimate of the replacement cost function will be biased.
- Parameters RC and θ_1 would be biased too
  (RC is upward biased and θ_1 is downward biased.)
- Predictions using the estimated model will be biased for two reasons:
  1. parameter estimates are biased
  2. the static model is not correct.
- Though the biases introduced by (1) and (2) might partly compensate each other, it will be a very unlikely coincidence that they compensate each other to make the bias negligible.
- Effect on equilibrium demand and hazard functions are very different!

Estimated Hazard Functions



Equilibrium bus mileage and demand for engines

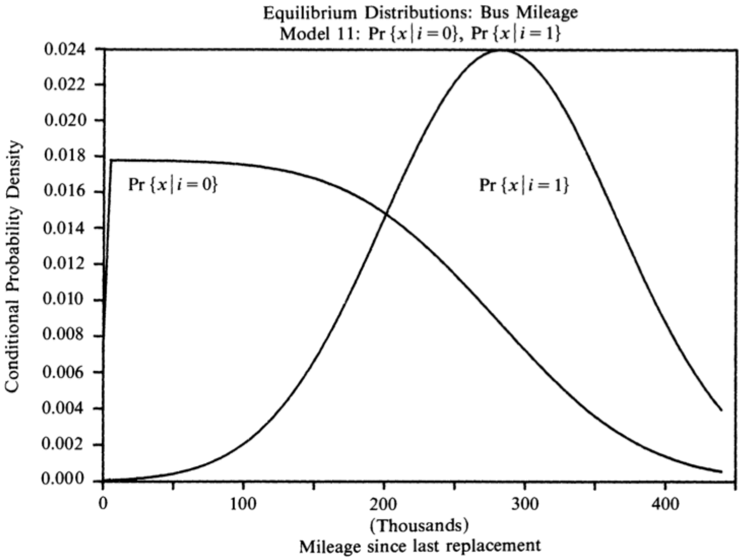
- Let π be the long run stationary (or equilibrium) distribution of the controlled process $\{i_t, x_t\}$
- π is then given by the unique solution to the functional equation

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

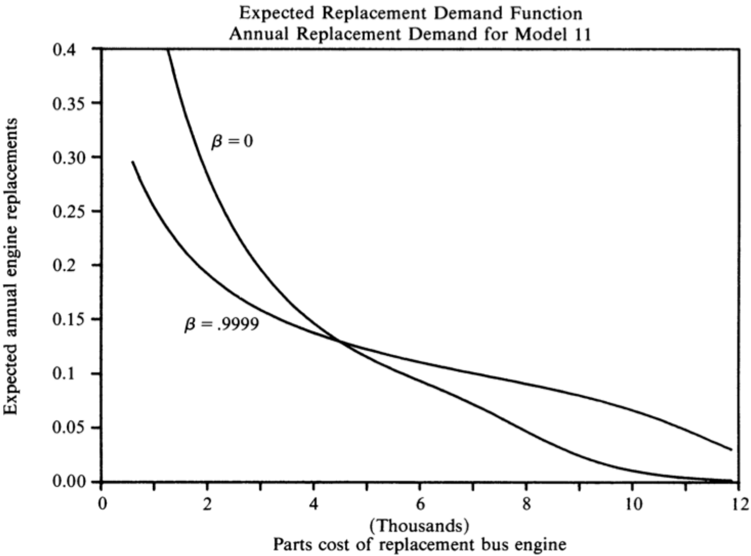
- π is the ergodic distribution of the controlled state transition matrix
- Carly the equilibrium distribution of π is an implicit function of the structural parameters θ, which we emphasize by the notation π_θ
- Given π_θ, we can also obtain the following simple formula for annual equilibrium demand 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

Python resources: Ruspy

https://github.com/OpenSourceEconomics/ruspy

ruspy

Anaconda.org 1.1 | Platforms noarch | docs passing | License MIT | Continuous Integration Workflow passing | codecov 54% | code style black

`ruspy` is an open-source package for the simulation and estimation of a prototypical infinite-horizon dynamic discrete choice model based on

> Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica, 55* (5), 999-1033.

You can install `ruspy` via conda with

```
$ conda config --add channels conda-forge
$ conda install -c opensourceeconomics ruspy
```

Please visit our online documentation for tutorials and other information.