The Bus Engine Replacement Model with Serially Correlated Unobserved State Variables: A Deterministic Approach

Gregor Reich

Computation in CA
July 24, 2013

PRELIMINARY – PLEASE DO NOT CIRCULATE

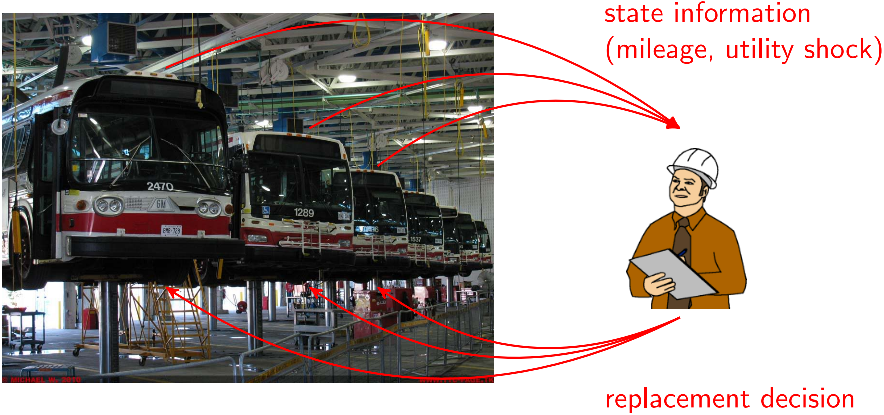
Abstract

- Dynamic Discrete Choice Models (DDCM) usually make strong **distributional assumptions** for the unobserved state variables (aka errors):
  - Extreme value type I (iid) distributed errors
  - Conditional independence (i. e. no serial correlation of errors)
- This assumption ensures **closed form solutions** of (potentially high-dimensional) integrals in both likelihood and expected value function
- This paper proposes a combination of numerical methods to solve these integrals numerically, allowing for:
  - **serially correlated errors**
  - variety of (conditional) distributions of the errors (e. g. normal)

Outline

The Bus Engine Replacement Model (Rust, 1987)

John Rust: *Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher.* Econometrica, 1987.



Formal Model in a Nutshell

- Utility per individual bus at time t

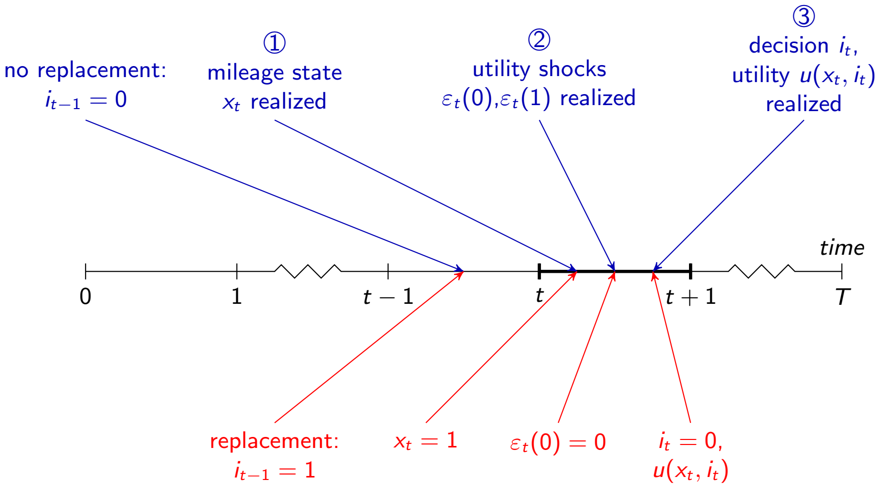
$$u(i, x_t, \theta_1) + \varepsilon_t(i) = \begin{cases} -RC + \varepsilon_t(1) & \text{if } i = 1 \\ -c(x_t, \theta_1) + \varepsilon_t(0) & \text{if } i = 0 \end{cases} \quad (1)$$

- State variables:
  - x_t observed; discretized; Markovian with probability vector θ_3
  - ε_t observable to agent, but not to econometrician; continuous (usual assumption: $\varepsilon_t(i) \sim EV1$ iid)
- Decision variable: i_t observed
- Bellman equation

$$V_\theta(x_t, \varepsilon_t) = \max_{i \in \{0,1\}} \{u(i, x_t, \theta_1) + \varepsilon_t(i) + \beta \mathbb{E}[V_\theta(x_{t+1}, \varepsilon_{t+1}) | i, x_t, \varepsilon_t]\}$$

- Estimation: Given data $\{x_t, i_t\}$, estimate model (1) using maximum likelihood

Timeline


no replacement:
$i_{t-1} = 0$
①
mileage state
x_t realized
②
utility shocks
$\varepsilon_t(0)$,$\varepsilon_t(1)$ realized
③
decision i_t,
utility $u(x_t, i_t)$
realized
time
0
1
$t-1$
t
$t+1$
T
replacement:
$i_{t-1} = 1$
$x_t = 1$
$\varepsilon_t(0) = 0$
$i_t = 0$,
$u(x_t, i_t)$

The Role of the Extreme Value Type 1 Distribution

- Problem: Computing EV and the likelihood function generally involves (high-dimensional) integration over the unobserved state variables $\varepsilon(i)$
- "Solution": Assume $\varepsilon_t(i) \sim EV1$ iid to get closed form solutions for these integrals
- Few empirical justification for this assumption (see Larsen et al, 2012)
- Misspecification can lead to biased parameter estimates

Outline

The Role of the Conditional Independence Assumption

- Conditional Independence assumption (CI)

$$Pr(x_{t+1}, \varepsilon_{t+1}|i, x_t, \varepsilon_t) = q(\varepsilon_{t+1} \mid x_{t+1})p(x_{t+1}|i, x_t)$$

  - ε: independent of ε_t (no serial correlation)
  - x: mileage transition independent of ε

- Decision probabilities under CI ($m_{it} \equiv u_{it} + \beta EV_{it}$)

$$Pr(i = 1|x_t, \theta) = Pr(\varepsilon_t(1) + m_{1t} > \varepsilon_t(0) + m_{0t})$$

- If $i = 1$ is rare (optimal stopping problem), the whole model is driven by the tail of the distribution of $\varepsilon_t(1) - \varepsilon_t(0)$ ("the agent is taken off-guard")

Serially Correlated Unobserved State Variables

- Serially correlated unobserved state variables

$$\begin{aligned} &\varepsilon_t(0) = \rho\varepsilon_{t-1}(0) + \tilde{\varepsilon}_t(0)\,, \quad \tilde{\varepsilon}_t(0) \sim f \text{ iid} \\ &\varepsilon_t(1) \sim f \text{ iid} \end{aligned} \tag{2}$$

- Remarks
  - Definition (2) nests the original model for $\rho = 0$ and f density of $EV1$
  - serial correlation only for $i = 0$
    - Serial correlation for the unobserved state variable in case of continuing is intuitive, but less in case of stopping
    - Interpolation is one-dimensional
    - Norets (2009) has an equivalent specification

  To be extended ...

Motivation for Serially Correlated Errors

- Decision probabilities with SCE (conditional)

$$Pr(i = 1|x_t, \theta, \varepsilon_{t-1}) = Pr(\varepsilon_t(1) + m_{1t} > \rho\varepsilon_{t-1}(0) + \tilde{\varepsilon}_t(0) + m_{0t})$$

- Conditional on ε_{t-1}, $Pr(i = 1 \mid \cdot)$ can be large, even if $i = 1$ is rare ("agent can anticipate replacement event")
- Rust (1987) does a specification test of CI, and concludes:

  "for groups 1, 2, and 3 and the combined groups 1-4 there is strong evidence that (CI) does not hold. The reason for rejection in the latter cases may be due to the presence of "fixed-effects" heterogeneity which induces serial correlation in the error terms."

Main Computational Tasks

- Serial correlation directly violates the CI assumption, thus no closed form solutions for integrals available
- Main computational tasks:
  - Expected value function
  - Likelihood function

Literature

- Norets (2009):
  - Instead of explicit likelihood integration, a MCMC approach is used to obtain distribution of parameters
  - The expected value function is obtained using random grids (small size), and value function iteration (few iterations).
  - Convergence of this method is proved

Outline

Approximating the Expected Value Function

$$\mathbb{E}[V(x',\varepsilon') \mid x, \varepsilon(0)] = \sum_{x'} \iint_{\varepsilon'} V(x',\varepsilon')\, q(d\varepsilon' \mid \varepsilon, i)\, Pr(x' \mid x, i)$$

$$= \sum_{x'} \int_{\tilde{\varepsilon}'(0)} \int_{\tilde{\varepsilon}'(1)} \max\Big\{ u(1,0) + \tilde{\varepsilon}'(1) + \beta\, \mathbb{E}[V(x'',\varepsilon'') \mid 0, 0],$$

$$u(0,x') + \rho\varepsilon + \tilde{\varepsilon}'(0) + \beta\, \mathbb{E}[V(x'',\varepsilon'') \mid x', \varepsilon'(0)] \Big\}\, q(d\tilde{\varepsilon}'(0))\, q(d\tilde{\varepsilon}'(1))\, Pr(x' \mid x, i)$$

- Note: $\mathbb{E}[V(x',\varepsilon') \mid x, \varepsilon(0)]$ is a function of $\varepsilon(0)$ only
- Computation
  1. Numerical integration over $\tilde{\varepsilon}$
  2. Discretization Γ_ε and interpolation $EV_\Gamma(x,\varepsilon)$ (only if integration over $\tilde{\varepsilon}$ rather than ε)
  3. Solution of the fixed point problem $EV_\Gamma(x,\varepsilon) = TEV_\Gamma(x,\varepsilon) \quad \forall x \in \Gamma_x, \forall \varepsilon \in \Gamma_\varepsilon$

Outline

Numerical Quadrature for the Expected Value Function

- Method: Gaussian Quadrature
  - $\tilde{\varepsilon}_t \sim N$: Gauss-Hermite quadrature
  - $\tilde{\varepsilon}_t \sim EV1$:
    - Change of variables

$$\int_{-\infty}^{+\infty} g(x)f(x) = \int_0^1 g(F^{-1}(x))$$

      where x is RV with density f and (invertible) distribution function F
    - Gauss-Legendre quadrature (unity weighting function)
- Issues
  - Integration over max function
    (singularity, potentially more nodes needed)
  - Conditional integration over $\tilde{\varepsilon}$ rather than ε

Discretization of the Support of ε

- Discretization: adaptive grid (Gruene and Semmler, 2004)
  - refinement:

    initialize η, Γ_ε
    while $\eta >$ threshold **do**
        solve $EV_\Gamma(x,\varepsilon) = TEV_\Gamma(x,\varepsilon)$
        for all grid cells l in Γ_ε **do**
            approximate $\eta_l = \max_{x,\varepsilon\in l} |EV_\Gamma(x,\varepsilon) - TEV_\Gamma(x,\varepsilon)|$
        end for
        $\eta = \max_l \eta_l$
        if $\eta_l > \theta\eta$ **then**
            insert node in cell l
        end if
    end while

  - coarsening: similar (using $\max_{x,\varepsilon\in l} |EV_\Gamma - EV_{\tilde{\Gamma}}|$, where $\tilde{\Gamma}_\varepsilon$ is "thinned grid")
- error bound: $\max_{x,\varepsilon} |EV(x,\varepsilon) - EV_\Gamma(x,\varepsilon)| \leq \frac{1}{1-\beta}\eta$

Interpolation of EV_Γ

- Interpolation: "Kindergarten"-method (Judd, 1998), aka piecewise linear interpolation
- Higher order methods
  - higher order splines caused instability in FX problem
  - general polynomial interpolation too "wiggly" (conjecture)

Solution of $EV_\Gamma(x, \varepsilon) = TEV_\Gamma(x, \varepsilon)$

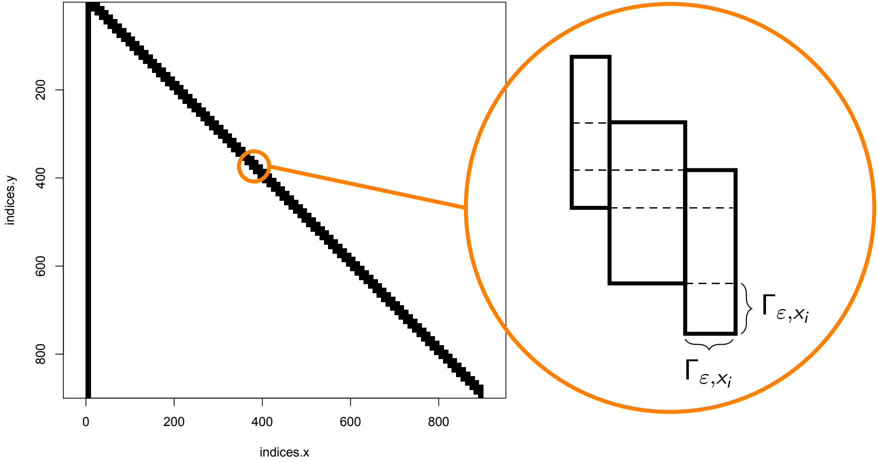
- General approach: solve NLES

$$0 = EV_\Gamma(x, \varepsilon) - TEV_\Gamma(x, \varepsilon)$$

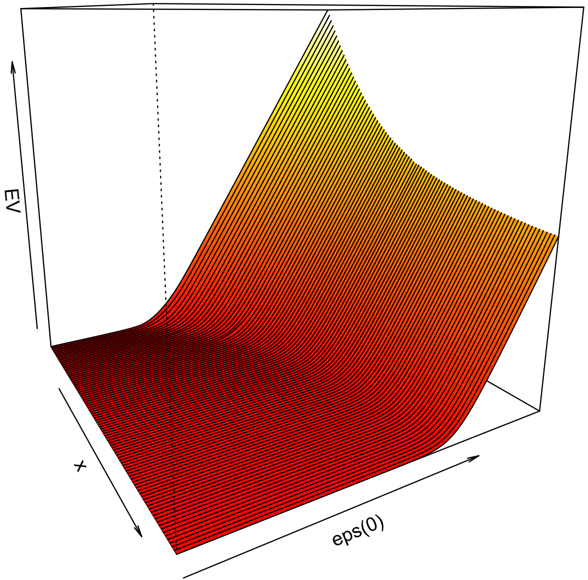
  Note: high accuracy is needed to have convergence in likelihood ("outer loop")

- Dimension: usually around $10 - 20,000$, precision: 1e-12
- Method and Solver (all parallel):
  - Newton ("ipopt" + "pardiso"; sparse)
  - Quasi-Newton (Broyden; R-package "nleqslv"; dense)
  - Quasi-Newton for sparse Jacobian (PETSc??)
- Sparsity: mileage transition probabilities imply sparsity of the Jacobian J (similar to block diagonal)

Sparsity Pattern of J



The Expected Value Function



Outline

The Likelihood Function

- General assessment of likelihood integration with serially correlated errors in DDCM:

  > *"In DDCMs, the likelihood function is an integral over the unobserved state variables. If the unobserved state variables are serially correlated, computing this integral is generally infeasible." (Norets, 2009)*

- However, for this model, a feasible and accurate approximation procedure exists.

Outline

Deriving the Likelihood Function (1)

$$
\begin{aligned}
L(\theta \mid \{x_t, i_t\}_{t=0,\dots,T}) &\equiv Pr(\{x_t, i_t\}_{t=0,\dots,T} \mid \theta) \\
&= \iint_{\varepsilon_0,\dots,\varepsilon_T} Pr(\{x_t, i_t, \varepsilon_t\}_{t=0,\dots,T})\, \mathrm{d}\varepsilon_0 \dots \mathrm{d}\varepsilon_T \\
&= \iint_{\varepsilon_0,\dots,\varepsilon_T} \prod_{t=1,\dots,T} Pr(\{x_t, i_t, \varepsilon_t\} \mid \{x_{t-1}, i_{t-1}, \varepsilon_{t-1}\})\, \mathrm{d}\varepsilon_0 \dots \mathrm{d}\varepsilon_T \\
&= \prod_{t=1,\dots,T} Pr(x_t \mid x_{t-1}, i_{t-1}) \iint_{\varepsilon_0,\dots,\varepsilon_T} \prod_{t=1,\dots,T} Pr(i_t \mid x_t, \varepsilon_t) Pr(\varepsilon_t \mid i_{t-1}, \varepsilon_{t-1})\, \mathrm{d}\varepsilon_0 \dots \mathrm{d}\varepsilon_T \\
&= \ell_1 \int_{\varepsilon_0} \dots \int_{\varepsilon_{T-2}} \dots \int_{\varepsilon_{T-1}} \dots \int_{\varepsilon_T} Pr(i_T \mid x_T, \varepsilon_T) Pr(\varepsilon_T \mid i_{T-1}, \varepsilon_{T-1})\, \mathrm{d}\varepsilon_0 \dots \mathrm{d}\varepsilon_{T-1} \mathrm{d}\varepsilon_T
\end{aligned}
$$

Deriving the Likelihood Function (2)

$$\int_{\varepsilon_t} Pr(i_t = 1 \mid x_t, \varepsilon_t)\ Pr(\varepsilon_t \mid i_{t-1}, \varepsilon_{t-1}) d\varepsilon_t$$

$$= \int_{\varepsilon_t(0)} Pr(\varepsilon_t(0) \mid i_{t-1}, \varepsilon_{t-1}(0)) \int_{\varepsilon_t(1)} Pr(i_t = 1 \mid x_t, \varepsilon_t(0), \varepsilon_t(1))\ Pr(\varepsilon_t(1))\ d\varepsilon_t(1) d\varepsilon_t(0)$$

$$Pr(i_t = 1 \mid x_t, \varepsilon_t(0), \varepsilon_t(1)) = \mathbb{1}(m_{1t} + \varepsilon_t(1) > m_{0t} + \varepsilon_t(0)) \qquad (m_{it} \equiv u_{it} + \beta EV_{it})$$

$$\int_{-\infty}^{\infty} \mathbb{1}(\varepsilon_t(1) > m_{0t} - m_{1t} + \varepsilon_t(0))\ Pr(\varepsilon_t(1))\ d\varepsilon_t(1)$$

$$= \int_{m_{0t} - m_{1t} + \varepsilon_t(0)}^{\infty} Pr(\varepsilon_t(1))\ d\varepsilon_t(1) = 1 - F(m_{0t} - m_{1t} + \varepsilon_t(0)) \equiv g(x_t, \varepsilon_t(0))$$

$$L = \ell_1 \int_{\varepsilon_0(0)} \dots \int_{\varepsilon_T(0)} Pr(\varepsilon_T(0) \mid i_{T-1}, \varepsilon_{T-1}(0))\ g(x_T, \varepsilon_T(0))\ d\varepsilon_0(0) \dots d\varepsilon_T(0)$$

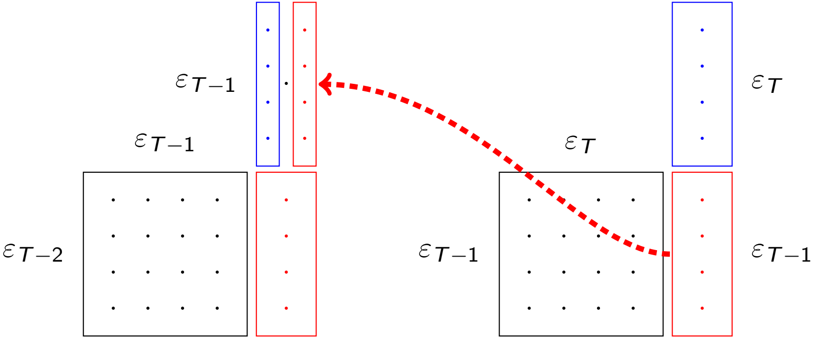
Outline

Numerical Quadrature for the Likelihood Function

- $\tilde{\varepsilon}_t \sim EV1$:
  - Gauss-Legendre quadrature
  - Change of variables
- $\tilde{\varepsilon}_t \sim N$
  - Gauss-Hermite quadrature

$\Rightarrow$ fixed set of integration nodes (no recursive schemes)

Computing the Likelihood Function (1)

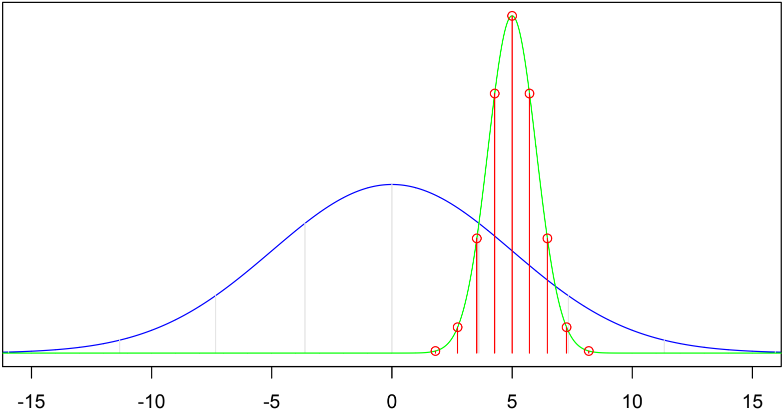


$$\dots \int_{\varepsilon_{T-1}} q(\varepsilon_{T-1} \mid \varepsilon_{T-2})\, g(x_{T-1}, \varepsilon_{T-1}) \int_{\varepsilon_T} q(\varepsilon_T \mid \varepsilon_{T-1})\, g(x_T, \varepsilon_T)\, d\varepsilon_0 \dots$$

complexity: $O(N) + O(N)$

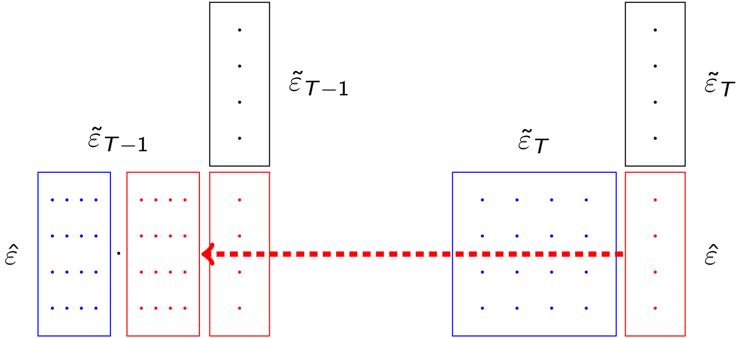
(Un-)Conditional Integration

placing integration nodes according to distribution of $\tilde{\varepsilon}$



$$\ldots \int_{\tilde{\varepsilon}_T} \tilde{q}(\tilde{\varepsilon}_T)\, g\left(x_T, \rho\varepsilon_{T-1} + \tilde{\varepsilon}_T\right) \mathrm{d}\tilde{\varepsilon}_0 \ldots \mathrm{d}\tilde{\varepsilon}_T$$

Computing the Likelihood Function (2)



$$\dots \int_{\tilde{\varepsilon}_{T-1}} \tilde{q}(\tilde{\varepsilon}_{T-1})\, g(x_{T-1}, \tilde{\varepsilon}_{T-1} + \rho\hat{\varepsilon}) \quad \int_{\tilde{\varepsilon}_{T}} \tilde{q}(\tilde{\varepsilon}_{T})\, g(x_{T}, \tilde{\varepsilon}_{T} + \rho\hat{\varepsilon})\, d\tilde{\varepsilon}_0 \dots$$

complexity: $O(N^2) + O(N^2)$

Outline

Maximization of the Likelihood Function

- Quasi-Newton trust-region method (R-package “trustOptim”)
- finite difference gradient approximation (GSL)
- Issues
  - problem scaling
  - Finite difference gradient approximation step length
  - numerical precision (??)

Outline

A Nested Fixed Point Algorithm

initialize θ, D (scaling matrix)
while $\nabla \ell_2 \neq 0$ **do**
    initialize Γ_ε, ε_{FD}, B_{BFGS}
    while $\nabla \ell_2 \neq 0$ **or** iter $<$ maxIter **do**
        while $\eta >$ threshold **do**
            solve $TEV_\theta(x, \Gamma_\varepsilon) = EV_\theta(x, \Gamma_\varepsilon)$
            update Γ_ε (coarsening and refinement)
        end while
        evaluate ℓ_2
        compute $\nabla \ell_2$ (update ε_{FD} if necessary)
        compute next θ (QNTR, updating B_{BFGS}, scaled by D)
    end while
    compute next D from $\nabla^2 \ell_2$
    update η
end while

Some Implementation Details

- Most code is written in R
- Time-critical components (*TEV*, *EV*, Jacobian) are written in C++ (Intel)
- Parallelization using OpenMP
- Computations are carried out on AMD Opteron (AbuDhabi) 4x16 core workstation

Outline

Replication: Rust (1987) Table IX, and Simulated Data

	Bus Groups 1-4 ($N = 8,156$)		Simulated Data ($N = 106,132$)	
	Rust (1987)	estimated	true	estimated
RC	9.7558	9.7557	14.0000	13.9959
θ_1	2.6275	2.6274	2.0000	2.0390
ρ	0	0	0.6000	0.5997
θ_{30}	0.3489	0.3489	0.3489	0.3489
θ_{31}	0.6394	0.6394	0.6394	0.6394
L	-6055.250	-6055.250		-81749.86
$\|\|\nabla L\|\|$		1e-9		1e-5

$$\beta = .9999$$

Estimation: $\tilde{\varepsilon} \sim EV1(-\gamma, 1)$

	Bus Groups 1-3 ($N = 3,864$)		Bus Group 1-4 ($N = 8,156$)	
RC	11.8270	25.0000	9.7557	26.4972
θ_1	4.6724	9.8347	2.6274	7.2392
RC/θ_1	2.5313	2.5420	3.7130	3.6602
ρ	0	0.6894	0	0.7366
L	-2708.335	-2707.765	-6055.250	-6053.341
$\|\|\nabla L\|\|$	1e-7	1e-6	1e-9	1e-5
p (LR)		0.2854		0.0507

$\beta = .9999$, p (LR) is p-value of likelihood ratio test $H_0 : \rho = 0$

Estimation: $\tilde{\varepsilon} \sim N(0,1)$

	Bus Groups 1-3 ($N = 3,864$)		Bus Group 1-4 ($N = 8,156$)	
RC	7.0870	13.9130	6.0047	18.4240
θ_1	2.4586	5.4257	1.4011	5.1150
RC/θ_1	2.8826	2.5643	4.2857	3.6020
ρ	0	0.5230	0	0.6623
L	-2707.877	-2707.820	-6054.084	-6053.685
$\lvert\lvert \nabla L \rvert\rvert$	1e-5	1e-5	1e-6	1e-5
p (LR)		0.7354		0.3713

$\beta = .9999$, p (LR) is p-value of likelihood ratio test $H_0 : \rho = 0$

Summary and Outlook

- Summary
  - Estimation of a popular DBCM with serially correlated unobserved state variables, using a fully deterministic approach
  - For some datasets, significant serial correlation could be identified
- Outlook
  - Resolve some technicalities (standard errors, distribution of ε_0, try PETSc etc.)
  - Generalize to DDCM

L. Grüne and W. Semmler.
Using dynamic programming with adaptive grid scheme for optimal control problems in economics.
Journal of Economic Dynamics and Control, 28(12):2427–2456, 2004.

B. Larsen, F. Oswald, G. Reich, and D. Wunderli.
A test of the extreme value type I assumption in the bus engine replacement model.
Economics Letters, 116(2):213–216, 2012.

A. Norets
Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables.
Econometrica, 77 (5): 1665-1682, 2009

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