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

Gregor Reich

Quantitative Business Administration
University of Zurich

October 25, 2013
Dynamic Discrete Choice Models (DDCM) usually make strong distributional assumptions.

They ensure closed form solutions of potentially high-dimensional integrals.

Most of these assumptions have been rejected.

Potential problems

- Bias from mis-specification
- Dynamic nature of decision problem can get lost

This paper proposes a procedure to approximate these integrals using deterministic numerical quadrature, allowing

- serially correlated errors (e.g. AR(1))
- variety of distributions of the errors (e.g. normal)

Application: Bus engine replacement model of Rust (1987)
Outline

1. Introduction

2. The Bus Engine Replacement Model (Rust, 1987)
   - Model and Common Assumptions
   - Serially Correlated Unobserved State Variables

3. The Expected Value Function

4. The Likelihood Function

5. Estimation Results
Formal Model in a Nutshell

- Single period utility (cost) per individual bus at time $t$

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

- $i = 1$: engine replacement
- $i = 0$: regular maintenance only

- State and decision variables:
  - $x_t$ observed; discretized; Markovian with probability vector $\theta_3$
  - $\varepsilon_t$ observable to agent, but not to econometrician; continuous
    (note: agent observes $\varepsilon_t(i)$ for all $i$ before decision)
  - $i_t$ 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
Common Distribution Assumption

- Problem: Computing EV and the likelihood function generally involves high-dimensional integration over the unobserved state variables $\varepsilon(i)$
- Two important assumptions
  - Conditional Independence assumption of Rust (1987)
    \[
    Pr(x_{t+1}, \varepsilon_{t+1} | i, x_t, \varepsilon_t) = q(\varepsilon_{t+1} | x_{t+1})p(x_{t+1} | i, x_t) \quad \text{(CI)}
    \]
  - $\varepsilon$: $\varepsilon_{t+1}$ independent of $\varepsilon_t$ (no serial correlation)
  - $x$: mileage transition independent of $\varepsilon$
  - Extreme value type I distributed errors
    \[
    \varepsilon_t(i) \sim EV1 \text{ iid}
    \]
- Under these assumptions, closed form solutions for these integrals exist
Serially Correlated Unobserved State Variables

- Extending the model: AR(1) unobserved state variables (Norets, 2009)

\[
\varepsilon_t(0) = \rho \varepsilon_{t-1}(0) + \tilde{\varepsilon}_t(0), \quad \tilde{\varepsilon}_t(0) \sim f \text{ iid}
\]

\[
\varepsilon_t(1) = \tilde{\varepsilon}_t(1), \quad \tilde{\varepsilon}_t(1) \sim f \text{ iid}
\]  

- Remarks
  - (2) violates (CI) assumption, thus no closed form solutions to integrals available
  - Serial correlation only for utility shock in case of no replacement \((i = 0)\)
  - Definition (2) nests the original model for \(\rho = 0\) and \(f\) density of \(EV1\)
Motivation for Serial Correlation: Intrinsic

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

\[
Pr(i = 1|x_t, \theta) = Pr(\epsilon_t(1) + m_{1t} > \epsilon_t(0) + m_{0t})
\]

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

- Decision probabilities with serially correlated errors

\[
Pr(i = 1|x_t, \theta, \epsilon_{t-1}) = Pr(\epsilon_t(1) + m_{1t} > \rho \epsilon_{t-1}(0) + \tilde{\epsilon}_t(0) + m_{0t})
\]

Conditional on $\epsilon_{t-1}$, $Pr(i = 1 | \cdot)$ 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 [...] serial correlation in the error terms.”
“the likelihood function for a DDCM can be thought of as an integral over latent variables (the unobserved state variables). If the unobservables are serially correlated, computing this integral is very hard.” (Norets, 2009)

- **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
    - proved for AR(1)
    - potentially slow (Monte Carlo error: $C/\sqrt{N}$)

- **Reich (2013)**
  - Decompose the integral of the likelihood function
  - Approximate the low-dimensional integrals with efficient, deterministic quadrature rules
Outline

1. Introduction

2. The Bus Engine Replacement Model (Rust, 1987)

3. The Expected Value Function
   - Definition
   - Numerical Approximation

4. The Likelihood Function

5. Estimation Results
Approximating the Expected Value Function

\[ EV(x, \varepsilon) = \mathbb{E}[V(x', \varepsilon') | x, \varepsilon] \]

\[ = \sum_{x'} \int \max\left\{ u(0, x') + \rho \varepsilon(0) + \tilde{\varepsilon}'(0) + \beta EV(x', \varepsilon'), \right. \\
\left. u(1, 1) + \tilde{\varepsilon}'(1) + \beta EV(1, 0) \right\} q(d\tilde{\varepsilon}') Pr(x' | x, i) \]

- **Computation**
  1. Numerical integration over \( \tilde{\varepsilon} \)
  2. Discretization of \( \varepsilon \) (\( \Gamma_\varepsilon \)) and interpolation of \( EV(x, \varepsilon) \)
  3. Solution of the fixed point problem
    \[ EV(x, \Gamma_\varepsilon) = T(EV)(x, \Gamma_\varepsilon) \]

- This function is needed for likelihood function computation later.
Numerical Quadrature

- **Method:** Gaussian Quadrature
  - if $\tilde{\epsilon}_t \sim N(0, 1)$: Gauss-Hermite quadrature
  - if $\tilde{\epsilon}_t \sim EV1$:
    - Change of variables to map domain of integration $[-\infty, \infty] \rightarrow [-1, 1]$
    - Gauss-Legendre quadrature (unity weighting function)
  - Product rule for multi-dimensional integrals
    (only 2-dimensional for binary choice)
- **Note:** Integration over max function
  (kink $\rightarrow$ potentially more nodes needed)
Discretization and Interpolation

- Discretization of $EV(x, \varepsilon)$
  - adaptive grid (Gruene and Semmler, 2004)
  - refinement and coarsening $\rightarrow$ “update”
  - error bound: $\max_{x,\varepsilon} |EV(x, \varepsilon) - EV(x, \Gamma_{\varepsilon})| \leq \frac{1}{1-\beta} \eta$

- Interpolation: piecewise linear interpolation
Solution of Fixed Point Problem

- General approach: solve NLES

\[ 0 = EV(x, \Gamma_\epsilon) - T(EV)(x, \Gamma_\epsilon) \quad (3) \]

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

- usually around 10 – 20,000 number of equations and unknowns
- Sparsity: System (3) has sparse Jacobian \( J \)
  - implied by mileage transition probabilities
  - similar to block diagonal
Sparsity Pattern of $J$
The Expected Value Function
Outline

1. Introduction
2. The Bus Engine Replacement Model (Rust, 1987)
3. The Expected Value Function
4. The Likelihood Function
   - Definition
   - Numerical Approximation
5. Estimation Results
Deriving the Likelihood Function

\[ L(\theta \mid \{x_t, i_t\}_{t=0,\ldots,T}) \equiv Pr(\{x_t, i_t\}_{t=0,\ldots,T} \mid \theta) \]

\[
= \int\int \cdots \int Pr(\{x_t, i_t, \varepsilon_t\}_{t=0,\ldots,T}) \, d\varepsilon_0 \cdots d\varepsilon_T \\
= \ell_1 \int \cdots \int \cdots \int \cdots \\
\varepsilon_0 \quad \varepsilon_1 \quad \varepsilon_2 \\
\int \cdots \int Pr(i_{T-1} \mid x_{T-1}, \varepsilon_{T-1})Pr(\varepsilon_{T-1} \mid i_{T-2}, \varepsilon_{T-2}) \\
\varepsilon_{T-2} \quad \varepsilon_{T-1} \\
\int Pr(i_T \mid x_T, \varepsilon_T)Pr(\varepsilon_T \mid i_{T-1}, \varepsilon_{T-1}) \, d\varepsilon_0 \cdots d\varepsilon_{T-1}d\varepsilon_T \]
Computing the Likelihood Function: Backward Induction

- Backward induction (analogous to finite horizon, discrete time DP)
- Define $g_t(\varepsilon) : \mathbb{R}^N \rightarrow \mathbb{R}$

$$g_t(\varepsilon) = \begin{cases} 1 & t > T \\ \int_{\varepsilon'} Pr(i' \mid x, \varepsilon') Pr(\varepsilon' \mid i, \varepsilon) g_{t+1}(\varepsilon') \, d\varepsilon' & \text{otherwise} \end{cases}$$

- Algorithm:
  1: discretize support of $\varepsilon \rightarrow \Gamma_\varepsilon \in \mathbb{R}^D$
  2: initialize interpolant $\hat{g}(\cdot)$ with nodes $\{\hat{g}_e\}_{e \in \Gamma_\varepsilon} \in \mathbb{R}^D$ to 1
  3: for $t \in T, \ldots, 1$ do
  4: for $e \in \Gamma_\varepsilon$ do
  5: $\hat{g}_e \leftarrow$ approximate $\int_{\varepsilon'} Pr(i' \mid x, \varepsilon') Pr(\varepsilon' \mid i, e) \hat{g}(\varepsilon') \, d\varepsilon'$
  6: end for
  7: $\hat{g}(\cdot) \leftarrow$ interpolant with nodes $\{\hat{g}_e\}_{e \in \Gamma_\varepsilon}$
  8: end for

- Integral is approximated using Gaussian quadrature again
Computing the Likelihood Function: Comments

- \((N \cdot T)\)-dimensional integral is split up into \(D \cdot T\) integrals of dimension \(N\):

\[
\mathcal{O}(\exp(T \cdot N)) \gg \mathcal{O}(D \cdot T \exp(N))
\]

- Algorithm is generic for DDCM, with AR(1) errors as in (2)
- Likelihood function maximization is done using a nested fixed point algorithm
Some Implementation Details

- Most code is written in R
- Time-critical components ($T$ operator, Jacobian) are written in C++
- Libraries:
  - GSL (interpolation)
  - ipopt+pardiso (sparse fixed-point problem)
  - trustOptim (likelihood maximization)
- Parallelization using OpenMP; partially SIMD-vectorized
- Computations are carried out on 64 core workstation
Outline

1. Introduction
2. The Bus Engine Replacement Model (Rust, 1987)
3. The Expected Value Function
4. The Likelihood Function
5. Estimation Results
Replication: Rust (1987) Table IX

<table>
<thead>
<tr>
<th></th>
<th>Rust (1987)</th>
<th>estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RC$</td>
<td>9.7558</td>
<td>9.7557</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>2.6275</td>
<td>2.6274</td>
</tr>
<tr>
<td>$\rho$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\theta_{30}$</td>
<td>0.3489</td>
<td>0.3489</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.6394</td>
<td>0.6394</td>
</tr>
<tr>
<td>$L$</td>
<td>-6055.250</td>
<td>-6055.250</td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\nabla L</td>
</tr>
</tbody>
</table>

$\beta = .9999$
Validation: Simulated Datasets

<table>
<thead>
<tr>
<th>Simulated Data</th>
<th>true</th>
<th>estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EV1</td>
<td>$N(0, 1)$</td>
</tr>
<tr>
<td>$RC$</td>
<td>14.0000</td>
<td>13.9959</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>2.0000</td>
<td>2.0390</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.6000</td>
<td>0.5997</td>
</tr>
<tr>
<td>$\theta_{30}$</td>
<td>0.3489</td>
<td>0.3489</td>
</tr>
<tr>
<td>$\theta_{31}$</td>
<td>0.6394</td>
<td>0.6394</td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\nabla L</td>
</tr>
</tbody>
</table>

$\beta = .9999$
Estimation: $\xi \sim EV1$

<table>
<thead>
<tr>
<th></th>
<th>Bus Groups 1-3 $(N = 3,864)$</th>
<th>Bus Group 1-4 $(N = 8,156)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RC$</td>
<td>11.8270</td>
<td>9.7557</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>4.6724</td>
<td>2.6274</td>
</tr>
<tr>
<td>$RC/\theta_1$</td>
<td>2.5313</td>
<td>3.7130</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-</td>
<td>0.6894</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td>-2708.335</td>
<td>-6055.250</td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\nabla L</td>
</tr>
<tr>
<td>$p$ (LR)</td>
<td>0.2854</td>
<td>0.0507</td>
</tr>
</tbody>
</table>

$\beta = .9999$, $p$ (LR) is $p$-value of likelihood ratio test $H_0 : \rho = 0$
### Estimation: $\tilde{\varepsilon} \sim N(0, 1)$

<table>
<thead>
<tr>
<th></th>
<th>Bus Groups 1-3 $(N = 3,864)$</th>
<th>Bus Group 1-4 $(N = 8,156)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$RC$</td>
<td>7.0870</td>
<td>6.0047</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>2.4586</td>
<td>1.4011</td>
</tr>
<tr>
<td>$RC/\theta_1$</td>
<td>2.8826</td>
<td>4.2857</td>
</tr>
<tr>
<td>$\rho$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$L$</td>
<td>-2707.877</td>
<td>-6054.084</td>
</tr>
<tr>
<td>$</td>
<td></td>
<td>\nabla L</td>
</tr>
<tr>
<td>$p$ (LR)</td>
<td>0.7354</td>
<td>0.3713</td>
</tr>
</tbody>
</table>

For Bus Groups 1 to 3, the $\rho$ value is 0.5230, and for Bus Group 1 to 4, it is 0.6623. The $p$ (LR) value for the likelihood ratio test with $H_0 : \rho = 0$ is $\beta = .9999$.
Conclusion

- Estimation of a popular DBCM with serially correlated unobserved state variables, using deterministic quadrature rules
- For some datasets, significant serial correlation could be identified
L. Grüne and W. Semmler.
Using dynamic programming with adaptive grid scheme for optimal control problems in economics.

B. Larsen, F. Oswald, G. Reich, and D. Wunderli.
A test of the extreme value type I assumption in the bus engine replacement model.

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

J. Rust.