A Primer on Recursive Likelihood Function Integration

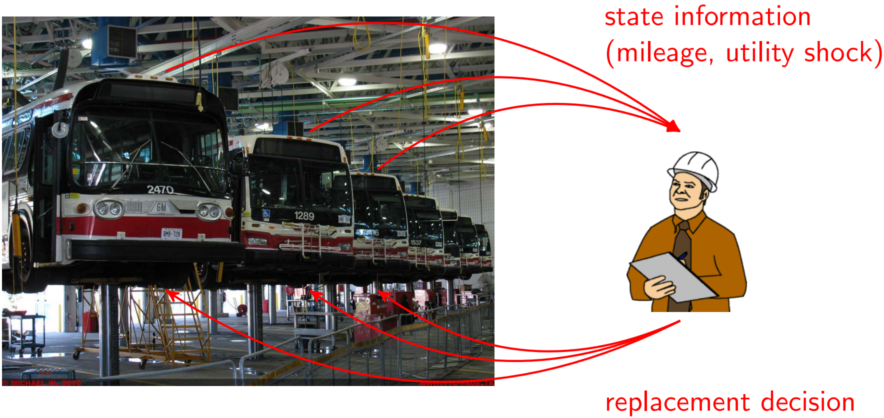
Jasmin Maag Gregor Reich

University of Zurich

September 13, 2019

The Bus Engine Replacement Model (Rust, 1987)

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



Dynamic Discrete Choice Models (1)

- Single period payoff/costs:

$$\begin{cases} -\text{price of new engine} & \textit{decision} = \textit{renewal} \\ -\text{maintenance cost (depending on age)} & \textit{decision} = \textit{continuation} \end{cases}$$

- More formally, model is **parametrized** by θ:

$$\pi(x, y; \theta) + \varepsilon(y) = \begin{cases} -RC & y = 1 \\ -\theta_{11} \cdot x & y = 0 \end{cases}$$

where x is the *state*, and y is the *decision* or *control*, and ε is a random utility component; RC and θ_{11} are elements of θ.

Dynamic Discrete Choice Models (2)

- Dynamic optimization problem

$$V_\theta(x_0,\varepsilon_0) = \max_{U_\theta(x,\varepsilon)} \mathbb{E}\Big[\sum_{t=0}^{K\leq\infty} \beta^t(\pi(x_t, U(x_t,\varepsilon_t);\theta) + \varepsilon_t)\Big]$$

given $P(x_t,\varepsilon_t|x_{t-1},\varepsilon_{t-1},y_{t-1};\theta)$, $\beta \in [0,1)$.

$V_\theta(x,\varepsilon)$ is the *value* function, $U_\theta(x,\varepsilon)$ is the *policy* function

- **Bellman** equation: Sufficient optimality condition for (4) (here: infinite horizon $K=\infty$)

$$\begin{aligned} V_\theta(x,\varepsilon) &= \max_y \Big\{\pi(x,y;\theta) + \varepsilon(y) + \beta\, \underbrace{\mathbb{E}\big[V_\theta(x',\varepsilon') \mid x,\varepsilon,y;\theta\big]}_{EV_\theta(x,\varepsilon,y)}\Big\} \\ &\equiv TV_\theta(x,\varepsilon) \end{aligned}$$

Note: The unknown is a *function*, $V\theta(\cdot)$ or $EV_\theta(\cdot)$, not x,ε!

Maximum Likelihood Estimation of Dynamic Models (1)

- To us as "econometricians", the models parameters θ are unknown.
- Panel **data** on controls y and states x (incomplete: no ε)

$$\boldsymbol{D} = \{(x_t^i, y_t^i)_{t=1}^T\}_{i=1}^I$$

- Likelihood function:

$$\begin{aligned} L(\theta) &= p(\{(x_t^i, y_t^i)_{t=1}^T\}_{i=1}^I; \theta) \\ &= \prod_{i=1}^I \prod_{t=1}^T p(x_t^i, y_t^i | x_{t-1}^i, y_{t-1}^i; \theta) \\ &= \prod_{i=1}^I \prod_{t=1}^T p(x_t^i | x_{t-1}^i, y_{t-1}^i; \theta) Pr(y_t^i | x_t^i; \theta) \end{aligned}$$

Maximum Likelihood Estimation of Dynamic Models (2)

- **Conditional choice probabilities**: Define

$$m(x, y; \theta) \equiv \pi(x, y; \theta) + \beta \mathbb{E}\left[V_\theta(x', \varepsilon') \mid x, \varepsilon, y; \theta\right]$$

  Under extreme value type I iid errors ε:

$$\begin{aligned} Pr(y = 1|x; \theta) &= Pr(m(1, x; \theta) + \varepsilon(1) > m(0, x; \theta) + \varepsilon(0)) \\ &= \frac{\exp(m(1, x; \theta))}{\sum_{j \in \{0,1\}} \exp(m(j, x; \theta))} \end{aligned}$$

- **Estimation** of θ by maximum likelihood:

$$\left.\begin{array}{rcl} \theta^* &=& \arg\max L(\theta; V_\theta) \\ V_\theta(x, \varepsilon) &=& TV_\theta(x, \varepsilon)\Big|_{\theta=\theta^*} \end{array}\right\}$$

  - Nested fixed point algorithm (NFXP; Rust, 1987)
  - Mathematical programming with equilibrium constraints (MPEC; Su and Judd, 2013)

Serially Correlated Unobserved States

- Problem: **Unobservable states**, ε, need to be integrated out to obtain the *marginal* likelihood function (for one individual i)

$$L^i(\theta) \equiv p(\{x_t^i, y_t^i\}_{t=1}^T | x_0^i, y_0^i; V_\theta, \theta)$$
$$= \int \cdots \int \prod_{t=1}^T p(x_t, y_t, \varepsilon_t \mid x_{t-1}, y_{t-1}, \varepsilon_{t-1}; \theta) \, d\varepsilon_1 \ldots \varepsilon_T$$

- Solution approaches
  - Assume ε are *serially uncorrelated* (conditional on x and y)

$$L^i(\theta) = \prod_{t=1}^T \int p(x_t, y_t, \varepsilon_t \mid x_{t-1}, y_{t-1}; \theta) \, d\varepsilon_t$$

  - Monte Carlo approaches
    - "Plain" MC (e.g. Pakes, 1986; Keane and Wolpin, 1994; etc.)
    - Bayesian inference using MCMC (Norets, 2009)
    - Sequential Monte Carlo (Blevins, 2016)
  - Recursive likelihood function integration (**RLI**; Reich, 2018)

Recursive Likelihood Function Integration (RLI)

- Recursive approach to likelihood function integration (omitting observed variables for brevity):

$$L^i(\theta) = \int \cdots \int \prod_{t=1}^{T} p(\varepsilon_t \mid \varepsilon_{t-1})\, d\varepsilon_T \ldots d\varepsilon_1$$

$$= \int \cdots \int \prod_{t=1}^{T-1} p(\varepsilon_t \mid \varepsilon_{t-1}) \underbrace{\int p(\varepsilon_T \mid \varepsilon_{T-1}) d\varepsilon_T}_{\equiv I_{f_T}(\varepsilon_{T-1})} d\varepsilon_{T-1} \ldots d\varepsilon_1$$

$$= \int \cdots \int \prod_{t=1}^{T-2} p(\varepsilon_t \mid \varepsilon_{t-1}) \underbrace{\int p(\varepsilon_{T-1} \mid \varepsilon_{T-2}) I_{f_T}(\varepsilon_{T-1}) d\varepsilon_{T-1}}_{\equiv I_{f_{T-1}}(\varepsilon_{T-2})} d\varepsilon_{T-2} \ldots d\varepsilon_1$$

$$= \cdots = I_{f_1}(\varepsilon_0)$$

- Recursive formulation:

$$I_{f_t}(\varepsilon_{t-1}) = \begin{cases} 1 & t > T \\ \int p(\varepsilon_t \mid \varepsilon_{t-1}) I_{f_{t+1}}(\varepsilon_t)\, d\varepsilon_t & 0 < t \leq T \end{cases}$$

A simple (iterative) implementation of RLI

initialize interpolation grid $(\varepsilon_i)_{i=1}^N$;
$\mathcal{I}(\cdot) \leftarrow$ initialize interpolant with $((\varepsilon_i, 1))_{i=1}^N$;
foreach $t \in T, \ldots, 1$ **do**
    foreach $\varepsilon_i \in (\varepsilon_i)_{i=1}^N$ **do**
        $I_{\varepsilon_i} \leftarrow$ approximate $\int p(\varepsilon \mid \varepsilon_i)\mathcal{I}(\varepsilon)\, d\varepsilon$;
    end
    $\mathcal{I}(\cdot) \leftarrow$ create new interpolant with $((\varepsilon_i, I_{\varepsilon_i}))_{i=1}^N$;
end
return $\mathcal{I}(\varepsilon_0)$

Numerical Considerations: Over- and Underflow

- Likelihood functions are usually optimized over in **log forms**, to avoid very small (or large) numbers, which create numerical issues (under- or overflow)
- However, the log operator cannot be moved into the integral.
- Solution: **Iteratively rescale interpolant** by moving a constant out of the integral:

$$I_{f_t}(\varepsilon_{t-1}) = \begin{cases} 1 & t > T \\ \int p(\varepsilon_t \mid \varepsilon_{t-1}) \frac{I_{f_{t+1}}(\varepsilon_t)}{\delta_{t+1}} \, d\varepsilon_t & 0 < t \leq T \end{cases}$$

with

$$\log(L(\theta)) = \log(I_{f_1}(\varepsilon_0)) + \sum_{i=2}^{T} \log(\delta_t)$$

Choose $\delta_t = O(I_{f_t}(\cdot))$, so $O(I_{f_t}(\cdot)/\delta_t) = 1$.

RLI: Convergence

- Question: How *fast* does RLI converge

Theorem

Given suitable choices of quadrature and interpolation methods, the convergence rate of RLI in terms of total integrand evaluations is:

$$|L^T - \hat{L}^T| = O\left(T\left(\frac{n}{T}\right)^{-\min\{s_Q\theta, s_I(1-\theta)\}}\right)$$

where s_Q and s_I are the convergence rates of the quadrature and interpolation methods, respectively, and $n_{Q,t} = (n/T)^\theta$ and $n_{I,t} = (n/T)^{1-\theta}$ with $\theta \in (0,1)$.

- Important special case: $s_Q = s_I = s \Rightarrow \theta = 0.5$ and convergence rate is $O(T(n/T)^{-s/2})$.

RLI: Convergence—Numerical Example

- Numerical experiment, cont.: Method comparison (T fixed)
  - Simpson + Cubic spline: $s_Q = s_I = 4 \Rightarrow s_{RLI} = 2$
  - Chebyshev poly. + Gauss–Hermite: exponential
  - Monte Carlo: $stdev(I_n) = O(n^{-1/2})$

- Confirmative result: all convergence rates seem to realize