Solution and Estimation of Dynamic Discrete Choice Structural Models Using Euler Equations

Victor Aguirregabiria (Toronto) & Arvind Magesan (Calgary)

November 2, 2016

Duke - UNC

Context: Hansen and Singleton (1982)

- The Euler equation – GMM approach (Hansen and Singleton, 1982) was a key contribution in the econometrics of continuous decision dynamic structural models.
- The method avoids the computation of discounted present values, and so the curse of dimensionality in estimation of structural parameters.
- Two **limitations** though.
- [1] Cannot be applied to discrete choice models.
- [2] Cannot be used for solution / counterfactuals: Euler equation-policy operator is not a contraction (Coleman, 1990).

Context: Hotz and Miller (1993)

- **Hotz-Miller** estimation methods avoid repeated solution of DP problem, allowing for richer specifications of observable state variables.
- Two limitations:
- [1] Hotz-Miller method (efficient version) is subject to curse of dimensionality: requires solving for present values. [Arcidiacono & Miller (2011, 2016) finite dependence representation avoids this computational cost at the price of losing estimation efficiency]
- [2] It does not deal with curse of dimensionality in solution of and computation of counterfactual experiments.

Our Contributions

- Aguirregabiria and Magesan (2013) derive **Euler equations** for a general class of **dynamic discrete choice models**.

1. **Euler fixed point operator**, that can be used to solve the DDC model. Show that this operator **is a contraction**.

   Euler operator is a stronger contraction than Value iteration or Relative Value iterations (and cheaper to evaluate) and implies substantial computational advantages for the solution of the DP.

2. **Sample-based Euler operator**. We use it to construct **estimators** of structural parameters and counterfactual experiments. These estimators imply **substantial improvements in computational cost and statistical efficiency** relative to existing methods.

Outline

[1] **Model and Basic Assumptions**

[2] **Euler Equations in DDC Models**

[3] **Euler operator: Contraction property**

[4] **Sample-based Euler operator: Estimators**

[5] **Numerical Example & Monte Carlo Experiments**

1. Model and Basic Assumptions

Dynamic DC Single-Agent Models

- Every time period t an agent makes a choice $a_t \in A = \{0, 1, ..., J\}$ to maximize $\mathbb{E}_t \left(\sum_{j=0}^{\infty} \beta^s \ \Pi(a_{t+j}, s_{t+j}) \right)$.

- s_t follows a controlled Markov process with transition $f(s_{t+1}|a_t, s_t)$.

- The value function $V(s_t)$ solves the Bellman equation:

$$V(s_t) = \max_{a_t \in A} \left\{ \Pi(a_t, s_t) + \beta \int V(s_{t+1}) \ f(s_{t+1}|a_t, s_t) \ ds_{t+1} \right\}$$

Observable and unobservable state variables

- $s_t = (x_t, \varepsilon_t)$: x_t observable; ε_t unobservable to researcher.
- **Additive separability (AS).** $\Pi(a_t, x_t, \varepsilon_t) = \pi(a_t, x_t) + \varepsilon_t(a_t)$
- **Conditional independence (CI).**
  $f(x_{t+1}, \varepsilon_{t+1} | a_t, x_t, \varepsilon_t) = f_x(x_{t+1} | a_t, x_t) \; g(\varepsilon_{t+1})$
- Optimal decision rule:

$$\{\alpha^*(x_t, \varepsilon_t) = a\} \quad \text{iff}$$

$$\{v(a, x_t) + \varepsilon_t(a) \geq v(j, x_t) + \varepsilon_t(j) \text{ for any } j \neq a\}$$

Decision rules as Conditional Choice Probabilities

- Given an arbitrary decision rule $\alpha(x_t, \varepsilon_t)$ from $X \times \mathbb{R}^{J+1}$ into A, we can defined **Conditional Choice Probability (CCP)** function:

$$P(a \mid x) \quad \equiv \quad \Pr\left(\alpha(x_t, \varepsilon_t) = a \mid x_t = x\right) = \Lambda\left(a, \widetilde{\mathbf{v}}(x)\right)$$

- *PROPOSITION 1 [Hotz-Miller Inversion]. Mapping* $\Lambda\left(.\right)$ *is invertible such that there is a one-to-one relationship between the vector of value differences* $\widetilde{\mathbf{v}}(x)$ *and the vector of optimal choice probabilities* $\mathbf{P}(x)$*, i.e.,* $\widetilde{\mathbf{v}}(x) = \Lambda^{-1}(\mathbf{P}(x))$*.*

Dynamic probability-choice problem

- Given a vector of CCPs $\mathbf{P}_t \equiv \{P_t(a) : a \in \mathcal{A} - \{0\}\}$, we define the **expected payoff function**:

$$\Pi^P(\mathbf{P}_t, x_t) \equiv \sum_{a=0}^{J} P_t(a) \left[\pi_t(a, x_t) + e_t(a, \mathbf{P}_t)\right],$$

where

$$e_t(a, \mathbf{P}_t) = \mathbb{E}\left[\varepsilon_t(a) \mid \Lambda^{-1}(a, \mathbf{P}_t) + \varepsilon_t(a) \geq \Lambda^{-1}(j, \mathbf{P}_t) + \varepsilon_t(j) \quad \forall j\right]$$

- And define the **expected transition probability**,

$$f^P(x_{t+1}|\mathbf{P}_t, x_t) \equiv \sum_{a=0}^{J} P_t(a) \; f(x_{t+1}|a, x_t).$$

- The Bellman equation of the probability-choice problem is:

$$V^P(x_t) = \max_{\mathbf{P}_t \in [0,1]^J} \left\{ \Pi^P(\mathbf{P}_t, x_t) + \beta \sum_{x_{t+1} \in \mathcal{X}} V^P_{t+1}(x_{t+1}) \; f^P(x_{t+1}|\mathbf{P}_t, x_t) \right\}$$

Equivalence of discrete-choice and probability-choice problems

- Define: $W(\mathbf{P}_t, x_t) \equiv \Pi^P(\mathbf{P}_t, x_t) + \beta \sum_{x_{t+1}} V^P(x_{t+1})\, f^P(x_{t+1}|\mathbf{P}_t, x_t)$.

PROPOSITION 2.

(A) the optimal decision in probability-choice problem is equal to the optimal choice probability in discrete choice problem, i.e., $\mathbf{P}^*(x_t) = \Lambda\left(\widetilde{\mathbf{v}}(x_t)\right)$.

(B) $W_t(\mathbf{P}_t, x_t)$ *is twice continuously differentiable and globally concave in* $\mathbf{P}_t$*, and the optimal decision rule* $\mathbf{P}^*(x_t)$ *is uniquely characterized by* $\partial W(\mathbf{P}^*, x_t)/\partial \mathbf{P}_t = 0$;

2. Euler Equations in DDC Models

Euler Equations in DDC models

- **Constrained optimization problem:**

$$\max_{\{\mathbf{P}_t,\mathbf{P}_{t+1}\}} \quad \Pi^P(\mathbf{P}_t, x_t) + \beta \sum_{x_{t+1}} \Pi^P(\mathbf{P}_{t+1}, x_{t+1}) \; f^P(x_{t+1} \mid \mathbf{P}_t, x_t)$$

$$\text{subject to:} \quad \begin{array}{l} f_{(2)}^P(x_{t+2} \mid \mathbf{P}_t, \mathbf{P}_{t+1}, x_t) = f^P(x_{t+2} \mid \mathbf{P}_t^*, \mathbf{P}_{t+1}^*, x_t) \\ \text{for any } x_{t+2} \end{array}$$

- By construction, $\mathbf{P}_t^*$ and $\mathbf{P}_{t+1}^*$ is the unique solution to this constrained optimization problem.

- Lagrange marginal conditions of optimilatiy characterize the solution.

Necessary and Sufficient Condition for Euler Equations

- Let $\widetilde{\mathbf{F}}_{t+1}$ be the matrix with elements:

$$\widetilde{f}\left(x_{t+2} \mid a_{t+1}, x_{t+1}\right) \equiv f\left(x_{t+2} \mid a_{t+1}, x_{t+1}\right) - f\left(x_{t+2} \mid 0, x_{t+1}\right)$$

  where the columns correspond to all the values $x_{t+2} \in \mathcal{X}_{(2)}(x_t)$ leaving out one, and the rows correspond to all the values $(a_{t+1}, x_{t+1}) \in [A - \{0\}] \times \mathcal{X}_{(1)}(x_t)$.

- *PROPOSITION 4: The model has an Euler equation representation if and only if* $\widetilde{\mathbf{F}}_{t+1}$ *is full column rank.*

- Relationship with Arcidiacono & Miller's finite state dependence.

General Form of the Euler Equations

The form of the Euler equation:

$$\pi\left(a, x_t\right) - \Lambda^{-1}(a, P_t)$$

$$\beta \sum_{x_{t+1}} \left[\pi\left(0, x_{t+1}\right) + e(0, \mathbf{P}_{t+1}) - \sum_{x_{t+2}} \lambda^{*}(x_{t+2})\ f\left(x_{t+2} | 0, x_{t+1}\right)\right] \widetilde{f}\left(x_{t+1} | a, x_t\right.$$

Example: Dynamic Multi-armed bandit problem

- Endogenous state variable: $f(y_{t+1}|a_t, y_t) = 1\{y_{t+1} = a_t\}$.
- The Euler equations, for any choice a:

$$\begin{aligned} & \pi(a, x_t) + e(a, P(x_t)) + \beta\ E_t\left[\pi(0, a, z_{t+1}) + e(0, a, P(x_{t+1}))\right] \\ + \quad & \pi(0, x_t) + e(0, P(x_t)) + \beta\ E_t\left[\pi(0, 0, z_{t+1}) + e(0, 0, P(x_{t+1}))\right] \end{aligned}$$

- With extreme value unobservables:

$$\begin{aligned} & \pi(a, x_t) - \ln P(a|x_t) + \beta\ E_t\left[\pi(0, a, z_{t+1}) - \ln P(0|a, z_{t+1})\right] \\ = \quad & \pi(0, x_t) - \ln P(0|x_t) + \beta\ E_t\left[\pi(0, 0, z_{t+1}) - \ln P(0|0, z_{t+1})\right] \end{aligned}$$

3. Euler Fixed Point Operator: Contraction Property

Euler operator

- The Euler equation implies a fixed point mapping in the space of the vector of value differences $\widetilde{\mathbf{v}}$

$$\Gamma_{EE-v}(\widetilde{\mathbf{v}}) \equiv \{\Gamma_{EE-v}(a, \mathbf{x}_t, \widetilde{\mathbf{v}}) : (a, \mathbf{x}_t) \in (\mathcal{A} - \{0\}) \times \mathcal{X}\}$$

where $\Gamma_{EE-v}(a, x_t, \widetilde{\mathbf{v}}) \equiv$

$$\widetilde{\pi}\,(a, x_t) + \beta \sum_{x_{t+1}} \left[\pi\,(0, x_{t+1}) + e(0, \Lambda\,(\widetilde{v}(x_{t+1}))) - \overline{\lambda^*}(x_{t+1}, \Lambda\,(\widetilde{\mathbf{v}})) \right] \widetilde{f}($$

Euler operator: Multi-armed bandit

- The Euler operator $\Gamma_{EE-v}(a, \mathbf{x}_t, \widetilde{\mathbf{v}})$ is:

$$\widetilde{\pi}(a, \mathbf{x}_t) + \beta \ \mathbb{E}_{\mathbf{z}_{t+1}|\mathbf{z}_t} \left[\pi\left(0, a, \mathbf{z}_{t+1}\right) - \pi\left(0, 0, \mathbf{z}_{t+1}\right)\right] +$$

$$\beta \ \mathbb{E}_{\mathbf{z}_{t+1}|\mathbf{z}_t} \left[\ln\left(1 + \sum_{j=1}^{J} \exp\{\widetilde{v}(j, 0, \mathbf{z}_{t+1})\}\right) - \ln\left(1 + \sum_{j=1}^{J} \exp\{\widetilde{v}(j, a, \mathbf{z}_{t\cdot}\right.\right.$$

Euler operator: Contraction

- We show that:

  1. Γ_{EE-v} *is a contraction mapping; its unique fixed point is the solution to the DP;*

  2. The contraction is stronger than value function iterations;

  3. A single evaluation of Γ_{EE-v} *is cheaper than a single value (or relative value) iteration.*

4. Sample-based Euler operator: Estimators

Sample-Based Euler Operator

- It is the sample counterpart of the Euler operator. We replace the conditional expectation at the population level $\mathbb{E}_{\{\mathbf{z}_{t+1}|\mathbf{z}_t\}}$ with its empirical counterpart $\mathbb{E}^{(N)}_{\{\mathbf{z}_{t+1}|\mathbf{z}_t\}}$.

- For the dynamic logit model,$\Gamma^{(N)}_{EE-v}(a, y, \mathbf{z}; \widetilde{\mathbf{v}})$ is:

$$[\pi(a, y, \mathbf{z}) - \pi(0, y, \mathbf{z})] + \beta\, \mathbb{E}^{(N)}_{\{\mathbf{z}'|\mathbf{z}_0\}} \left[\pi\left(0, a, \mathbf{z}'\right) - \pi\left(0, 0, \mathbf{z}'\right)\ \right]$$

$$\beta\, \mathbb{E}^{(N)}_{\{\mathbf{z}'|\mathbf{z}\}} \left[\ln\left(1 + \sum_{j=1}^{J} \exp\{\widetilde{v}_{t+1}(j, 0, \mathbf{z}')\}\right) - \ln\left(1 + \sum_{j=1}^{J} \exp\{\widetilde{v}_{t+1}(j, a,\right.\right.$$

- The dimension of the operator is given by the sample size and not by the dimension of the space of exogenous state variables $\mathbf{z}$.

Sample-Based Euler Operator: Properties

1. it is a contraction mapping; a stronger contraction than sample-based value function operator;

2. its evaluation does not suffer of any curse of dimensionality;

3. A fixed point of $\Gamma_{EE-v}^{(N)}(.;\theta)$ *is a* $\sqrt{N}-$*consistent and asymptotically normal estimator of the fixed point of* $\Gamma_{EE-v}(.;\theta)$

5. Numerical Examples

Data Generating Process

- Binary choice model of market entry-exit.

- Inactive firms get profit $\pi(0, \mathbf{x}_t) + \varepsilon_t(0)$, with $\pi(0, \mathbf{x}_t) = 0$, and active firms earn profit $\pi(1, \mathbf{x}_t) + \varepsilon_t(1)$, with:

$$\pi(1, \mathbf{x}_t) = VP_t - FC_t - EC_t * (1 - y_t)$$

and

$$\begin{aligned} VP_t &= \left[\theta_0^{VP} + \theta_1^{VP} z_{1t} + \theta_2^{VP} z_{2t}\right] \exp\left(\omega_t\right) \\ FC_t &= \theta_0^{FC} + \theta_1^{FC} z_{3t} \\ EC_t &= \theta_0^{EC} + \theta_1^{EC} z_{4t} \end{aligned}$$

Data Generating Process

Table 1
Parameters in the DGP

Payoff Parameters:	$\theta_0^{VP} = 0.5$; $\theta_1^{VP} = 1.0$; $\theta_2^{VP} = -1.0$
	$\theta_0^{FC} = 0.5$; $\theta_1^{FC} = 1.0$
	$\theta_0^{EC} = 1.0$; $\theta_1^{EC} = 1.0$
Each z_j state variable:	z_{jt} is AR(1), $\gamma_0^j = 0.0$; $\gamma_1^j = 0.6$
Productivity :	ω_t is AR(1), $\gamma_0^\omega = 0.2$; $\gamma_1^\omega = 0.9$
Low persistence model:	$\sigma_e^\omega = \sigma_e = 1$
High persistence model:	$\sigma_e^\omega = \sigma_e = 0.01$
Discount factor	$\beta = 0.95$

Comparing degree of contraction of operators

- Lipschitz constant of a fixed point operator $\Gamma(\mathbf{V}) = \{\Gamma(\mathbf{x}, \mathbf{V}) : \mathbf{x} \in \mathcal{X}\}$ is defined as:

$$
\begin{aligned}
L &\equiv \sup_{\mathbf{V},\mathbf{W}\in\mathbb{R}^{|\mathcal{X}|}} \left[\frac{\|\Gamma(\mathbf{V}) - \Gamma(\mathbf{W})\|}{\|\mathbf{V} - \mathbf{W}\|} \right] \\
&= \sup_{\mathbf{V},\mathbf{W}\in\mathbb{R}^{|\mathcal{X}|}} \left[\frac{\sup_{\mathbf{x}\in\mathcal{X}} |\Gamma(\mathbf{x}, \mathbf{V}) - \Gamma(\mathbf{x}, \mathbf{W})|}{\sup_{\mathbf{x}\in\mathcal{X}} |V(\mathbf{x}) - W(\mathbf{x})|} \right]
\end{aligned}
$$

Comparing contraction properties

Table 2
Degree of Contraction (Lipschitz)

	Low Persistence			**High Persistence**		
$\lvert\mathcal{X}\rvert$	**EE-v**	**VF**	**RVF**	**EE-v**	**VF**	**RVF**
64	0.20	0.95	0.59	0.34	0.95	0.95
486	0.18	0.95	0.54	0.34	0.95	0.95
2,032	0.18	0.95	0.53	0.31	0.95	0.95
6,250	0.18	0.95	0.53	0.32	0.95	0.95
15,552	0.18	0.95	0.53	0.28	0.95	0.95
200,000	0.18	0.95	0.53	0.28	0.95	0.95

Comparing computing times

Table 3(a)
Comparison of Solution Methods
(Model with low persistence)

# states	**Number iters.**				**Time per iter. (secs)**			
$\|\mathcal{X}\|$	EE-v	PF	VF	RVF	EE-v	PF	VF	RVF
64	13	5	351	36	$<$0.001	0.01	$<$0.001	$<$0.001
486	13	5	246	31	$<$0.001	0.60	$<$0.001	$<$0.001
2048	13	5	345	30	0.005	16.29	0.01	0.01
6250	13	5	344	30	0.04	150.0	0.08	0.08
15552	13	5	344	30	0.17	916.2	0.46	0.46
200,000	13	5	344	30	21.05	198,978	67.64	64.47

Comparing computing times

Table 3(b)
Comparison of Solution Methods
(Model with low persistence)

# states	**Total Time (in seconds)**				**Time Ratios**			
$\lvert\mathcal{X}\rvert$	EE-v	PF	VF	RVF	$\frac{\text{EE-v}}{\text{EE-v}}$	$\frac{\text{PF}}{\text{EE-v}}$	$\frac{\text{VF}}{\text{EE-v}}$	$\frac{\text{RVF}}{\text{EE-v}}$
64	<0.001	0.05	0.03	<0.001	1	60	37.0	4.0
486	0.01	3.02	0.35	0.03	1	300	26.6	3.4
2048	0.05	81.44	2.76	0.27	1	1320	53.1	5.2
6250	0.47	750.0	26.14	2.37	1	1,595	55.9	5.1
15552	2.25	4,581	158.9	13.89	1	2,040	70.7	6.2
200,000	273.6	$\simeq$ 1M*	23,270	1,934	1	3,630	85.0	7.1

Comparing computing times

Table 4
Comparison of EE-value and RVF Solution Methods
(Model with high persistence)

Number of states	**Euler**		**Relative value**		**Ratio total time**
$\lvert\mathcal{X}\rvert$	# iter.	Time-per-iter.	# iter.	Time-per-iter.	**RVF / EE-v**
64	24	<0.001	378	<0.001	24.5
486	24	<0.001	350	<0.001	30.4
2048	17	0.01	341	0.02	38.8
6250	18	0.06	322	0.12	38.0
15552	17	0.43	333	0.87	40.5
200,000	16	46.2	319	98.5	42.5

Finite Sample Properties: Parameter estimates

Table 5(c)
Monte Carlos: Estimation of Parameters
$N = 1,000$ & $T = 2$; Monte Carlo rep. $= 1,000$

	Root Mean Squared Error			
Parameter (True value)	2-step Eff. HM	2-step EE	MLE (NPL)	K-step EE
Total RMSE	0.703	0.935	0.680	0.684
Ratio $\frac{RMSE\ \text{HM}}{RMSE\ \text{EE}}$	0.75		0.99	
Time (in secs)	1067.80	0.218	7345.79	3.261
Ratio $\frac{Time\ \text{PF}}{Time\ \text{EE}}$	4898		2252	

Finite Sample Properties: Counterfactual

Table 6
Factual and Counterfactual Scenarios

	Prob. Being Active	Entry Prob.	Exit Prob.	State Persist.	Output
(A) Factual DGP	0.323	0.274	0.580	0.768	0.529
(B) Counterfactual DGP	0.258	0.157	0.513	0.884	0.423
Policy Effect: (B) - (A)	-0.065	-0.117	-0.068	+0.116	-0.106
(Percentage change)	(-20.1%)	(-42.7%)	(-11.7%)	(15.1%)	(-20.0%

Finite Sample Properties: Counterfactual

Table 8
Monte Carlo: Counterfactual Estimates

	Prob. Being Active	Entry Prob.	Exit Prob.	State Persist.	Total Output
True Policy Effect	-0.065	-0.117	-0.068	+0.116	-0.106
Root Mean Square Error					
VF iterations	67.5%	17.1%	30.1%	11.8%	82.8%
PF iterations	67.5%	17.1%	30.1%	11.8%	82.8%
EE-value iterations	37.1%	6.3%	14.8%	7.2%	38.6%

Conclusion

- Develop and study an alternative **EE fixed point mapping** which can be used to:
  - [1] Estimate parameters of a structural model;
  - [2] Obtain a solution of the model;
  - [3] Estimate counterfactual probabilities.
- Provide evidence that the EE-PI mapping is faster than the ST-PI mapping, and that the EE-PI mapping provides better estimates of counterfactual CPs when compared to alternative approximation methods that impose the same computational burden.