Estimation of DP Models

Computational Economics

Florian Oswald

December 9, 2021

Table of contents

Rust Bus Replacement

Berry, Levinsohn and Pakes (BLP) as MPEC

Estimation of Dynamic Programming Models

- Now that we know how to solve them, how do we estimate DP models?
- Examples
  - Rust (1987)
  - Berry et al. (1995)
- There are many different methods. We will introduce just a few. Look at the survey Aguirregabiria and Mira (2010) for more details.

Rust (1987)

- Each Bus comes in once a month for repair
  - Harold Zurcher decides after observing mileage x_t since last engine replacment **and** some other unobserved variable ε whether to replace or not:

$$u(x+t, d_t, \theta^c, RC) = \begin{cases} -c(x_t, \theta^c) & \text{if } d_t = 0 \\ -(RC + c(0, \theta^c)) & \text{if } d_t = 1 \end{cases}$$

  - He solves the DP

$$V_\theta(x_t) = \sup_{d_t} \mathbb{E} \left\{ \sum_{j=t}^{\infty} \beta^{j-t} u(x_j, d_j, \theta) + \varepsilon_t(d_t) | x_t \right\}$$

- Parameters to be estimated: $\theta = (\theta^c, RC, \theta^p)$
- This formulation results after making a set of simplifying assumptions.

Rust (1987)

- To simplify, the odometer progress is assumed to be a random process.
- that is, x_t evolves stochastically.
- The assumption is that $x_{t+1} \in \{s, s+1, s+2, s+3\}$ where s is the state of x_t, i.e. the bin it lies in.
- Move from one bin to the next with probabilities in θ^p.

Model and Data

- Data: a time series $\{x_t, d_t\}_{t=1}^T$
- Likelihood function is

$$\mathcal{L}(\theta) = \Pi_{t=2}^T P(d_t|x_t, \theta^c, RC)\pi(x_t|x_{t-1}, d_{t-1}, \theta^p)$$

where the conditional choice probabilities are given by

$$P(d_t|x_t, \theta^c, RC) = \frac{\exp[u(x, d, \theta^c, RC) + \beta EV_\theta(x, d)]}{\sum_{d' \in \{0,1\}} \exp[u(x, d', \theta^c, RC) + \beta EV_\theta(x', d')]}$$

and, importantly, EV is the solution to

$$\begin{aligned} EV_\theta(x, d) =& T_\theta(EV_\theta)(x, d) \\ \equiv& \int_{x'=0}^{\infty} \log \left(\sum_{d' \in 0,1} \exp[u(x, d', \theta^c, RC) + \beta EV_\theta(x', d')] \right) \end{aligned}$$

Nested Fixed Point Algorithm (NXFP)

1. **Outer Loop**: Solve the Likelihood function

$$\max_{\theta > 0} \mathcal{L}(\theta) = \Pi_{t=2}^{T} P(d_t | x_t, \theta^c, RC) \pi(x_t | x_{t-1}, d_{t-1}, \theta^p)$$

2. **Inner Loop**: Compute Expected value function EV_θ for a given guess θ

$$EV_\theta = T_\theta(EV_\theta)(x, d)$$

Potential Issues with NXFP

- We need a stopping rule for the likelihood function.
- We need one for the inner loop as well.
- Errors will propagate from the inner loop to the outer loop.
- Given that the search direction on $\mathcal{L}(\theta)$ depends on it's gradient, errors will matter a lot.
- the tolerance on the inner loop needs to be **tight**, like $1.0e^{-13}$

MPEC

Mathematical Programming with Equality constraints

- We can turn the problem around.
- Instead of asking *Whats the EV compatible with my guess θ?*, we could directly attack the likelihood:
- Maximize $\mathcal{L}(\theta)$ **subject to** the constraint, that behavior is **optimal** according to the model.
- in other words, augment the likelihood:

$$\mathcal{L}(\theta, EV; X) = \Pi_{t=2}^{T} P(d_t|x_t, \theta^c, RC)\pi(x_t|x_{t-1}, d_{t-1}, \theta^p)$$

$$P(d_t|x_t, \theta^c, RC) = \frac{\exp[u(x, d, \theta^c, RC) + \beta EV(x, d)]}{\sum_{d' \in \{0,1\}} \exp[u(x, d', \theta^c, RC) + \beta EV(x', d')]}$$

Different Optimization problems

NXFP solves the unconstrained optimization problem:

$$\max_{\theta} \mathcal{L}(\theta, EV_{\theta})$$

MPEC solves the constrained optimization problem:

$$\max_{\theta, EV} \mathcal{L}(\theta, EV; X)$$

$$\text{subject to} EV = T(EV, \theta)$$

Su and Judd (2012)

- Su and Judd (2012) perform MPEC on the bus model.
- the key difference to note is that EV now becomes a choice variable.
- In fact, the optimizer will be fed a vector

$$x = [RC, \theta^c, \mathbf{EV}]$$

where $\mathbf{EV}$ is an approximation to EV. In Su and Judd (2012), this is just going to be

$$\mathbf{EV} \equiv [EV(x_1), EV(x_2), \ldots, EV(x_n)]$$

i.e. the approximation needs to hold pointwise.

Su and Judd (2012)

TABLE II

NUMERICAL PERFORMANCE OF NFXP AND MPEC IN THE MONTE CARLO EXPERIMENTS[a]

β	Implementation	Runs Converged (out of 1250 runs)	CPU Time (in sec.)	# of Major Iter.	# of Func. Eval.	# of Contraction Mapping Iter.
0.975	MPEC/AMPL	1240	0.13	12.8	17.6	–
	MPEC/MATLAB	1247	7.90	53.0	62.0	–
	NFXP	998	24.60	55.9	189.4	134,748
0.980	MPEC/AMPL	1236	0.15	14.5	21.8	–
	MPEC/MATLAB	1241	8.10	57.4	70.6	–
	NFXP	1000	27.90	55.0	183.8	162,505
0.985	MPEC/AMPL	1235	0.13	13.2	19.7	–
	MPEC/MATLAB	1250	7.50	55.0	62.3	–
	NFXP	952	43.20	61.7	227.3	265,827
0.990	MPEC/AMPL	1161	0.19	18.3	42.2	–
	MPEC/MATLAB	1248	7.50	56.5	65.8	–
	NFXP	935	70.10	66.9	253.8	452,347
0.995	MPEC/AMPL	965	0.14	13.4	21.3	–
	MPEC/MATLAB	1246	7.90	59.6	70.7	–
	NFXP	950	111.60	58.8	214.7	748,487

[a]For each β, we use five starting points for each of the 250 replications. CPU time, number of major iterations, number of function evaluations and number of contraction mapping iterations are the averages for each run.

Performance

- In *general*, NXFP is a computationally expensive operation.
  - you have to solve a DP for many many many times in order to find your θ.
- However, there is much to qualify about this statement. The **details** matter here.
- For example, Su and Judd (2012) are very critical about NXFP in the Bus Model. They compare it to the performance of **MPEC**.
- But Iskhakov et al. (2016) redo the exercise with Rust's original method to solve EV and show that NXFP is still a very strong contender in this example.

Rust Bus Replacement

Berry, Levinsohn and Pakes (BLP) as MPEC

BLP after Dubé et al. (2012)

- Berry et al. (1995) is a model for automobile sales.
- It has become a very widely applied model and estimation technique, short: **BLP**.
- The original paper performs demand estimation with a large number of differentiated products:
  - characteristics approach
  - useful when only aggregate data are available
  - allows for flexible substitition patterns
  - controls for price endogeneity
- The computational algorithm derives moment conditions from a non-linear model
- The method is also known as **Random Coefficients Logit Demand**

Random Coefficients Logit Demand

- The Utility of i from purchasing product j in market t is

$$u_{ijt} = \beta_i^0 + x_{jt}\beta_i^x - \beta_i^p p_{jt} + \xi_{jt} + \varepsilon_{ijt} \tag{1}$$

- with product characteristics x_{jt}, p_{jt}, ξ_{jt}
  - x_{jt}, p_{jt}: ovserved with $cov(p_{jt}, \xi_{jt}) \neq 0$
  - ξ_{jt}: unovserved to econometrician.

- $\beta_i \equiv [\beta_i^0, \beta_i^x, \beta_i^p]$: random coefficients or individual specific tastes to be estimated.
  - We posit a distribution: $\beta_i \sim F_\beta(\beta, \theta)$
  - Goal of BLP: estimate θ in the above parametric distribution.
  - errors are assumed type 1 EV
  - Consumer picks product j if $u_{ijt} \geq u_{ij't}$

The Model: Market Shares

- The model predicts market shares:

$$s_j(x_t, p_t, x_{it}; \theta) = \int_{\{\beta_i, \varepsilon_i | u_{ijt} \geq u_{ij't}, \forall j' \neq j\}} dF_\beta(\beta, \theta) dF_\varepsilon(\varepsilon) \tag{2}$$

- with type 1 EV shocks ε, there is an analytical solution to one of those integrals:

$$s_j(x_t, p_t, x_{it}; \theta) = \int_\beta \frac{\exp(\beta^0 + x_{jt}\beta^x - \beta^p p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(\beta^0 + x_{kt}\beta^x - \beta^p p_{kt} + \xi_{kt})} dF_\beta(\beta, \theta \tag{3}$$

The Model: Market Shares

They use numerical integration:

$$\hat{s}_j(x_t, p_t, x_{it}; \theta) = \frac{1}{ns} \sum_{r=1}^{ns} \frac{\exp(\beta^{0r} + x_{jt}\beta^{xr} - \beta^{pr} p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(\beta^{0r} + x_{kt}\beta^{xr} - \beta^{pr} p_{kt} + \xi_{kt})} dF_{\beta}(\beta, \theta \tag{4}$$

to arrive at the market share (moment) conditions:

$$\hat{s}_j(x_t, p_t, \xi_{it}; \theta) = S_{jt}, \forall j \in J, t \in T \tag{5}$$

where S_{jt} is data.

GMM Estimator

- If firms can observe demand shocks ξ_t, they will set prices accordingly.
- There will be correlation between p_t and $\xi_t \Rightarrow$ Endogeneity Bias!
- BLP solve endogeneity of prices with a vector z_{jt} of IVs, which are **excluded** from the demand equation (1)
- they propose a moment condition $E[\xi_{jt}|z_{jt}, x_{jt}] = 0$
- z_{jt}: e.g. product-specific cost shifters, or K non-price characteristics in $x_{j,t}$ (assumed mean independent of ξ_t)
- We often form $E[\xi_{jt} \cdot h(z_{jt}, x_{jt})] = 0$ for some known function h.

Getting moment equations

- To get the sample analog of $E[\xi_{jt}|z_{jt}, x_{jt}] = 0$, we need to find ξ_{jt} corresponding to θ
- System (5) defines a mapping ξ_{jt} and S_t
- Berry proved that s has an inverse, hence any observed S_t can be explained by a **unique** $\xi_{jt}(\theta) = s^{-1}(S_t, \theta)$
- Sample analog of $E[\xi_{jt}|z_{jt}, x_{jt}] = 0$ is thus

$$g(\theta) = \frac{1}{TJ} \sum_{t,j} \xi_{jt}(\theta)' z_{jt}$$

GMM Estimator

- Data are $\{(x_{jt}, p_{jt}, S_{jt}, z_{jt})_{j \in J, t \in T}\}$
- We want to minimize the GMM objective

$$Q(\theta) = g(\theta)' W g(\theta)$$

- There is no analytic form for $\xi_{jt}(\theta)$, see previous slide

Berry et al. (1995) Estimation Algorithm - NFXP

- **Outer** Loop: $\min_{\theta} g(\theta)'Wg(\theta)$
  1. Guess vector θ to get $g(\theta) = \frac{1}{TJ}\sum_t \sum_j \xi_{jt}(\theta)'z_{jt}$
  2. Stop whenever $||\nabla_{\theta}(g(\theta)'Wg(\theta))|| \leq \epsilon_{\text{out}}$

Berry et al. (1995) Estimation Algorithm - NFXP

- **Outer** Loop: $\min_{\theta} g(\theta)' W g(\theta)$
  1. Guess vector θ to get $g(\theta) = \frac{1}{TJ} \sum_t \sum_j \xi_{jt}(\theta)' z_{jt}$
  2. Stop whenever $||\nabla_{\theta}(g(\theta)' W g(\theta))|| \leq \epsilon_{\text{out}}$
- **Inner** loop: compute $\xi_{jt}(\theta)$ given θ
  1. Solve system $s_j(x_t, p_t, x_{it}; \theta) = S_{\cdot t}$ by **Berry** constraction:

$$\xi_t^{h+1} = \xi_t^h + \log S_t - \log s_j(x_t, p_t, \xi_t^h; \theta)$$

  2. Stop whenever $||\xi_{\cdot t}^{h+1} - \xi_{\cdot t}^h|| \leq \epsilon_{\text{in}}$
  3. Call resulting demand shock $\xi_{jt}(\theta, \epsilon_{\text{in}})$
- Clearly, need to choose **both stopping rules** for inner and outer loop.

Knittel and Metaxoglou (2014)

- They perform an extensive investigation into BLP on two widely used datasets: cars and cereals.
- They use 10 free solvers and 50 starting points for each.
- Find: convergence occurs at several local extrema, saddles, and in regions where the FOC is not satisfied.
- Resulting parameter estimates of economic variables (market shares, price elasticiteis) exhibit **huge** variation depending on solver/starting point.
- All in all, they found 400 local solutions.

Dubé et al. (2012)'s concerns

1. Too much computation
   - need to know $\xi(\theta)$ only at true θ.
   - NFXP solves for $\xi(\theta)$ at each stage.
2. Stopping criteria
   - inner loop can be slow to converge
   - it's tempting to loosen ϵ_{in} (often see $\epsilon_{\text{in}} = 1e^{-6}$ or higher!)
   - outer loop may not converge with loose inner criterion
3. Inner loop error propagates to outer loop.

Errors from loose stopping

$$
\begin{aligned}
\theta^* &= \arg\max_{\theta} Q(\xi(\theta, 0)) \\
\hat{\theta} &= \arg\max_{\theta} Q(\xi(\theta, \epsilon_{\text{in}}))
\end{aligned}
$$

- Dubé et al. (2012) derive bounds on the order of estimatin error as a function of ϵ_{in}
- Consider Knittel and Metaxoglou (2014) for numerical experiments.

BLP as an MPEC

- Dubé et al. (2012) cast this as an MPEC:

$$\min_{\theta,\xi} \xi^T Z W Z^T \xi$$

$$\text{subject to } s(\xi, \theta) = S$$

- Advantages:

1. No need to set up 2 tolerances
2. no inner errors propagated
3. easy to code in AMPL
4. fewer iterations, given that AMPL provides analytic gradients/hessian

Exploring Sparsity in BLP

- The way this is formulated now, the Hessian of objective is dense. :-(

- They add an additional variable r and associated constraint $Z^T \xi = r$

$$\min_{\theta,\xi,r} r^T W r$$
$$\text{subject to } s(\xi,\theta) = S$$
$$\text{and } Z^T \xi = r$$

- advantages:

  1. Hessian of objective function is now sparse
  2. Very big saving in memory requirements.

Convergence and Loose vs Tight

TABLE II

THREE NFP IMPLEMENTATIONS: VARYING STARTING VALUES FOR NEVO'S CEREAL DATA SET, WITH CLOSED-FORM DERIVATIVES[a]

	NFP Loose Inner	NFP Loose Both	NFP Tight Both	NFP Tight Simplex
Fraction reported convergence	0.0	0.76	1.00	1.00
Frac. obj. fun. < 1% greater than "global" min.	0.0	0.0	1.00	0.0
Mean own-price elasticity across all runs	−3.82	−3.69	−7.43	−3.84
Std. dev. own price elasticity across all runs	0.4	0.07	∼0	0.35
Lowest objective function value	0.00213	0.00683	0.00202	0.00683
Elasticity for run with lowest obj. value	−6.71	−3.78	−7.43	−3.76

[a] We use the same 50 starting values for each implementation. The NFP loose inner-loop implementation has $\epsilon_{in} = 10^{-4}$ and $\epsilon_{out} = 10^{-6}$. The NFP loose-both implementation has $\epsilon_{in} = 10^{-4}$ and $\epsilon_{out} = 10^{-2}$. The NFP-tight implementation has $\epsilon_{in} = 10^{-14}$ and $\epsilon_{out} = 10^{-6}$. The Nelder–Meade or simplex method uses a tighter inner-loop tolerance of $\epsilon_{in} = 10^{-14}$ and MATLAB's default values for the simplex convergence criteria. We manually code closed-form derivatives for all methods other than for Nelder–Meade, which does not use derivative information.

Speed

TABLE III

MONTE CARLO RESULTS VARYING THE LIPSCHITZ CONSTANT[a]

Intercept $E[\beta_i^0]$	Lipsch. Const.	Imple.	Runs Converged	CPU Time (min)	Elasticities Bias	Elasticities RMSE	Elasticities Truth	Outside Share
−2	0.891	NFP-tight	100%	21.7	−0.077	0.14	−10.4	0.91
		MPEC	100%	18.3	−0.076	0.14		
−1	0.928	NFP-tight	100%	28.3	−0.078	0.15	−10.5	0.86
		MPEC	100%	16.3	−0.077	0.15		
0	0.955	NFP-tight	100%	41.7	−0.079	0.16	−10.6	0.79
		MPEC	100%	15.2	−0.079	0.16		
1	0.974	NFP-tight	100%	71.7	−0.083	0.16	−10.7	0.69
		MPEC	100%	11.8	−0.083	0.16		
2	0.986	NFP-tight	100%	103	−0.085	0.17	−10.8	0.58
		MPEC	100%	13.5	−0.085	0.17		
3	0.993	NFP-tight	100%	167	−0.088	0.17	−11.0	0.46
		MPEC	100%	10.7	−0.088	0.17		
4	0.997	NFP-tight	100%	300	−0.091	0.16	−11.0	0.35
		MPEC	100%	12.7	−0.090	0.16		

[a]There are 20 replications for each experiment and reported means are across these 20 replications. Each replication uses five starting values to do a better job at finding a global minimum. For each NFP starting value, we run the inner-loop once and use this vector of demand shocks and mean taste parameters as starting values for MPEC. The NFP-tight implementation has $\epsilon_{in} = 10^{-14}$ and $\epsilon_{out} = 10^{-6}$. There is no inner-loop in MPEC; $\epsilon_{out} = 10^{-6}$ and $\epsilon_{feasible} = 10^{-6}$. The same 1000 simulation draws are used to generate the data and to estimate the model. NFP and MPEC use closed-form first- and second-order derivatives. We supply the sparsity patterns of the constraints and derivatives to the optimization routine for both methods.

References I

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, 55(5):999–1033, 1987. ISSN 00129682. URL `http://www.jstor.org/stable/1911259`.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

Victor Aguirregabiria and Pedro Mira. Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, 156(1):38 – 67, 2010. ISSN 0304-4076. doi: DOI:10.1016/j.jeconom.2009.09.007. URL `http://www.sciencedirect.com/science/article/B6VC0-4X7R85S-3/2/b78d013f2b1f3dbce3f2325f0c41d4e4`. Structural Models of Optimization Behavior in Labor, Aging, and Health.

References II

Che-Lin Su and Kenneth L. Judd. Constrained optimization approaches to estimation of structural models. *Econometrica*, 80 (5):2213–2230, 2012. ISSN 1468-0262. doi: 10.3982/ECTA7925. URL `http://dx.doi.org/10.3982/ECTA7925`.

Fedor Iskhakov, Jinhyuk Lee, John Rust, Bertel Schjerning, and Kyoungwon Seo. Comment on "constrained optimization approaches to estimation of structural models". *Econometrica*, 84 (1):365–370, 2016.

Jean-Pierre Dubé, Jeremy T Fox, and Che-Lin Su. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica*, 80 (5):2231–2267, 2012.

Christopher R Knittel and Konstantinos Metaxoglou. Estimation of random-coefficient demand models: Two empiricists' perspective. *Review of Economics and Statistics*, 96(1):34–59, 2014.