Discrete Choice - PyBLP

Data Skills for Research
Kellogg Research Support

July 10, 2024

Northwestern | Kellogg

Roadmap

- Why use a BLP Framework?
  - What does it solve?
  - What are the limitations?

- PyBLP to the Rescue
  - Tutorials

- Using PyBLP on KLC
  - Running Tutorials
  - Using Knitro as an Optimization Routine

What is BLP?

Econometrica, Vol. 63, No. 4 (July, 1995), 841–890

AUTOMOBILE PRICES IN MARKET EQUILIBRIUM

BY STEVEN BERRY, JAMES LEVINSOHN, AND ARIEL PAKES[1]

- Models based on Berry, Levinsohn and Pakes (1995) and related papers have become 'workhorse' models in Empirical Industrial Organization, Marketing, Environmental Economics, Economics of Education, ….

- What do these models do?
  - Discrete choice models of differentiated product demand
  - Tackle Price Endogeneity
  - Can generate realistic substitution patterns by modeling unobserved attributes / quality and allowing for preference heterogeneity
  - Allow computation of counterfactuals for policy analysis (mergers, tariffs, regulation,…)

Dataset – Automobile Market 1970-1990

model_year	id	firmid	market	hpwt	space	air	mpd	price	mpg	share	share_out
71	129	15	1	0.528997	1.1502	0	1.888146	4.935802	1.697	0.001051	0.880106
71	130	15	1	0.494324	1.278	0	1.935989	5.516049	1.74	0.00067	0.880106
71	132	15	1	0.467613	1.4592	0	1.716799	7.108642	1.543	0.000341	0.880106
71	134	15	1	0.42654	1.6068	0	1.687871	6.839506	1.517	0.000522	0.880106
71	136	15	1	0.452489	1.6458	0	1.504286	8.928395	1.352	0.000442	0.880106
71	138	19	1	0.450871	1.6224	0	1.726813	7.153086	1.552	0.002756	0.880106
71	141	19	1	0.564002	1.768	0	1.727926	9.85679	1.553	0.002651	0.880106
71	143	19	1	0.731368	1.768	0	2.194121	11.269136	1.972	0.000487	0.880106
71	144	19	1	0.719014	1.816	0	2.056154	12.135802	1.848	0.002415	0.880106
71	145	19	1	0.728324	1.744	0	1.978269	12.97037	1.778	0.000551	0.880106
71	146	19	1	0.732484	1.808	0	1.918187	15	1.724	0.000127	0.880106
71	147	19	1	0.729387	1.808	0	1.912624	16.044444	1.719	0.002705	0.880106
71	148	19	1	0.780749	1.776	0	2.018324	18.22963	1.814	0.000534	0.880106
71	149	19	1	0.716511	1.832	0	1.892596	19.167901	1.701	0.000333	0.880106
71	150	19	1	0.419385	1.122	0	2.211923	5.160494	1.988	0.004005	0.880106
71	151	19	1	0.487231	1.387	0	1.802473	5.938272	1.62	0.003533	0.880106
71	153	19	1	0.451713	1.5352	0	1.682308	6.609877	1.512	0.00503	0.880106
71	158	19	1	0.468649	1.41	0	1.62	7.212346	1.456	0.001812	0.880106
71	160	19	1	0.702408	1.5732	0	1.604423	8.434568	1.442	0.002186	0.880106
71	161	19	1	0.388532	1.736	0	1.4175	7.644444	1.274	0.00038	0.880106
71	163	19	1	0.388532	1.736	0	1.4175	7.982716	1.274	0.000732	0.880106
71	165	19	1	0.385638	1.736	0	1.406374	8.37284	1.264	0.008658	0.880106
71	167	19	1	0.631188	1.736	0	1.463118	10.207407	1.315	0.002089	0.880106
71	169	19	1	0.843223	1.2627	0	1.872569	13.661728	1.683	0.000358	0.880106
71	170	16	1	0.659314	1.8	0	1.528764	10.345679	1.374	0.001038	0.880106
71	171	16	1	0.775283	1.8	0	1.65783	11.57284	1.49	0.00012	0.880106
71	172	16	1	0.77278	1.8	0	1.893709	12.123457	1.702	0.000339	0.880106
71	173	16	1	0.49505	1.0168	0	2.330975	4.925926	2.095	0.000426	0.880106
71	177	16	1	0.431034	1.379	0	1.949341	6.049383	1.752	0.003121	0.880106
71	179	16	1	0.413907	1.4784	0	1.654492	6.733333	1.487	0.00044	0.880106

Why BLP?

Berry, Levinsohn and Pakes (1995) logit results

AUTOMOBILE PRICES

TABLE III

RESULTS WITH LOGIT DEMAND AND MARGINAL COST PRICING
(2217 OBSERVATIONS)

Variable	OLS Logit Demand	IV Logit Demand	OLS ln (*price*) on *w*
Constant	−10.068	−9.273	1.882
	(0.253)	(0.493)	(0.119)
*HP/Weight**	−0.121	1.965	0.520
	(0.277)	(0.909)	(0.035)
Air	−0.035	1.289	0.680
	(0.073)	(0.248)	(0.019)
MP$	0.263	0.052	—
	(0.043)	(0.086)	
*MPG**	—	—	−0.471
			(0.049)
*Size**	2.341	2.355	0.125
	(0.125)	(0.247)	(0.063)
Trend	—	—	0.013
			(0.002)
Price	−0.089	−0.216	—
	(0.004)	(0.123)	
No. Inelastic Demands	1494	22	*n.a.*
(+/−2 *s.e.'s*)	(1429–1617)	(7–101)	
R^2	0.387	*n.a.*	.656

Notes: The standard errors are reported in parentheses.
*The continuous product characteristics—hp/weight, size, and fuel efficiency (*MP$* or *MPG*)—enter the demand equations in levels, but enter the column 3 price regression in natural logs.

This model predicts that the same price increase for a Tesla and a Yugo would equally decrease demand.

BLP fixes this!

TABLE VI

A SAMPLE FROM 1990 OF ESTIMATED OWN- AND CROSS-PRICE SEMI-ELASTICITIES: BASED ON TABLE IV (CRTS) ESTIMATES

	Mazda 323	Nissan Sentra	Ford Escort	Chevy Cavalier	Honda Accord	Ford Taurus	Buick Century	Nissan Maxima	Acura Legend	Lincoln Town Car	Cadillac Seville	Lexus LS400	BMW 735i
323	−125.933	1.518	8.954	9.680	2.185	0.852	0.485	0.056	0.009	0.012	0.002	0.002	0.000
Sentra	0.705	−115.319	8.024	8.435	2.473	0.909	0.516	0.093	0.015	0.019	0.003	0.003	0.000
Escort	0.713	1.375	−106.497	7.570	2.298	0.708	0.445	0.082	0.015	0.015	0.003	0.003	0.000
Cavalier	0.754	1.414	7.406	−110.972	2.291	1.083	0.646	0.087	0.015	0.023	0.004	0.003	0.000
Accord	0.120	0.293	1.590	1.621	−51.637	1.532	0.463	0.310	0.095	0.169	0.034	0.030	0.005
Taurus	0.063	0.144	0.653	1.020	2.041	−43.634	0.335	0.245	0.091	0.291	0.045	0.024	0.006
Century	0.099	0.228	1.146	1.700	1.722	0.937	−66.635	0.773	0.152	0.278	0.039	0.029	0.005
Maxima	0.013	0.046	0.236	0.256	1.293	0.768	0.866	−35.378	0.271	0.579	0.116	0.115	0.020
Legend	0.004	0.014	0.083	0.084	0.736	0.532	0.318	0.506	−21.820	0.775	0.183	0.210	0.043
TownCar	0.002	0.006	0.029	0.046	0.475	0.614	0.210	0.389	0.280	−20.175	0.226	0.168	0.048
Seville	0.001	0.005	0.026	0.035	0.425	0.420	0.131	0.351	0.296	1.011	−16.313	0.263	0.068
LS400	0.001	0.003	0.018	0.019	0.302	0.185	0.079	0.280	0.274	0.606	0.212	−11.199	0.086
735i	0.000	0.002	0.009	0.012	0.203	0.176	0.050	0.190	0.223	0.685	0.215	0.336	−9.376

Note: Cell entries i, j, where i indexes row and j column, give the percentage change in market share of i with a $1000 change in the price of j.

BLP Applications

- **Product Pricing:** Optimize prices based on market demand and competition, enhancing pricing strategies.
- **Market Competition**: Analyze competition, mergers, and entry decisions to predict market outcomes. (See Nevo, 2001a/2001b)
- **Policy Evaluation:** Assess policy impact on prices, consumer welfare, and industry performance. (See Bayer et al. 2007, Neilson 2017 for school choice examples)
- **New Product Launch**: Predict market share for new products, aiding launch strategies. (see Petrin, 2002)
- **Consumer Behavior:** Understand preferences across segments for targeted marketing.

So what's the problem? Nevo (2000)

A PRACTITIONER'S GUIDE TO ESTIMATION OF RANDOM-COEFFICIENTS LOGIT MODELS OF DEMAND

AVIV NEVO

University of California–Berkeley, Berkeley, CA 94720-3880 and NBER

Estimation of demand is at the heart of many recent studies that examine questions of market power, mergers, innovation, and valuation of new brands in differentiated-products markets. This paper focuses on one of the main methods for estimating demand for differentiated products: random-coefficients logit models. The paper carefully discusses the latest innovations in these methods with the hope of increasing the understanding, and therefore the trust among researchers who have never used them, and reducing the difficulty of their use, thereby aiding in realizing their full potential.

So what's the problem?

From the econometric model to pseudocode to code

- Outer loop: search over nonlinear parameters θ to minimize GMM objective:

$$\widehat{\theta_{BLP}} = \arg\min_{\theta_2} (Z'\hat{\xi}(\theta_2))W(Z'\hat{\xi}(\theta_2))'$$

- Inner Loop:
  - Fix a guess of $\widetilde{\theta}_2$.
  - Solve for $\boldsymbol{\delta_t}(\mathcal{S}_t, \widetilde{\theta}_2)$ which satisfies $\sigma_{jt}(\boldsymbol{\delta_t}, \widetilde{\theta}_2) = \mathfrak{s}_{jt}$.
    - Computing $s_{jt}(\boldsymbol{\delta_t}, \widetilde{\theta}_2)$ requires numerical integration (quadrature or monte carlo).
  - We can do IV-GMM to recover $\hat{\alpha}(\widetilde{\theta}_2), \hat{\beta}(\widetilde{\theta}_2), \hat{\xi}(\widetilde{\theta}_2)$.

$$\boldsymbol{\delta_t}(\mathcal{S}_t, \widetilde{\theta}_2) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

  - Use $\hat{\boldsymbol{\xi}}(\theta)$ to construct sample moment conditions $\frac{1}{N}\sum_{j,t} Z'_{jt}\xi_{jt}$
- When we have found $\hat{\theta}_{BLP}$ we can use this to update $W \to W(\hat{\theta}_{BLP})$ and do 2-stage GMM.

So what's the problem?

From the econometric model to pseudocode to code

(a) For each market t: solve $S_{jt} = s_{jt}(\delta_{\cdot t}, \theta_2)$ for $\widehat{\delta}_{\cdot t}(\theta_2)$.

(b) For each market t: use $\widehat{\delta}_{\cdot t}(\theta_2)$ to construct $\eta_{\cdot t}(\mathbf{q_t}, \mathbf{p_t}, \widehat{\delta}_{\cdot t}(\theta_2), \theta_2)$

(c) For each market t: Recover $\widehat{mc}_{jt}(\widehat{\delta}_{\cdot t}(\theta_2), \theta_2) = p_{jt} - \eta_{jt}(\widehat{\delta}_{\cdot t}(\theta_2), \theta_2)$

(d) Stack up $\widehat{\delta}_{\cdot t}(\theta_2)$ and $\widehat{mc}_{jt}(\widehat{\delta}_{\cdot t}(\theta_2), \theta_2)$ and use linear IV-GMM to recover $[\widehat{\theta}_1(\theta_2), \widehat{\theta}_3(\theta_2)]$ following the recipe on previous slide

(e) Construct the residuals:

$$\widehat{\xi}_{jt}(\theta_2) = \widehat{\delta}_{jt}(\theta_2) - [x_{jt}\ v_{jt}]\,\widehat{\beta}(\theta_2) + \alpha p_{jt}$$
$$\widehat{\omega}_{jt}(\theta_2) = f(\widehat{mc}_{jt}(\theta_2)) - [x_{jt}\ w_{jt}]\,\widehat{\gamma}(\theta_2)$$

(f) Construct sample moments

$$g_n^D(\theta_2) = \frac{1}{N}\sum_{jt} Z_{jt}^{D\prime}\widehat{\xi}_{jt}(\theta_2)$$
$$g_n^S(\theta_2) = \frac{1}{N}\sum_{jt} Z_{jt}^{S\prime}\widehat{\omega}_{jt}(\theta_2)$$

(g) Construct GMM objective $Q_n(\theta_2) = \begin{bmatrix} g_n^d(\theta_2) \\ g_n^s(\theta_2) \end{bmatrix}' W \begin{bmatrix} g_n^d(\theta_2) \\ g_n^s(\theta_2) \end{bmatrix}$

So what's the problem? K-M 2014 (!)

ESTIMATION OF RANDOM-COEFFICIENT DEMAND MODELS: TWO EMPIRICISTS' PERSPECTIVE

Christopher R. Knittel and Konstantinos Metaxoglou*

Abstract—We document the numerical challenges we experienced estimating random-coefficient demand models as in Berry, Levinsohn, and Pakes (1995) using two well-known data sets and a thorough optimization design. The optimization algorithms often converge at points where the first- and second-order optimality conditions fail. There are also cases of convergence at local optima. On convergence, the variation in the values of the parameter estimates translates into variation in the models' economic predictions. Price elasticities and changes in consumer and producer welfare following hypothetical merger exercises vary at least by a factor of 2 and up to a factor of 5.

PyBLP to the Rescue

- Paper by Conlon and Gortmaker (2020) with accompanying python package
- Expansive framework that nests many canonical BLP-style papers
  - Consistent notation helps understand differences between models
  - Very modular – elements such as supply-side restrictions, micro moments can be added as appropriate
- Using CS-inspired best practices on various subproblems appears to allay many of the Knittel-Metaxoglou (2014) concerns
- Choice of free (SciPy) or commercial (Artelys Knitro) optimization routines

PyBLP Modified Tutorials

- PyBLP offers tutorials here: https://pyblp.readthedocs.io/en/stable/tutorial.html
- In the training session, we will go through the logit and random coefficients tutorials using the supply and demand side data from the original BLP (1995) dataset.
- We will also demonstrate how to apply Artelys Knitro as an optimization solver on KLC

Creating a PyBLP Conda Environment

The PyBLP website provides explicit instructions for installing the package. Unlike most package installations, the environment needs to have certain related packages already installed before `pyblp` will install properly. To create a PyBLP conda environment, first load mamba:

```
module load mamba/23.1.0
```

Then use mamba to install a clean python environment:

```
mamba create -n pyblp_env python=3.10
```

Activate the environment and load the following:

```
source activate pyblp_env
```

Install the related packages:

```
conda install -c conda-forge numpy scipy sympy patsy notebook
   pandas statsmodels
module load gcc/11.2.0 # this is used for statsmodels

pip install --upgrade pyhdfe
```

Finally, install pyblp:

```
pip install pyblp
```

Activating a Conda Environment

To activate the environment in the future, either load mamba:

```
module load mamba/23.1.0
```

OR any version of conda:

```
module load python-anaconda3/2019.10
```

Then run this line to activate the environment:

```
source activate pyblp_env
```

To leave the environment:

```
source deactivate pyblp_env
```

To output the yaml file so you can share your environment with others:

```
conda env export > pyblp_env.yml
```

Alternatively, you can also use a pyblp environment we created:

```
module load mamba/23.1.0
source activate /kellogg/software/envs/pyblp_env
```

Jupyter Notebooks on KLC

Recall that we already installed jupyter notebooks into our PyBLP environment.

We need to perform a one-time installation of python kernel to be able to use the pyblp_env in jupyter notebook on KLC:

```
python -m ipykernel install --user --name pyblp_env --display-name "Python (pyblp_env)"
```

To open a notebook on KLC, we recommend using the default Firefox web browser that is already loaded on KLC for you. Just type:

```
jupyter notebook --browser=firefox
```

Jupyter Notebooks on Quest's Jupyterlab

If you have a Quest allocation, you can open the jupyter notebooks in your home directory by visting the URL:

https://jupyter.questanalytics.northwestern.edu

Remember to change Kernel to "Python (pyblp_env)"

PyBLP with Knitro

Once you activate your pyblp environment:

```
source activate /kellogg/software/envs/pyblp_env
```

Load a Knitro module on KLC:

```
module load knitro/12.4
```

Now you can run your pyblp while using Knitro as a solver:

```
python blp_optimizer.py --opt_method knitro
```

NUIT Knitro guide for Quest here:
https://services.northwestern.edu/TDClient/30/Portal/KB/ArticleDet?ID=1696

Runtime

- `python blp_optimizer.py`
- Default: using nested fixed-point (nfxp) optimizer
- 112.60 seconds

- `python blp_optimizer.py --opt_method knitro`
- Using the Knitro optimizer
- 48.93 seconds

PyBLP Optimization Routine Comparison

Criteria	Knitro (KN)	NFXP (Nested Fixed Point)	BFGS (Broyden-Fletcher-Goldfarb-Shanno)
Definition	Commercial solver with advanced algorithms for nonlinear optimization.	Iterative method for estimating equilibrium models with alternating fixed-point computations.	Gradient-based algorithm for unconstrained optimization problems.
Benefits	- Advanced techniques for various problems. - Handles large-scale nonlinear problems. - Global and local optimization.	- Specifically for equilibrium models. - Suited for nested demand and supply problems. - Converges to an equilibrium.	- Efficient for smooth unconstrained problems. - No derivatives needed. - Suitable for simpler models.
Drawbacks	- Commercial (may require license). - May need tuning. - Complex setup.	- More iterations to converge. - Limited to specific equilibrium structures. - May require adjustments.	- Not for constrained or non-smooth problems. - Local convergence. - Initial conditions matter.
Speed	Variable; depends on problem complexity.	Moderate; iterative process takes time.	Fast; quick convergence for smooth problems.