Introduction to Demand Estimation w/ `pyblp`

Chris Conlon and Jeff Gormaker

NYU Stern and NY Fed

January 13, 2019

Introduction

An Ongoing Project...

What do we have so far?

- Available on `PyPI`
  `pip install pyblp`
- Extensive documentation: `https://pyblp.readthedocs.io/en/stable/`
- Long list of features
- 6k downloads: who are these people?

Outline of Workshop

1. Review of Theory of Demand Estimation/BLP.
2. Logit and Nested Logit with `pyblp`.
3. Nevo (2000) and BLP (1995) with `pyblp`.
   BREAK?
4. What do we do with demand estimates? (Post-Estimation).
5. Under the Hood
6. Best Practices → some new econometric properties?

Each consumer i chooses a product j to maximize utility u_{ij}:

$$y_i = \arg\max_j u_{ij}$$

Begin with the basic logit model where ε I.I.D Type I EV (Gumbel):

$$u_{ij} = \beta x_j - \alpha p_j + \varepsilon_{ij}$$

Goal is to use data on observed choices (y_{ij}, x_j) to estimate (α, β).

Logit is convenient for two closed forms:

1. Expectation of the maximum

$$E_{\varepsilon}[\max_j u_{ij}] = \log\left(\sum_j \exp[\beta x_j - \alpha p_j]\right)$$

2. Choice Probability

$$s_j = \frac{\exp[\beta x_j - \alpha p_j]}{1 + \sum_k \exp[\beta x_k - \alpha p_k]}$$

Estimation is straightforward via Maximum Likelihood (MLE):

$$
\begin{aligned}
L(\mathbf{y}|\mathbf{x},\theta) &= \prod_{i=1}^{N} \underbrace{\frac{n_i!}{\prod_{j=1}^{J} y_{ij}!}}_{C(\mathbf{y})} \prod_{j=1}^{J} s_{ij}(x_{ij},\theta)^{y_{ij}} \\
ll(\mathbf{y}|\mathbf{x},\theta) &= \sum_{i=1}^{N} \log(C(\mathbf{y})) + \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij}\log(s_{ij}(x_{ij},\theta)) \\
l(\mathbf{y}|\mathbf{x},\theta) &\approx \sum_{i=1}^{N}\sum_{j=1}^{J} y_{ij}\log(s_{ij}(x_{ij},\theta))
\end{aligned}
$$

- We can ignore the combinatorial term (with the factorials) since it does not affect the location of the maximum (it is additive and doesn't depend on θ).

Estimation is just like before

- Suppose that all consumers had the same $x_{ij} = x_j$ (Choices depended only on products not on income, education, etc.)
- We can construct $y_j^* = \sum_{i=1}^N y_{ij}$.

$$l(\mathbf{y}|\mathbf{x},\theta) \approx \sum_{j=1}^J y_j^* \log(s_j(\mathbf{x},\theta))$$

- When each consumer i faces the same choice environment, we can aggregate data into sufficient statistics.

Heterogeneity and Endogeneity: Separately

- Suppose that the number of individuals $N \to \infty$.
- Suppose we observe aggregate sales $q_j = N\mathbf{s}_j$ where q_j is finite.
- Imagine I observe $\mathbf{s}_j$ without any measurement error
  - Usual variance $Var(q_j) = Ns_j(1 - s_j)$ and $Var(\frac{q_j}{N}) = \frac{1}{N}s_j(1 - s_j)$

- Take log of choice probabilty

$$s_j = \frac{\exp[\beta x_j - \alpha p_j]}{1 + \sum_k \exp[\beta x_k - \alpha p_k]} \quad s_0 = \frac{1}{1 + \sum_k \exp[\beta x_k - \alpha p_k]}$$

$$\log s_j = \beta x_j - \alpha p_j - \underbrace{\log\left(1 + \sum_k \exp[\beta x_k - \alpha p_k]\right)}_{-\log s_0}$$

- And we get the linear estimating equation

$$\underbrace{\log s_j - \log s_0}_{\text{data!}} = \beta x_j - \alpha p_j + \xi_j$$

- This is why logit/multinomial logit often referred to as generalized linear model (GLM).
- Also if you needed to instrument you already know how!

We relax the IIA property by mixing over various logits:

$$\begin{aligned} u_{ij} &= x_j\beta + \mu_{ij} + \varepsilon_{ij} \\ s_{ij} &= \int \frac{\exp[x_j\beta + \mu_{ij}]}{1 + \sum_k \exp[x_k\beta + \mu_{ik}]} f(\mu_i|\theta) \end{aligned}$$

- Each individual draws a vector μ_i of μ_{ij} (separately from ε).
- Conditional on μ_i each person follows an IIA logit model.
- However we integrate (or mix) over many such individuals giving us a mixed logit or heirarchical model (if you are a statistician)
- In practice these are not that different from linear random effects models you have learned about previously.
- It helps to think about fixing μ_i first and then integrating out over ε_i

Mixed Logit: Interpretations

An alternative is to allow for individuals to have random variation in β_i:

$$U_{ij} = \beta_i x_{ij} + \varepsilon_{ij}$$

Which is the random coefficients formulation.
As an alternative, we could have specified an error components structure on ε_i.

$$U_{ij} = \beta x_{ij} + \underbrace{\nu_i z_{ij} + \varepsilon_{ij}}_{\tilde{\varepsilon}_{ij}}$$

- The key is that ν_i is unobserved and mean zero. But that x_{ij}, z_{ij} are observed per usual and ε_{ij} is IID Type I EV.
- This allows for a heteroskedastic structure on ε_i, but only one which we can project down onto the space of z.

Heterogeneity and Endogeneity: Together

Putting it Together

- Now we want to have both price endogeneity and flexible substitution in the same model.
- We are ultimately going with the random coefficients logit model, but we will start with the logit and nested logit.

Basic Idea from Price Endogeneity

$$s_{jt} = \int \frac{\exp[x_{jt}\beta_i]}{1 + \sum_k \exp[x_{kt}\beta_i]} f(\beta_i|\theta)$$

- We know prices are set with demand in mind and this can create an endogeneity problem.
- How do we deal with it?
- We would like to instrument in this world but what is the error term exactly?
- An obvious choice might be $\eta_{jt} = (s_{jt}(\theta) - \tilde{s}_{jt})$
- Can we find things that are orthogonal to the error between observed and predicted market shares?
- Do we have the usual IV conditions (exogeneity, relevance, monotonicity, etc.)

Basic Idea from Price Endogeneity

- We need to add an unobservable quality term ξ_{jt} to our model

$$
\begin{aligned}
u_{ijt} &= x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ij} \\
s_{jt} &= \int \frac{\exp[x_{jt}\beta_i + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta_i + \xi_{kt}]} f(\beta_i|\theta)
\end{aligned}
$$

- The idea is that ξ_{jt} is observed to the firm when prices are set, but not to us the econometricians.
- We call ξ_{jt} a vertical component, because all consumers agree on its value.
- This allows for products j to better than some other product in a way that is not fully explained by differences in x_j and x_k.
- Basically there is something about a BMW that makes it better than a Peugeot in a way that is not fully captured by its mileage, weight, horsepower, etc. that leads to it having higher sales and/or higher prices.

Inversion: IIA Logit

- Think about the plain IIA logit for a minute:

$$
\begin{aligned}
u_{ijt} &= x_{jt}\beta + \xi_{jt} + \varepsilon_{ij} \\
s_{jt} &= \frac{\exp[x_{jt}\beta + \xi_{jt}]}{1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]}
\end{aligned}
$$

- Take logs

$$
\begin{aligned}
\ln s_{ot} &= -\log\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right) \\
\ln s_{jt} &= [x_{jt}\beta + \xi_{jt}] - \log\left(1 + \sum_k \exp[x_{kt}\beta + \xi_{kt}]\right) \\
\ln s_{jt} - \ln s_{ot} &= x_{jt}\beta - \alpha p_{jt} + \xi_{jt}
\end{aligned}
$$

Inversion: IIA Logit

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{data!} = \underbrace{x_{jt}\beta - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}}$$

- The LHS is data! The RHS is now a linear IV problem!
- α the price coefficient is the endogenous parameter.
- We know how to solve this. We need instruments that shift p_{jt} but are orthogonal to ξ_{jt}.
- Economic theory tells us how: cost shifters, markup shifters.
- Markups in IIA logit are pretty boring since they only depend on your shares and α.
- If number of products varies across markets, that works. Otherwise you want cost shifters in cross section or time series.

Was that magic?

- No. It was just a nonlinear change of variables from $\eta_{jt} \to \xi_{jt}$.
- Our moment condition is just that $E[\xi_{jt}|x_{jt}, z_{jt}] = 0$.
- We moved from the space of shares and MLE for the logit to the space of utilities and an IV model.
- We are losing some efficiency – but now we are able to estimate under weaker conditions.

CAVEATS

- We do need a technical condition. This only works if the market size $N \to \infty$.
- That is our data/shares we must believe we are observing without any sampling error.
- This is not necessary for the multinomial MLE where shares have some natural sampling variation.
- In our IV/GMM approach we cannot have this sampling error. (Why?).

Inversion: Nested Logit (Berry 1994 / Cardell 1991)

This takes a bit more algebra but not much

$$\underbrace{\ln s_{jt} - \ln s_{0t}}_{data!} = x_{jt}\beta - \alpha p_{jt} - \sigma \underbrace{\log(s_{j|gt})}_{data!} + \xi_{jt}$$

- Same as logit plus an extra term $\log(s_{j|g})$ the within group share.
- We now have a second endogenous parameter.
- If you don't see it – realize we are regressing Y on a function of Y. This should always make you nervous.
- If you forget to instrument for σ you will get $\sigma \to 1$ because of attenuation bias.
- A good instrument for σ is the number of products within the nest. Why?

We can't solve for δ_{jt} directly this time. We often exploit a trick when β_i, ν_i is normally distributed:

$$s_{jt} = \int \frac{\exp[\delta_{jt} + x_{jt} \cdot \Sigma \cdot \nu_i]}{1 + \sum_k \exp[\delta_{kt} + x_{kt} \cdot \Sigma \cdot \nu_i]} f(\nu_i|\theta)$$

- This is a $J \times J$ system of equations for each t.
- It is diagonally dominant.
- There is a unique vector ξ_t that solves it for each market t.
- If you can work out $\frac{\partial s_{jt}}{\partial \delta_{kt}}$ (easy) you can solve this using Newton's Method.

Contraction: BLP

BLP actually propose an easy solution to find δ_t. Fix θ and solve for δ. Think about doing this one market at a time:

$$\delta^{(k)}(\theta) = \delta^{(k-1)}(\theta) + \log(\tilde{s}_j) - \log(s_j(\delta_t^{(k-1)}, \theta)$$

- They prove (not easy) that this is a contraction mapping.
- If you keep iterating this equation enough $|\delta^{(k)}(\theta) - \delta^{(k-1)}(\theta)| < \epsilon_{tol}$ you can recover the δ's so that the observed shares and the predicted shares are identical.
- Practical tip: ϵ_{tol} needs to be as small as possible. ($\approx 10^{-13}$).
- Practical tip: Contraction isn't as easy as it looks: $\log(s_j(\delta_t^{(k-1)}, \theta)$ requires computing the numerical integral each time (either via quadrature or monte carlo).

BLP Pseudocode

From the outside, in:

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

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

- Inner Loop:
  - Fix θ.
  - Solve for δ so that $s_{jt}(\delta, \theta) = \tilde{s}_{jt}$.
    - Computing $s_{jt}(\delta, \theta)$ requires numerical integration (quadrature or monte carlo).
  - We can do IV-GMM to recover $\hat{\alpha}(\theta), \hat{\beta}(\theta), \hat{\xi}(\theta)$.

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

  - Use $\hat{\xi}(\theta)$ to construct moment conditions.
- When we have found $\hat{\theta}_{BLP}$ we can use this to update $W \to W(\hat{\theta}_{BLP})$ and do 2-stage GMM.

- Now that you have done change of variables to get:

$$\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$$

- We can do IV-GMM to recover $\hat{\alpha}(\theta), \hat{\beta}(\theta), \hat{\xi}(\theta)$.
- Outer Loop update guess θ, solve for δ and repeat.

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

- When we have found $\hat{\theta}_{BLP}$ we can use this to update $W \rightarrow W(\hat{\theta}_{BLP})$ and do 2-stage GMM.

BLP Alternatives

- BLP give us both a statistical estimator and an algorithm to obtain estimates.
- Plenty of other algorithms exist
  - We could solve for δ using the contraction mapping, using `fsolve` / Newton's Method / Guess and Check (not a good idea!).
  - We could try and consider a non-nested estimator for the BLP problem instead of solving for $\delta(\theta), \xi(\theta)$ we could let $\delta, \xi, \alpha, \beta$ be free parameters.
- We could think about different statistical estimators such as K-step GMM, Continuously Updating GMM, etc.

- Advantages
  - Concentrate out all of the linear in utility parameters (ξ, δ, β) so that we only search over Σ. When $\dim(\Sigma) = K$ is small (few dimensions of unobserved heterogeneity) this is a big advantage. For $K \leq 3$ this is my preferred approach.
  - When T (number of markets/periods) is large then you can exploit solving in parallel for δ market by market.
- Disadvantages
  - Small numerical errors in contraction can be amplified in the outer loop, $\rightarrow$ tolerance needs to be very tight.
  - Errors in numerical integration can also be amplified in the outer loop $\rightarrow$ must use a large number of draws/nodes.
  - Hardest part is working out the Jacobian via IFT.

$$D\delta_{\cdot t} = \begin{pmatrix} \frac{\partial \delta_{1t}}{\partial \theta_{21}} & \cdots & \frac{\partial \delta_{1t}}{\partial \theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \delta_{Jt}}{\partial \theta_{21}} & \cdots & \frac{\partial \delta_{Jt}}{\partial \theta_{2L}} \end{pmatrix} = -\begin{pmatrix} \frac{\partial s_{1t}}{\partial \delta_{1t}} & \cdots & \frac{\partial s_{1t}}{\partial \delta_{Jt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_{Jt}}{\partial \delta_{1t}} & \cdots & \frac{\partial s_{Jt}}{\partial \delta_{Jt}} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial s_{1t}}{\partial \theta_{21}} & \cdots & \frac{\partial s_{1t}}{\partial \theta_{2L}} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_{Jt}}{\partial \theta_{21}} & \cdots & \frac{\partial s_{Jt}}{\partial \theta_{2L}} \end{pmatrix},$$

BLP Extensions: Demographics

- It is helpful to allow for interactions with consumer demographics (such as income).

$$\alpha_{it} = \overline{\alpha} + \sigma_p \nu_i + \pi_p y_{it}$$

- A few ways to do this:
  - You could just use cross sectional variation in s_{jt} and $\overline{y}_t$ (mean or median income).
  - Better: Draw y_{it} from a geographic specific income distribution. Draw ν_i from a general distribution of unobserved heterogeneity.
- Ex: Nevo (2000) Cereal demand sampled individual level D_i from geographic specific CPS data
- Joint distribution of income, income-squared, age, child at home.

$$\beta_i = \overline{\beta} + \Pi D_i + \sigma \nu_i$$

- with enough observations on the same product it is possible to include fixed effects

$$\delta_{jt}(\Sigma) = x_{jt}\beta - \alpha p_{jt} + \underbrace{\xi_{jt}}_{\xi_j + \xi_t + \Delta\xi_{jt}}$$

- What does ξ_j mean in this context?
- What would ξ_t mean in this context?
- $\Delta\xi_{jt}$ is now the structural error term, this changes our identification strategy a little.

Extensions: Micro Data (Petrin 2002), (microBLP 2004)

Suppose we had additional data on behavior of individuals (in addition to aggregate market).

- Examples:
  - For some customers have answer to "Which car would you have purchased if the car you bought was not available?"
  - Demographic data on purchasers of a single brand.
  - Full individual demographic and choice data.

Extensions: Micro Data (Petrin 2002), (microBLP 2004)

- Previously we had moment conditions from orthogonality of structural error (ξ) and (X,Z) in order to form our GMM objective.

$$E[\xi_{jt}|x_{jt},z_{jt}] = 0 \rightarrow E[\xi'[Z X]] = 0$$

- We can incorporate additional information using "micro-moments" or additional moment conditions to match the micro data.
    - $Pr($ i buys j $|y_i \in [0, \$20K]) = c_1$
    - $Cov(d_i, s_{ijt}) = c_2$
    - Construct an additional error term ζ_1, ζ_2 and interact that with instruments to form additional moment conditions.
    - Econometrics get tricky when we have a different number of observations for $E[\zeta'[XZ]] = 0$ and $E[\xi'[XZ]] = 0$.

Adding Supply

Supply

- Economic theory gives us some additional powerful restrictions.
- We may want to impose $MR = MC$.
- Alternatively, we can ask – what is a good instrument for demand? something from another equation (ie: supply).

Bertrand Nash Pricing

Consider the problem of firm f which sets prices p_j for products in the set $\mathcal{J}_f$:

$$\begin{aligned} \arg\max_{p \in \mathcal{J}_f} \pi_f(\mathbf{p}) &= \sum_{j \in \mathcal{J}_f} (p_j - c_j) \cdot q_j(\mathbf{p}) \\ \rightarrow 0 &= q_j(\mathbf{p}) + \sum_{k \in \mathcal{J}_f} (p_k - c_k) \frac{\partial q_k}{\partial p_j}(\mathbf{p}) \end{aligned}$$

We can re-write the FOC in matrix form:

$$q(\mathbf{p}) = \Omega(\mathbf{p}) \cdot (\mathbf{p} - \mathbf{mc})$$

It is helpful to define two $J_t \times J_t$ matrices. The first is an *ownership matrix*:

$$O_{(j,k)} = \left\{ \begin{array}{ll} 1 & \text{for } (j,k) \in \mathcal{J}_f \\ 0 & \text{for } (j,k) \notin \mathcal{J}_f \end{array} \right\}$$

And the second is the matrix of demand derivatives $\tilde{\Omega}(\mathbf{p})$ with entries $\tilde{\Omega}_{(j,k)}(\mathbf{p}) = \frac{\partial q_j}{\partial p_k}(\mathbf{p})$. We are mainly interested in the Hadamard (element-wise) product of the two matrices $\Omega = O \odot \tilde{\Omega}$.

$$\Omega_{(j,k)}(\mathbf{p},\theta) = \left\{ \begin{array}{ll} -\frac{\partial q_j}{\partial p_k}(\mathbf{p},\theta) & \text{for } (j,k) \in \mathcal{J}_f \\ 0 & \text{for } (j,k) \notin \mathcal{J}_f \end{array} \right\}$$

Recover implied markups/ marginal costs, and assume a functional form for $mc_{jt}(x_{jt}, w_{jt})$.

$$\begin{aligned}\widehat{\mathbf{mc}}(\theta) &= \mathbf{p} - \Omega(\mathbf{p},\theta)^{-1} q(\mathbf{p},\theta)\\ f(mc_{jt}) &= x_{jt}\gamma_1 + w_{jt}\gamma_2 + \omega_{jt}\end{aligned}$$

Which we can solve for ω_{jt}:

$$\omega_{jt} = f(\mathbf{p} - \Omega(\mathbf{p},\theta)^{-1} q(\mathbf{p},\theta)) - x_{jt}\gamma_1 - w_{jt}\gamma_2$$

- $f(\cdot)$ is usually $\log(\cdot)$ or identity.
- I can use this to form additional moments: $E[\omega_{jt}' Z_{jt}^s] = 0$.
- I can just stack these up with the demand moments $E[\xi_{jt}' Z_{jt}^d] = 0$.
- Now I have $\dim(Z^d) + \dim(Z^s)$ moments altogether.
- This step is optional but can aid in identification (if you believe it).

Supply Side as Instruments

- Instruments for demand depend on exclusion restrictions
- Where do instruments come from? Something omitted that appears in another equation.

$$p_{jt} = c_{jt}(w_{jt}, x_{jt}) + \frac{s_{jt}(\mathbf{p_t})}{\left|\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}\right|} + \sum_{k \in \mathcal{J}_f} (p_k - c_k) \frac{\frac{\partial s_{kt}}{\partial p_{jt}}(\mathbf{p_t})}{\left|\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}\right|}$$

1. Exogenous regressors x_{jt}.
2. Cost shifters: w_{jt} (hard to find in practice), Hausman instruments.
3. Markup shifters: $\frac{s_{jt}(\mathbf{p_t})}{\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}}$

   (function of $(p_j, x_j, \xi_j, p_{-j}, x_{-j}, \xi_{-j})$).

Instruments and Identification

- Recall the nested logit, where there are two separate endogeneity problems
  - Price: this is the familiar one!
  - Nonlinear characteristics σ this is the other one.
- We are doing nonlinear GMM: Start with $E[\xi_{jt}|x_{jt}, z_{jt}] = 0$ use $E[\xi'[ZX]] = 0$.
  - In practice this means that for valid instruments (x, z) any function $f(x, z)$ is also a valid instrument $E[\xi_{jt} f(x_{jt}, z_{jt})] = 0$.
  - We can use $x, x^2, x^3, \ldots$ or interactions $x \cdot z, x^2 \cdot z^2, \ldots$.
  - What is a reasonable choice of $f(\cdot)$?
  - Where does z come from?

Identifcation

- Once we have $\delta_{jt}(\theta)$ identification of linear parameters is pretty straightforward

$$\delta_{jt}(\theta) = x_{jt}\beta - \alpha p_{jt} + \xi_j + \xi_t + \Delta\xi_{jt}$$

- This is either basic linear IV or panel linear IV.
- How are σ taste parameters identified?
  - Consider increasing the price of j and measuring substitution to other products k, k' etc.
  - If sales of k increase with p_j and $(x_j^{(1)}, x_k^{(1)})$ are similar then we increase the σ that corresponds to $x^{(1)}$.
  - Price is the most obvious to vary, but sometimes this works for other characteristics (like distance).
  - Alternative: vary the set of products available to consumers by adding or removing an option.

Extensions: Supply Moments

- We can also impose the Bertrand FOC as a set of additional moments.
- First parametrize marginal cost

$$\ln mc_{jt} = \gamma_1 x_{jt} + \gamma_2 w_{jt} + \omega_{jt}$$

- helpful to constrain MC to be positive always.
- Note that for any vector of prices p and demand parameters θ we can recover a unique vector of marginal costs (by solving the system of linear equations).
- Imposing the supply side only helps if we have information about the marginal costs / production function that we would like to impose
- Imposing these restrictions is helpful in constraining markups (so that implied MC are always positive, etc.).
- Misspecified functional forms for costs can cause problems!

- Common choices are average characteristics of other products in the same market $h(x_{-j,t})$. BLP instruments
  - Same firm $z_{1jt} = \overline{x}_{-j_f,t} = \frac{1}{|F_j|}\sum_{k\in\mathcal{F}_j} x_{kt} - \frac{1}{|F_j|}x_{jt}$.
  - Other firms $z_{2jt} = \overline{x}_{\cdot t} - \overline{x}_{-j_f,t} - \frac{1}{J}x_{jt}$.
  - Plus regressors $(1, x_{jt})$.
  - Plus higher order interactions
- Technically linearly independent for large (finite) J, but becoming highly correlated.
  - Can still exploit variation in number of products per market or number of products per firm.
- Correlated moments $\rightarrow$ "many instruments".
  - May be inclined to "fix" correlation in instrument matrix directly.

Demo: Logit/Nested Logit / Nevo/BLP