Supply and Instruments

Chris Conlon

September 19, 2020

Grad IO

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).

Some setup

We can break up the parameter space into three parts:

- θ_1: linear exogenous demand parameters,
- θ_2: parameters including price and random coefficients (endogenous / nonlinear)
  - $\theta_2 = [\alpha, \widetilde{\theta}_2]$
- θ_3: linear exogenous supply parameters.

Supply Side

Consider the multi-product Bertrand FOCs:

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

It is helpful to define the ownership matrix $\Omega_{(j,k)}(\mathbf{p}) = -\frac{\partial s_j}{\partial p_k}(\mathbf{p})$:

$$A(\kappa)_{(j,k)} = \left\{ \begin{array}{ll} 1 & \text{for } (j,k) \in \mathcal{J}_f \text{ for any } f \\ 0 & \text{o.w} \end{array} \right\}$$

We can re-write the FOC in matrix form where $\odot$ denotes Hadamard product (element-wise):

$$\begin{aligned}
s(\mathbf{p}) &= (A \odot \Omega(\mathbf{p})) \cdot (\mathbf{p} - \mathbf{mc}), \\
\mathbf{mc} &= \mathbf{p} - \underbrace{(A \odot \Omega(\mathbf{p}))^{-1} s(\mathbf{p})}_{\eta(\mathbf{p},\mathbf{s},\theta_2)}.
\end{aligned}$$

Recovering Marginal Costs

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

$$\begin{aligned}
\widehat{\mathbf{mc}}(\theta_2) &= \mathbf{p} - \Omega(\mathbf{p}, \theta_2)^{-1} q(\mathbf{p}, \theta_2) \\
f(mc_{jt}) &= h_s(x_{jt}, w_{jt}, \theta_3) + \omega_{jt}
\end{aligned}$$

Which we can solve for ω_{jt}:

$$\omega_{jt} = f(\mathbf{p} - \Omega(\mathbf{p}, \theta_2)^{-1} q(\mathbf{p}, \theta_2)) - h_s(x_{jt}, w_{jt}, \theta_3)$$

- $f(\cdot)$ is usually $\log(\cdot)$ or identity.
- $h_s(x_{jt}, w_{jt}, \theta_3) = [x_{jt},\, w_{jt}]\gamma$ is usually linear
- I can use this to form additional moments: $E[\omega'_{jt} Z^s_{jt}] = 0$.

Simultaneous Supply and Demand

(a) For each market t: solve $\mathcal{S}_{jt} = s_{jt}(\delta_{\cdot t}, \theta_2)$ 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 in Appendix

(e) Construct the residuals:

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

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

Additional Details

Some different definitions:

$$
\begin{aligned}
y_{jt}^D &:= \widehat{\delta}_{jt}(\theta_2) + \alpha p_{jt} &= (x_{jt}\ v_{jt})'\beta + \xi_t =: x_{jt}^{D\prime}\beta + \xi_{jt} \\
y_{jt}^S &:= \widehat{mc}_{jt}(\theta_2) &= (x_{jt}\ w_{jt})'\gamma + \omega_t =: x_{jt}^{S\prime}\gamma + \omega_{jt}
\end{aligned}
\tag{1}
$$

Stacking the system across observations yields:[1]

$$
\underbrace{\begin{bmatrix} y_D \\ y_S \end{bmatrix}}_{2N\times 1} = \underbrace{\begin{bmatrix} X_D & 0 \\ 0 & X_S \end{bmatrix}}_{2N\times(K_1+K_3)} \underbrace{\begin{bmatrix} \beta \\ \gamma \end{bmatrix}}_{(K_1+K_3)\times 1} + \underbrace{\begin{bmatrix} \xi \\ \omega \end{bmatrix}}_{2N\times 1}
\tag{2}
$$

Instruments and Identification

Parametric 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.

Instruments

- 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?

Exclusion Restrictions

$$\begin{aligned}\delta_{jt}(\mathcal{S}_t, \widetilde{\theta}_2) &= [x_{jt}, v_{jt}]\beta - \alpha p_{jt} + \xi_{jt} \\ f(p_{jt} - \eta_{jt}(\theta_2, \mathbf{p}, \mathbf{s})) &= h(x_{jt}, w_{jt}; \theta_3) + \omega_{jt}\end{aligned}$$

The first place to look for exclusion restrictions/instruments:

- Something in another equation!
- v_j shifts demand but not supply
- w_j shifts supply but not demand
- If it doesn't shift either is it really relevant?

Markup Shifters

The equilibrium markup is a function of everything! $\eta_{jt}(\mathbf{p}, \mathbf{s}, \xi_t, \omega_t, x_t, w_t, v_t, \theta_2)$:

- It is literally endogenous (depends on error terms)!
- But lots of potential instruments beyond excluded v_t or w_t.
- Also v_{-j} and w_{-j} and x_{-j}.
- Not p_{-j} or ξ_{-j}, etc.
- The idea is that these instruments shift the marginal revenue curve.
- What is a good choice of $f(x_{-j})$? etc.

BLP Instruments

- Common choices are average characteristics of other products in the same market $f(x_{-j,t})$. 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.

Armstrong (2016): Weak Instruments?

Consider the limit as $J \to \infty$

$$\frac{s_{jt}(\mathbf{p_t})}{\left|\frac{\partial s_{jt}(\mathbf{p_t})}{\partial p_{jt}}\right|} = \frac{1}{\alpha}\frac{1}{1-s_{jt}} \to \frac{1}{\alpha}$$

- Hard to use markup shifting instruments to instrument for a constant.
- How close to the constant do we get in practice?
- Average of x_{-j} seems like an especially poor choice. Why?
- Shows there may still be some power in: products per market, products per firm.
- Convergence to constant extends to mixed logits (see Gabaix and Laibson 2004).
- Suggests that you really need cost shifters.

Differentiation Instruments: Gandhi Houde (2017)

- Also need instruments for the Σ or σ random coefficient parameters.
- Instead of average of other characteristics $f(x) = \frac{1}{J-1}\sum_{k \neq j} x_k$, can transform as distance to x_j.

$$d_{jt}^{k} = x_k - x_j$$

- And use this transformed to construct two kinds of IV (Squared distance, and count of local competitors)

$$\begin{aligned} DIV_1 = &\quad \textstyle\sum_{j \in F} d_{jt}^2, & \sum_{j \notin F} d_{jt}^2 \\ DIV_2 = &\quad \textstyle\sum_{j \in F} I[d_{jt} < c] & \sum_{j \notin F} I[d_{jt} < c] \end{aligned}$$

- They choose c to correspond to one standard deviation of x across markets.

Optimal Instruments

- Since any $f(x, z)$ satisfies our orthogonality condition, we can try to choose $f(x, z)$ as a basis to approximate optimal instruments.
- This is challenging in practice – and in fact suffers from a curse of dimensionality.
- This is frequently given as a rationale behind higher order x's.
- When the dimension of x is low – this may still be feasible. $(K \leq 3)$.

Optimal Instruments

Chamberlain (1987) tells us the optimal instruments for this supply-demand system of $G\Omega^{-1}$ where for a given observation n,

$$G_n := \underbrace{\begin{bmatrix} \frac{\partial\xi}{\partial\beta} & \frac{\partial\omega}{\partial\beta} \\ \frac{\partial\xi}{\partial\alpha} & \frac{\partial\omega}{\partial\alpha} \\ \frac{\partial\xi}{\partial\sigma} & \frac{\partial\omega}{\partial\sigma} \\ \frac{\partial\xi}{\partial\gamma} & \frac{\partial\omega}{\partial\gamma} \end{bmatrix}_n}_{(K_1+K_2+K_3)\times 2} = \begin{bmatrix} -x & 0 \\ \xi_\alpha & \omega_\alpha \\ \xi_\sigma & \omega_\sigma \\ 0 & -x \\ 0 & -w \end{bmatrix}_n \qquad \Omega := \underbrace{\begin{bmatrix} v_\xi^2 & v_{\xi\omega} \\ v_{\xi\omega} & v_\omega^2 \end{bmatrix}}_{2\times 2}$$

#4: Optimal Instruments

$$G_n \Omega^{-1} = \frac{1}{v_\xi^2 v_\omega^2 - (v_{\xi\omega})^2} \times \begin{bmatrix} -v_\omega^2 x & v_{\xi\omega} x \\ v_\omega^2 \xi_\alpha - v_{\xi\omega}\omega_\alpha & v_\xi^2 \omega_\alpha - v_{\xi\omega}\xi_\alpha \\ v_\omega^2 \xi_\sigma - v_{\xi\omega}\omega_\sigma & v_\xi^2 \omega_\sigma - v_{\xi\omega}\xi_\sigma \\ v_{\xi\omega} x & -v_\xi^2 x \\ v_{\xi\omega} w & -v_\xi^2 w \end{bmatrix}_n$$

Clearly rows 1 and 4 are co-linear.

#4: Optimal Instruments

$$(G_n\Omega^{-1}) \circ \Theta = \frac{1}{v_\xi^2 v_\omega^2 - (v_{\xi\omega})^2} \times \begin{bmatrix} -v_\omega^2 x & 0 \\ v_\omega^2 \xi_\alpha - v_{\xi\omega}\omega_\alpha & v_\xi^2 \omega_\alpha - v_{\xi\omega}\xi_\alpha \\ v_\omega^2 \xi_\sigma - v_{\xi\omega}\omega_\sigma & v_\xi^2 \omega_\sigma - v_{\xi\omega}\xi_\sigma \\ 0 & -v_\xi^2 x \\ v_{\xi\omega} w & -v_\xi^2 w \end{bmatrix}_n$$

Now we can partition our instrument set by column into "demand" and "supply" instruments as

$$z_{nD} := (G_n\Omega^{-1} \circ \Theta)_{\cdot 1}$$
$$z_{nS} := (G_n\Omega^{-1} \circ \Theta)_{\cdot 2}$$

Aside: What does Supply tell us about Demand?

$$
\begin{aligned}
\partial\alpha : v_\omega^2 \xi_\alpha - v_{\xi\omega}\omega_\alpha \quad v_\xi^2 \omega_\alpha - v_{\xi\omega}\xi_\alpha \\
\partial\sigma : v_\omega^2 \xi_\sigma - v_{\xi\omega}\omega_\sigma \quad v_\xi^2 \omega_\sigma - v_{\xi\omega}\xi_\sigma
\end{aligned}
$$

- Under optimal IV these are overidentifying restrictions
- Maybe cases where one part of these instruments is trivial.

Optimal Instruments

How to construct optimal instruments in form of Chamberlain (1987)

$$E\left[\frac{\partial \xi_{jt}}{\partial \theta}|X_t, w_{jt}\right] = \left[\beta, E\left[\frac{\partial \xi_{jt}}{\partial \alpha}|X_t, w_{jt}\right], E\left[\frac{\partial \xi_{jt}}{\partial \sigma}|X_t, w_{jt}\right]\right]$$

Some challenges:

1. p_{jt} depends on X_t, w_t, ξ_t in a highly nonlinear way (no explicit solution!).
2. $E[\frac{\partial \xi_{jt}}{\partial \sigma}|X_t, w_t] = E[[\frac{\partial \mathbf{s_t}}{\partial \delta_\mathbf{t}}]^{-1}[\frac{\partial \mathbf{s_t}}{\partial \sigma}]|X_t, w_t]$ (not conditioned on endogenous p!)

"Feasible" Recipe:

1. Fix $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\sigma})$ and draw ξ_t from empirical density
2. Solve fixed point equation for $\hat{p_{jt}}$
3. Compute necessary Jacobian
4. Average over all values of ξ_t. (Lazy approach: use only $\xi = 0$).

Simplified Version: Reynaert Verboven (2014)

- Optimal instruments are easier to work out if $p = mc$.

$$c = p + \underbrace{\Delta^{-1} s}_{\to 0} = X\gamma_1 + W\gamma_2 + \omega$$

- Linear cost function means linear reduced-form price function.

$$\begin{aligned}
E\left[\frac{\partial \xi_{jt}}{\partial \alpha} | z_t\right] &= E[p_{jt}|z_t] = x_{jt}\gamma_1 + w_{jt}\gamma_2 \\
E\left[\frac{\partial \omega_{jt}}{\partial \alpha} | z_t\right] &= 0, \quad E\left[\frac{\partial \omega_{jt}}{\partial \sigma} | z_t\right] = 0 \\
E\left[\frac{\partial \xi_{jt}}{\partial \sigma} | z_t\right] &= E\left[\frac{\partial \delta_{jt}}{\partial \sigma} | z_t\right]
\end{aligned}$$

- If we are worried about endogenous oligopoly markups is this a reasonable idea?
- Turns out that the important piece tends to be shape of jacobian for σ_x.

Optimal Instruments: Reynaert Verboven (2014)

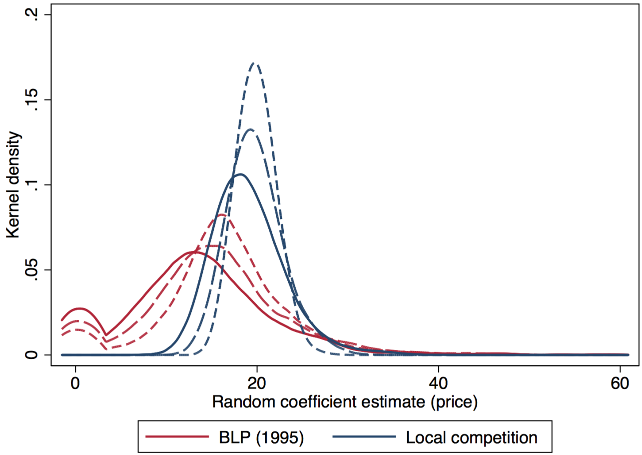
Table 2: Bias and Efficiency with Imperfect Competition

		Single Equation GMM								
			g_{jt}^1			g_{jt}^2			g_{jt}^3	
	True	Bias	St Err	RMSE	Bias	St Err	RMSE	Bias	St Err	RMSE
β^0	2	-0.127	0.899	0.907	-0.155	0.799	0.814	-0.070	0.514	0.519
β^1	2	-0.068	0.899	0.901	0.089	0.766	0.770	-0.001	0.398	0.398
α	-2	0.006	0.052	0.052	0.010	0.049	0.050	0.010	0.043	0.044
σ^1	1	-0.162	0.634	0.654	-0.147	0.537	0.556	-0.016	0.229	0.229
		Joint Equation GMM								
			g_{jt}^1			g_{jt}^2			g_{jt}^3	
	True	Bias	St Err	RMSE	Bias	St Err	RMSE	Bias	St Err	RMSE
β^0	2	-0.095	0.714	0.720	-0.103	0.677	0.685	0.005	0.459	0.459
β^1	2	0.089	0.669	0.675	0.098	0.621	0.628	-0.009	0.312	0.312
α	-2	0.001	0.047	0.047	0.002	0.046	0.046	-0.001	0.043	0.043
σ^1	1	-0.116	0.462	0.476	-0.110	0.418	0.432	0.003	0.133	0.133

Bias, standard errors (St Err) and root mean squared errors (RMSE) are computed from 1000 Monte Carlo replications. Estimates are based on the MPEC algorithm and Sparse Grid integration. The instruments g_{jt}^1, g_{jt}^2.and g_{jt}^3 are defined in section 2.4 and 2.5.

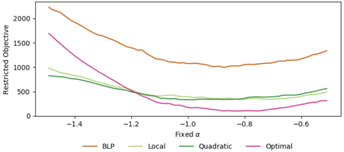
Differentiation Instruments: Gandhi Houde (2016)

Figure 4: Distribution of parameter estimates in small and large samples

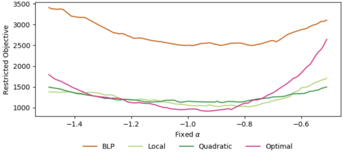


IV Comparison: Conlon and Gortmaker (2019)

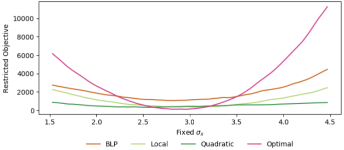
Figure 2: Profiled GMM Objective with Alternative IV ("Simple" simulation)



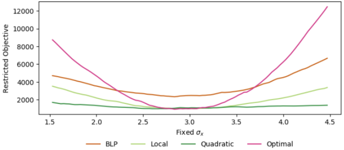
(a) α: Demand-Only



(b) α: Supply and Demand



(c) σ_x: Demand-Only



(d) σ_x: Supply and Demand

Each plot profiles the GMM objective $Q(\theta)$ with respect to a single parameter for our "Simple" simulation scenario and a single simulation. We fix either σ_x or α and re-optimize over other parameters and plot the restricted objective in each subplot. The top row profiles the objective over the price parameter α, while the bottom row profiles over the random coefficient σ_x. The left column uses moments from demand alone, while the right column uses both supply and demand moments.

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.

Dube Fox Su (2012)

$$
\begin{aligned}
\arg\min_{\theta_2} & \quad \psi'\Omega^{-1}\psi \quad \text{s.t.} \\
\psi &= \xi(\theta_2)'Z \\
\xi_{jt}(\theta) &= \delta_{jt}(\theta_2) - x_{jt}\beta - \alpha p_{jt} \\
\log(S_{jt}) &= \log(s_{jt}(\delta, \theta_2))
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
\arg\min_{\theta_2,\alpha,\beta,\xi,\psi} & \quad \psi'\Omega^{-1}\psi \quad \text{s.t.} \\
\psi &= \xi'Z \\
\xi_{jt} &= \delta_{jt} - x_{jt}\beta - \alpha p_{jt} \\
\log(S_{jt}) &= \log(s_{jt}(\theta_2, \delta))
\end{aligned} \tag{4}
$$

Comparing Approaches

- The original BLP paper and the DFS paper define different algorithms to produce the same statistical estimator.
    - The BLP algorithm is a nested fixed point (NFP) algorithm.
    - The DFS algorithm is a mathematical program with equilibrium constraints (MPEC).
    - The unknown parameters satisfy the same set of first-order conditions. (Not only asymptotically, but in finite sample).
    - $\hat{\theta}_{NFP} \approx \hat{\theta}_{MPEC}$ but for numerical differences in the optimization routine.
- Our choice of algorithm should mostly be about computational convenience.

BLP: NFP Advantages/Disadvantages

- 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.

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

BLP: MPEC Advantages/Disadvantages

- Advantages
    - Problem scales better in $\dim(\Sigma)$.
    - Because all constraints hold at the optimum only: less impact of numerical error in tolerance or integration.
    - Derivatives are less complicated than $\frac{\partial \delta}{\partial \theta}$ (no IFT).
- Disadvantages
    - We are no longer concentrating out parameters, so there are a lot more of them! Storing the (Hessian) matrix of second derivatives can be difficult on memory.
    - We have to find the derivatives of the shares with respect to all of the parameters β, ξ, θ. (The other derivatives are pretty easy).
    - Parallelizing the derivatives is trickier than NFP case.