# CompeteGrok Analysis Report

**Query:** explain BLP demand estimation step by step

**Timestamp:** 2025-12-17 15:37:33.728195

**Routes:** []

## Agent Output

### BLP Demand Estimation: Step-by-Step Explanation

The Berry-Levinsohn-Pakes (BLP) model (from Berry, Levinsohn, and Pakes, *Econometrica* 1995) is a cornerstone of empirical industrial organization (IO) economics. It estimates **demand for differentiated products** in markets with thousands of consumers, accounting for:

- **Price endogeneity** (prices correlate with unobserved quality shocks $\xi$).
- **Heterogeneity** in consumer preferences (via random coefficients).
- **Market equilibrium** (supply side often estimated jointly, but focus here is demand).

It extends the **Berry (1994) inversion** from the homogeneous logit and uses **nonlinear Generalized Method of Moments (GMM)** for estimation. The core challenge: demand is nonlinear in parameters, so we **invert** observed market shares to recover mean utilities $\delta$, then use GMM to fit the model.

I'll explain **demand estimation only** (supply can be added post-demand). Assume: - **Markets** $t = 1$ to $T$ (e.g., city-quarters). - **Products** $j = 1$ to $J_t$ in each market. - Data: prices $p_{jt}$, characteristics $x_{jt}$ (e.g., size, HP), **observed market shares** $s_{jt}$ (e.g., from aggregate sales / total potential sales). - **Outside good** share $s_{0t} = 1 - \sum_j s_{jt}$.

**Step 1: Specify the Random Coefficients Logit Utility** Consumer *i*'s utility for product *j* in market *t*:

```
u_{ijt} =  _{jt} +  _{ijt}(p_{jt}, x_{jt},  _i;  _Σ) +  _{ijt}
```

- **Mean utility** $\delta_{jt} = x_{jt} \beta_x + \alpha p_{jt} + \xi_{jt}$
  - $x_{jt}$: Observed product chars (e.g., size, dummy for air conditioning).
  - $p_{jt}$: Price.
  - $\beta_x$, $\alpha$: Mean tastes ($\alpha < 0$ for downward-sloping demand).
  - $\xi_{jt}$: Unobserved (mean-zero) product quality shock → **endogenous**.
- **Heterogeneity** $\mu_{ijt} = \sum_k \sigma_k v_{ik} x_{jkt}$ (often includes price: $\sigma_p v_{ip} p_{jt}$).
  - $v_i \sim G(v)$ (e.g., i.i.d. standard normal; simulated with H=100-500 draws).
  - $\theta_\Sigma = \{\sigma_k\}$: Nonlinear **random coefficient** parameters → allow flexible substitution patterns.
- $\varepsilon_{ijt}$ ~ Type I extreme value (Gumbel) → induces logit form conditionally on $v_i$.

**Structural parameters** $\theta = (\theta_1, \theta_2) = (\beta_x, \alpha; \theta_\Sigma)$.

**Step 2: Derive Predicted Market Shares** Integrate over consumer heterogeneity:

```
ŝ_{jt}( _t, p_t, x_t;  _2) =   [exp( _{jt} +  _{ijt}) / (1 +  _k exp( _{kt} +  _{ikt})) ] dG( _i)
```

- Denominator = 1 (outside good) + ∑ inside goods.
- **Simulate** the integral: Draw H vectors $v^{(h)}$, average over h=1 to H.
- Observed $s_{jt}$ **must equal** $\hat{s}_{jt}$ for the true $\delta_t$ and $\theta_2$ → **implicit equation**.

**Step 3: Invert Shares to Recover Mean Utilities $\delta_t(\theta_2)$ (Contraction Mapping)** For fixed $\theta_2$, solve **nonlinear system** for $\delta_t$ ($J_t$ equations per market t):

```
s_{jt} = ŝ_{jt}( _t, p_t, x_t;  _2)     j,t
```

Equivalent to (taking log-odds vs. outside good):

```
log(s_{jt} / s_{0t}) = log( ŝ_{jt}( _t) / (1 -  _k ŝ_{kt}( _t)) )
```

- **BLP contraction mapping** (guaranteed to converge; Picard fixed-point theorem):
  1. Initialize: $\delta^{(0)}_{jt} = \log(s_{jt} / s_{0t})$.
  2. Iterate r=1,2,… until $||\delta^{(r)} - \delta^{(r-1)}|| < \varepsilon$ (e.g., $10^{-12}$; ~20 iterations):

     ```
     ^{(r)}_{jt} = log(s_{jt} / s_{0t}) - log( E_  [ exp( ^{(r-1)}_{jt} +  _{jt}( )) / D_t(
     ```

     - $D_t(\delta, v) = 1 + \sum_k \exp(\delta_{kt} + \mu_{kt}(v))$ (choice-specific denom).
     - $E_v[\cdot] = (1/H) \sum_h [\cdot]$ over simulation draws.
  3. Output: $\delta_t(\theta_2)$ → "unobserved" mean utilities backed out from data.

This is **fast** per candidate $\theta_2$ (vectorized in code like MATLAB/Python).

**Step 4: Form Residuals and Moment Conditions**

```
 _{jt}( ) =  _{jt}( _2) - x_{jt}  _x -   p_{jt}
```

- Orthogonality: $E[\xi_{jt} | Z_{jt}] = 0$ (instruments $Z_{jt}$ uncorrelated with ξ).
- **Instruments Z** (critical for price endogeneity): | Type | Examples | |——|———-| | **Product chars** | Own $x_{jt}$ (for non-price). | | **Cost shifters** | Materials/labor costs (if available). | | **Hausman-style** | $\sum_{k \neq j} x_{kt}$ (rival chars; orthogonal under monopoly pricing). | | **BLP** | ∑*{k□ same nest/group}* x{kt} (e.g., by firm/size). |
- Stack: Let $Z_t$ be ($J_t \times L$) instruments matrix per market.
- Sample moments: $\psi(\theta) = (1/\sum J_t) \sum_t Z_t' \xi_t(\theta)$.

**Step 5: Nonlinear GMM Optimization** Minimize quadratic form:

```
Q( ) = ( )' W ( )
```

- **W**: Weighting matrix.
  - 1st stage: Identity (or diagonal).
  - 2nd stage: Optimal $W = [\sum Z_t' \xi \xi' Z_t]^{-1}$ (iterate).
- **Nested optimization** ($\theta_2$ high-dimensional, e.g., 5-10 sigmas):
  1. **Outer loop**: Grid/search/optimize $\theta_2$ (e.g., Nelder-Mead, or Bayesian).
     - For each $\theta_2$ candidate:
       * Run contraction (Step 3) $\rightarrow \delta_t(\theta_2)$.
       * **Inner linear GMM**: Solve $\theta_1(\theta_2) = \text{argmin}\ \psi(\theta_1,\theta_2)' W \psi = (Z'X W Z'X)^{-1}$ etc. (2SLS projection: $\theta_1 = (X' Z W Z' X)^{-1} X' Z W Z' \delta$).
       * Compute $Q(\theta_1(\theta_2), \theta_2)$.
  2. Converged $\theta \rightarrow$ estimates.
- Software: Nevo's MATLAB code, pyblp (Python), or R packages.

**Step 6: Inference and Diagnostics**

- **Standard errors**: GMM variance-covariance:

  ```
  Var( ) = (1/N) [D' W D]^{-1} (D' W Ω W D) [D' W D]^{-1}
  ```

  - $D = \partial\psi/\partial\theta$ (numerical Jacobian).
  - $\Omega = \text{var}(\sqrt{N}\ \psi) \approx (1/N) \sum Z_t \xi_t \xi_t' Z_t$.
- **Diagnostics**:
  - Overidentification: Hansen J-test (should not reject).
  - Monotonicity: Elasticities ↓ in price.
  - OWBP test (O'Brien-Wansbeek-Pagan): Check inversion.
  - Plots: $\delta$ vs. fitted; shares fit.

**Key Insights & Caveats**

- **Why it works**: Inversion linearizes heterogeneity; GMM handles endogeneity.
- **Computation**: $O(T \times J^2 \times H \times \text{outer its}) \rightarrow$ feasible with $J \sim 20$, $T \sim 50$, $H=200$.
- **Extensions**: Nested logit (group sigmas), full supply (markup eq: $p - c = \mu(s,\theta)$), dynamics.
- **Common pitfalls**: Bad instruments → bias; few sims H → noisy; collinear x.
- **Empirical use**: Autos (BLP original), cereals, airlines, pharma.

For code/examples, see pyblp docs or Nevo (2000) handbook chapter. If you need supply-side, simulations, or caveats for a specific app, ask!

### Agent Output

**Synthesis:** See above.

**Privacy:** Ephemeral RAG; zero retention.
**Disclaimer:** Not legal advice. Models have caveats (e.g. IIA assumption). Verify 2025 data.
**LaTeX:** Inline $x$, display

$$E = mc^2$$

.