# Identification and Estimation of Demand for Differentiated Products

Jean-François Houde
Cornell University & NBER

September 29, 2017

# Starting Point: The Characteristic Approach

- **Ultimate goal:** Measure the elasticity of substitution between goods.
- **Why?**
  - Measure the degree of market power
  - Predict the effect of proposed mergers
  - Evaluate the value of new goods
  - ...
- **Natural starting point:** Linear inverse demand

$$p_{jt} = \alpha_0 + \sum_{k \in \mathcal{J}} \alpha_{j,k} q_{kt} + \epsilon_{jt}$$

- **Curse of dimensionality:** Even with *exogenous* variation in $q$'s, the number of parameters to estimate grows with the number of products.
- **Solution:** The characteristic approach (Lancaster 1966)
  - Each product is defined as a bundle of characteristics: $x_j = \{x_1, \ldots, x_k\}$
  - Consumers assigned values to each characteristics, and choose the product maximize their utility given their budget.
  - Demand for product $j$ is function of the distribution of valuations for each attribute (finite dimension).

# Discrete Choice Models

- **Two classes of models**
  - *Local:* Each product has only few close substitutes
  - *Global:* Each product is equally substitutable with all others
- **Global competition:** Logit (McFadden 1974, Anderson et al. 1992)
  - Indirect utility:

$$u_{ij} = \underbrace{\delta_j - p_j}_{\text{average}} + \underbrace{\epsilon_{ij}}_{\text{T1EV}}$$

  - Demand (shares):

$$s_j = \frac{\exp((\delta_j - p_j)/\sigma_\epsilon)}{\sum_{j' \in \mathcal{J}} \exp((\delta_{j'} - p_{j'})/\sigma_\epsilon)}$$

  - Elasticity of substitution:

$$\eta_{jk} = \left(\frac{1}{\sigma_\epsilon} s_j s_k\right) \frac{p_k}{s_j} = \frac{1}{\sigma_\epsilon} p_k s_j$$

  - Two key properties: (i) Independence of Irrelevance Alternatives (IIA), and (ii) monopolistic competition in the limit (i.e. constant markup).

# Local competition: Vertical or horizontal differentiation?

- **Definition:**
  - *Horizontal differentiation:* Consumers have different valuations of products available (even with equal prices)
  - *Vertical differentiation:* Consumers would choose the **same** product if prices were equal.
- **Horizontal Example:** Linear city model (Hotelling 1929)
  - Preferences for product $j = 1, 2$:

    $$u_{ij} = \delta - p_j - \lambda |l_i - x_j|$$

    where $l_i \sim U(0, 1)$ is the *location* (or taste) of consumer $i$, and $x_j \in (0, 1)$ is the "address" of products (i.e. characteristic) and $x_2 > x_1$.
  - Demand:

    $$s_1 = \int_0^{\bar{l}} f(l_i) dl_i = \frac{\lambda(x_2 - x_1) - p_1 + p_2}{2}$$

    where $\bar{l}$ is the "marginal" consumer (i.e. indifferent type).

# Local competition: Vertical or horizontal differentiation?

- **Vertical Example:** Quality ladder (Shaked and Sutton 1982, Bresnahan 1987)
  - Preferences for product $j = 1, 2$:

    $$u_{ij} = \theta_i x_j - p_j$$

    where $\theta_i \sim U(0, \lambda)$ is the *type* of consumer $i$, and $x_j \in (0, 1)$ is the "quality" of products (i.e. characteristic) and $x_2 > x_1$.
  - Demand:

    $$\text{Type } \bar{\theta}: \qquad \bar{\theta} x_1 - p_1 = \bar{\theta} x_2 - p_2$$

    $$s_1 = \int_0^{\bar{\theta}} f(\theta_i) d\theta_i = \bar{\theta}/\lambda = \frac{1}{\lambda} \frac{p_2 - p_1}{x_2 - x_1}$$

# Hybrid model: Random-coefficients

- The *work-horse* model is the random-utility model with heterogenous valuations for characteristics (aka random-coefficients):

$$\max_{j \in \mathcal{J}} x_j \beta_i - \alpha_i p_j + \epsilon_{ij}$$

- **Note 1:** Despite the multiplicative form, this functional form nests *some* Hotelling-style models.

$$\text{Example:} \quad u_{ij} = \delta_j - p_j - \lambda(z_j - l_i)^2 = x_j \beta_i - \alpha p_j.$$

- **Note 2:** The nested-logit model is a special case of the random-coefficient model:

$$u_{ij} = x_j \beta + \sum_k 1(j \in \mathcal{G}_k) \nu_{ik} + \epsilon_{ij}$$

where $\mathcal{G}_k$ is a nest or product-segment (i.e. discrete attribute).

# From McFadden to BLP

- **Data:** Panel of market shares and product characteristics:

$$\{s_t, p_t, x_t\}_{t=1,\ldots,T}$$

  where $t$ indexes a market, $x_t = \{x_{jt,1}, \ldots, x_{jt,K}\}_{j=1,\ldots,n_t}$ is a matrix of observed characteristics, and $\{p_t, s_t\} = \{p_{jt}, s_{jt}\}_{j=1,\ldots,n_t}$ is a matrix of endogenous prices and market shares.
- I will use $w_{jt}$ to denote a vector of price instruments (e.g. cost shifters, ownership, etc.)
- Market shares:

$$s_{jt} = q_{jt}/M_t$$

  where $M_t$ is the **observed** market size (exogenous). Examples:
    - **Cars:** U.S. population
    - **Cereals:** Population $\times$ Avg. servings per month.
    - **Gasoline:** Transportation needs (including car, bus, walk, etc.)

# From McFadden to BLP

- **Assumption:** $s_{jt}$ is measured without error (very important).
  - This often mean that you should **aggregate** data across time periods or markets to reduce the importance of measurement error in $q_{jt}$.
  - The data requirement in BLP is often underestimated... You don't need consumer-level data, but you need data on **every** consumers!
- Indirect utility function (Nevo 2001):

$$u_{ijt} = \begin{cases} \sum_{k=1}^{K} \beta_{i,k} x_{jt,k} - \alpha_i p_{jt} + \xi_{jt} + \epsilon_{ij} & \text{If } j \neq 0 \\ \epsilon_{i0} & \text{Else.} \end{cases}$$

$$\beta_{i,k} = \beta_k + z_i \pi_k + \eta_{i,k}$$

$$\alpha_i = \alpha + z_i \pi_p + \eta_{i,p}$$

$$z_i \sim F(\cdot) \text{ (known) and } \nu_i \sim N(0, \Sigma) \text{ (unknown)}$$

- Average indirect utility: $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$.

# Baseline Model: Exogenous Characteristics

- Consider first a model without prices and without demographics characteristics
- **Demand:** Linear random-coefficient with T1EV random utility shocks

$$\sigma_j\left(\delta_t, x_t^{(2)}; \Sigma\right) = \int \frac{\exp\left(\delta_{jt} + \nu_i^T x_{jt}^{(2)}\right)}{1 + \sum_{j'=1}^{J_t} \exp\left(\delta_{j't} + \nu_i^T x_{j't}^{(2)}\right)} dF(\nu_i|\Sigma)$$

  where $\delta_{jt} = \beta_0 + x_{jt}^{(1)}\beta_1 + x_{jt}^{(2)}\beta_2 + \xi_{jt}$.
- The **residual** of the model is obtained from the inverse-demand function:

$$\rho_j(s_t, x_t; \theta) = \sigma_j^{-1}\left(s_t, x_t^{(2)}; \Sigma\right) - x_{jt}\beta, \quad \text{where } \theta = (\beta, \Sigma).$$

# Identifying Assumption

- **Assumption:** The unobserved attribute of each product is independent of the **menu**, $x_t$, of characteristics available in market $t$,

$$E[\xi_{jt}|x_t] = 0 \quad \text{(CMR)}.$$

- In practice, the model is estimated using a finite number ($L$) of unconditional moment restrictions, $A_j(x_t)$:

$$\begin{aligned} & E\left[\rho_j(s_t, x_t; \theta^0) \cdot A_j(x_t)\right] = 0 \\ \leftrightarrow \quad & E\left[\left(\sigma_j^{-1}\left(s_t, x_t^{(2)}; \Sigma^0\right) - x_{jt}\beta\right) \cdot A_j(x_t)\right] = 0. \end{aligned}$$

- **Gandhi & Houde (2016):** How to construct relevant instruments to identify $\Sigma$?
  - **Stock & Wright (2000)**: $A_j(x_t)$ is weak if the moment conditions are *almost* satisfied away from the true parameters.

# What in the data identifies the model?

- Like any IV problem, we need to have enough instruments to estimate $(\bar{\beta}, \Sigma)$. But what is a relevant/valid instrument?
- The CMR suggests that $\theta$ is identified by variation in the choice-set of consumers.
  - Most identification sections describe the "ideal" experiment in which a new product enters, and steal market shares from existing products with similar attributes.
- This "red-bus/blue-bus" logic is correct, but we need to use it to construct IVs, otherwise the model will be weakly identified (even if the ideal experiment is present the data-set!)

# Illustration of the Weak IV Problem

- Two detection tests:
  1. Testing the **wrong** model: IIA hypothesis

$$H_0 : E\left[\rho_j(s_t, x_t|\beta, \Sigma = 0) \cdot z_{jt}\right] = 0$$

$$\leftrightarrow \qquad \ln s_{jt}/s_{0t} = x_{jt}\beta + \gamma z_{jt} + \xi_{jt}$$
$$H_0 : \gamma = 0$$

  2. Local identification: Cragg-Donald rank test

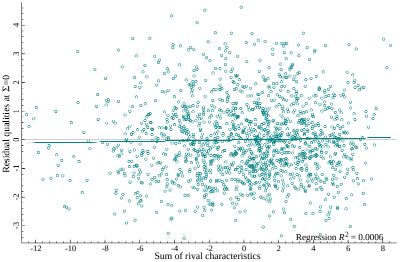
$$\text{rank}\left(E\left[\partial\rho_j\left(s_t, x_t; \theta\right)/\partial\theta^T \cdot z_{jt}\right]\right) = m$$

  This test can be implemented in STATA (ivreg2 or ranktest).

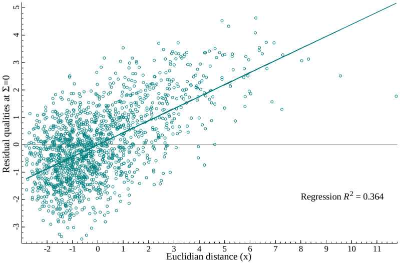
- Monte-Carlo design:
  - Sample: $T = 100$ and $J = 15$
  - Random utility with (independent) normal random-coefficients ($K_2$)
  - DGP: $(x_{jt,k}, \xi_{jt}) \sim N(0, I)$

# Weak Identification in a Picture: IIA Test

(A) IV: Sum of rivals' characteristics

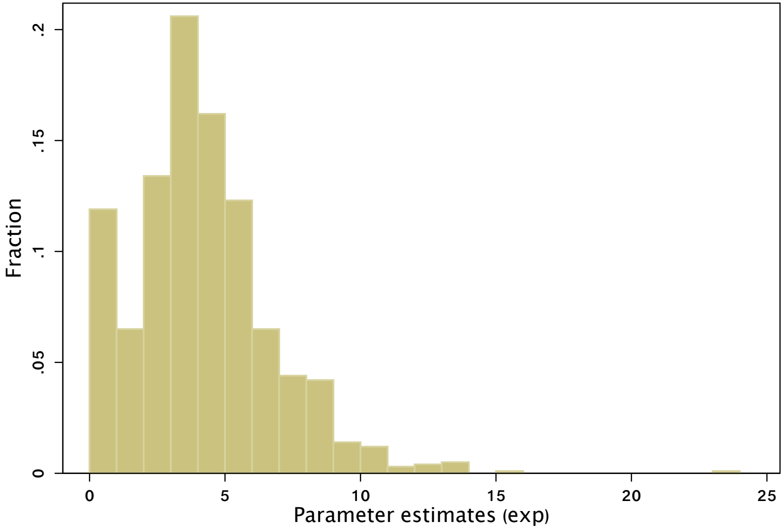



(B) IV: Euclidean distance in $x$




- **Takeaway:** Independence of $\xi_{jt}$ and the *distance of rival characteristics* rules out the IIA hypothesis, but **not** the *sum of rival characteristics*.

# Distribution of $\hat{\sigma}_2$ with weak IVs




Shapiro-Wilk test for normality: 15.71 (0). Width = 1.

# GMM Estimates with Weak IVs

| | $K_2 = 1$ | | $K_2 = 2$ | | $K_2 = 3$ | | $K_2 = 4$ | |
|---|---|---|---|---|---|---|---|---|
| | bias | rmse | bias | rmse | bias | rmse | bias | rmse |
| $\log \sigma_1$ | -11.29 | 95.93 | -5.43 | 74.95 | -1.15 | 5.50 | -8.40 | 229.67 |
| $\log \sigma_2$ | | | -4.69 | 58.31 | -1.36 | 6.26 | -1.10 | 6.17 |
| $\log \sigma_3$ | | | | | -1.41 | 9.20 | -4.66 | 112.64 |
| $\log \sigma_4$ | | | | | | | -0.93 | 4.02 |
| $\sigma_1$ | 0.14 | 2.64 | -0.01 | 2.49 | -0.03 | 2.19 | 0.22 | 2.35 |
| $\sigma_2$ | | | 0.12 | 2.42 | -0.01 | 2.27 | 0.10 | 2.30 |
| $\sigma_3$ | | | | | 0.18 | 2.38 | 0.11 | 2.38 |
| $\sigma_4$ | | | | | | | 0.08 | 2.21 |
| 1(Local-min) | 0.19 | | 0.51 | | 0.59 | | 0.66 | |
| Range(J) | 0.74 | | 1.15 | | 1.64 | | 1.51 | |
| Range(pv) | 0.17 | | 0.19 | | 0.21 | | 0.21 | |
| Range($\log \sigma$) | 11.74 | | 6.64 | | 6.58 | | 4.86 | |
| Rank-test | 1.26 | | 0.46 | | 0.26 | | 0.18 | |
| p-value | 0.62 | | 0.81 | | 0.89 | | 0.92 | |
| IIA-test | 1.33 | | 1.30 | | 1.49 | | 1.94 | |
| p-value | 0.43 | | 0.42 | | 0.36 | | 0.24 | |

# GMM Estimates with Weak IVs

| | $K_2 = 1$ | | $K_2 = 2$ | | $K_2 = 3$ | | $K_2 = 4$ | |
|---|---|---|---|---|---|---|---|---|
| | bias | rmse | bias | rmse | bias | rmse | bias | rmse |
| $\log\sigma_1$ | -11.29 | 95.93 | -5.43 | 74.95 | -1.15 | 5.50 | -8.40 | 229.67 |
| $\log\sigma_2$ | | | -4.69 | 58.31 | -1.36 | 6.26 | -1.10 | 6.17 |
| $\log\sigma_3$ | | | | | -1.41 | 9.20 | -4.66 | 112.64 |
| $\log\sigma_4$ | | | | | | | -0.93 | 4.02 |
| $\sigma_1$ | 0.14 | 2.64 | -0.01 | 2.49 | -0.03 | 2.19 | 0.22 | 2.35 |
| $\sigma_2$ | | | 0.12 | 2.42 | -0.01 | 2.27 | 0.10 | 2.30 |
| $\sigma_3$ | | | | | 0.18 | 2.38 | 0.11 | 2.38 |
| $\sigma_4$ | | | | | | | 0.08 | 2.21 |
| 1(Local-min) | 0.19 | | 0.51 | | 0.59 | | 0.66 | |
| Range(J) | 0.74 | | 1.15 | | 1.64 | | 1.51 | |
| Range(pv) | 0.17 | | 0.19 | | 0.21 | | 0.21 | |
| Range($\log\sigma$) | 11.74 | | 6.64 | | 6.58 | | 4.86 | |
| Rank-test | 1.26 | | 0.46 | | 0.26 | | 0.18 | |
| p-value | 0.62 | | 0.81 | | 0.89 | | 0.92 | |
| IIA-test | 1.33 | | 1.30 | | 1.49 | | 1.94 | |
| p-value | 0.43 | | 0.42 | | 0.36 | | 0.24 | |

# GMM Estimates with Weak IVs

| | $K_2 = 1$ | | $K_2 = 2$ | | $K_2 = 3$ | | $K_2 = 4$ | |
|---|---|---|---|---|---|---|---|---|
| | bias | rmse | bias | rmse | bias | rmse | bias | rmse |
| $\log\sigma_1$ | -11.29 | 95.93 | -5.43 | 74.95 | -1.15 | 5.50 | -8.40 | 229.67 |
| $\log\sigma_2$ | | | -4.69 | 58.31 | -1.36 | 6.26 | -1.10 | 6.17 |
| $\log\sigma_3$ | | | | | -1.41 | 9.20 | -4.66 | 112.64 |
| $\log\sigma_4$ | | | | | | | -0.93 | 4.02 |
| $\sigma_1$ | 0.14 | 2.64 | -0.01 | 2.49 | -0.03 | 2.19 | 0.22 | 2.35 |
| $\sigma_2$ | | | 0.12 | 2.42 | -0.01 | 2.27 | 0.10 | 2.30 |
| $\sigma_3$ | | | | | 0.18 | 2.38 | 0.11 | 2.38 |
| $\sigma_4$ | | | | | | | 0.08 | 2.21 |
| 1(Local-min) | 0.19 | | 0.51 | | 0.59 | | 0.66 | |
| Range(J) | 0.74 | | 1.15 | | 1.64 | | 1.51 | |
| Range(pv) | 0.17 | | 0.19 | | 0.21 | | 0.21 | |
| Range($\log\sigma$) | 11.74 | | 6.64 | | 6.58 | | 4.86 | |
| Rank-test | 1.26 | | 0.46 | | 0.26 | | 0.18 | |
| p-value | 0.62 | | 0.81 | | 0.89 | | 0.92 | |
| IIA-test | 1.33 | | 1.30 | | 1.49 | | 1.94 | |
| p-value | 0.43 | | 0.42 | | 0.36 | | 0.24 | |

# Non-Parametric Identification (Berry and Haile (2014))

- To get a sense of "how" the model is identified, it is useful to consider an **ideal** setting.
  - Infinitely many markets with exogenous changes in the choice-set facing consumers.
- **Note:** BH are interested in studying the identification of the demand system $\sigma_j(x_t)$; not the distribution $f(\nu_i|\Sigma)$.
- New normalization:

$$\sigma_j^{-1}\left(s_t, x^{(2)}\right) = x_{jt}^{(1)} + \xi_{jt}$$

  where $x_{jt}^{(1)}$ is a scalar (i.e. special regressor).
- The BLP model written in this form correspond to a non-parametric instrumental variable model (Newey and Powell (2003)):

$$x_{jt}^{(1)} = \sigma_j^{-1}\left(s_t, x^{(2)}\right) + \xi_{jt}$$

# Non-Parametric Identification (Berry and Haile (2014))

- **Question 1:** Is it possible to identify $\sigma_j^{-1}(\cdot)$ from the conditional mean restriction (i.e. $E(\xi_{jt}|x_t) = 0$)?
- To answer this we need to rely on the non-parametric analog of a **rank** condition. Do we have "enough" independent excluded instruments?
- **Reduced-form:** Applying the CMR to the inverse-demand equation,

$$E[x_{jt}^{(1)}|x_t] = E\left[\sigma_j^{-1}\left(s_t, x^{(2)}\right)|x_t\right] - E[\xi_{jt}|x_t]$$

$$x_{jt}^{(1)} = E\left[\sigma_j^{-1}\left(s_t, x^{(2)}\right)|x_t\right]$$

- **Completeness condition:** For all functions $B(s_t)$ with finite expectations, if $E[B(s_t)|x_t] = 0$ almost surely, then $B(s_t) = 0$ a.s.
- If the demand function and distribution of $x_t$ satisfy this condition, then there is a unique $\sigma_j^{-1}(\cdot)$ that can be "inverted" from the reduced-form (see Theorem 1).

# Non-Parametric Identification (Berry and Haile (2014))

- **Question 2:** Is there a unique demand function associated with $\sigma_j^{-1}(s_t, x_t)$?
- The answer is yes in most mixed-logit models (see Berry (1994))
- Fore more general demand systems, Berry, Gandhi, and Haile (2013) defined a new condition that is sufficient for existence and uniqueness of an inverse demand: *connected substitutes*
  - Require that the index $\delta_{jt}$ weakly lower the market shares of all goods $k \neq j$.
  - See paper for more details...
- If both conditions are satisfied, the model is non-parametrically identified.
- The argument is only slightly more complicated with prices: The completeness condition (i.e. full rank) extends to the price IVs.

# Identification

- Back to the parametric model...
- The model is identified by the same logic:

$$E[\rho_j(s_t, x_t; \theta)|x_t] = 0, \text{ iff } \theta = \theta^0$$

$$\Leftrightarrow \underbrace{E\left[\sigma_j^{-1}\left(s_t, x_t^{(2)}; \Sigma^0\right)|x_t\right]}_{\text{Reduced-form}} - \beta_0 - x_{jt}^{(1)}\beta_1 - x_{jt}^{(2)}\beta_2 = 0$$

- **Takeaway:** The presence of a *special regressor* $x_{jt}^{(1)}$ implies that $x_{-j,t}^{(1)}$ can be used as *excluded instruments* for the endogenous shares.

# How to select the instruments?

- Since $\dim(x_t) >> dim(\Sigma) = m$, any transformation of $x_t = \{x_{1t}, \ldots, x_{J_t,t}\}$ can be used to construct valid moments:

$$E[\rho_j(s_t, x_t; \theta) \times A_j^L(x_t)] = 0, \text{ where } dim(A_j^L(x_t)) = L \geq m$$

- **Donald, Imbens, and Newey (2003):** Conditional moment restriction is equivalent to a countable number of unconditional moment restrictions (aka IVs),

$$E[\rho_j(s_t, x_t; \theta) \times A_j^L(x_t)] = 0 \text{ Iff } E[\rho_j(s_t, x_t; \theta) | x_t] = 0,$$

  where the instruments $A_j^L(x_t)$ correspond to **basis functions** spanning the space of $x_t$ (dimension $L$).
- In our context, a necessary condition for this equivalence is that the reduced-form of the model can be approximated by $A_j^L(x_t)$ as $L \rightarrow \infty$:

$$E\left[\{g_j(x_t) - A_j^L(x_t)\gamma_L\}^2\right] \rightarrow 0$$

# Curse of Dimensionality Problem

- **Curse of Dimensionality:** The reduced-form is a *product-specific* function of the entire menu of product characteristics.
  - As $J \uparrow$, both the number of arguments **and** the number of functions to approximate increase.
  - This is a *different* point than the one raised by Armstrong (2015), which is about the weakness of the price instruments as $J \to \infty$.
- Without further restrictions, we cannot directly use the insights of BH to construct relevant IVs
- **Our approach:** Reduce the dimensionality of the problem by exploiting the symmetry of the demand function (implied by the linearity of the random utility model)

# What does the characteristic structure imply for the reduced-form of the model?

- Market-structure facing product $j$ (dropping $t$):

$$(\boldsymbol{w}_j, \boldsymbol{w}_{-j}) \equiv \left(\left(\delta_j, \boldsymbol{x}_j^{(2)}\right), \left(\boldsymbol{\delta}_{-j}, \boldsymbol{x}_{-j}^{(2)}\right)\right)$$

- Properties of the linear-in-characteristics model:
  - Symmetry:
$$\sigma_j(\boldsymbol{w}_j, \boldsymbol{w}_{-j}) = \sigma_k(\boldsymbol{w}_j, \boldsymbol{w}_{-j}) \quad \forall k \neq j$$
  - Anonymity:
$$\sigma(\boldsymbol{w}_j, \boldsymbol{w}_{-j}) = \sigma\left(\boldsymbol{w}_j, \boldsymbol{w}_{\rho(-j)}\right) \forall \rho$$
  - Translation invariant: for any $\boldsymbol{c} \in \mathbb{R}^K$
$$\sigma(\boldsymbol{w}_j + (0, \boldsymbol{c}), \boldsymbol{w}_{-j} + (0, \boldsymbol{c})) = \sigma(\boldsymbol{w}_j, \boldsymbol{w}_{-j})$$

# Re-Express the Demand System

- Express the "state" of the market in *differences* relative to $j$ and treat the outside option just like any other product.
  - Characteristic differences:

$$\boldsymbol{d}_{j,k}^{(2)} = \boldsymbol{x}_k^{(2)} - \boldsymbol{x}_j^{(2)}$$

  - New normalization:

$$\tau_j = \frac{\exp(\delta_j)}{1 + \sum_{j'} \exp(\delta_{j'})}, \forall j = 0, \ldots, n.$$

  - Product $k$ attributes: $\boldsymbol{\omega}_{j,k} = \left(\tau_k, \boldsymbol{d}_{jt,k}^{(2)}\right)$
- Demand for product $j$ is a fully exchangeable function of $\boldsymbol{\omega}_j$:

$$\sigma\left(\boldsymbol{w}_j, \boldsymbol{w}_{-j}\right) = \mathcal{D}(\boldsymbol{\omega}_j)$$

where $\boldsymbol{\omega}_j = \{\omega_{j,0}, \ldots, \omega_{j,j-1}, \omega_{j,j+1}, \ldots, \omega_{j,n}\}$.

# Main Theory Result

- Define the *exogenous* state of the market facing product $j$:

$$\begin{aligned} \boldsymbol{d}_{j,k} &= \boldsymbol{x}_k - \boldsymbol{x}_j \\ \boldsymbol{d}_j &= (\boldsymbol{d}_{j,0}, \ldots, \boldsymbol{d}_{j,j-1}, \boldsymbol{d}_{j,j+1}, \ldots, \boldsymbol{d}_{j,n}) \end{aligned}$$

**Theorem**

*If the distribution of $\{\xi_j\}_{j=1,\ldots,n}$ is exchangeable (conditional on $x_{jt}$), then the reduced form becomes*

$$E\left[\sigma_j^{-1}\left(\boldsymbol{s}, \boldsymbol{x}^{(2)}; \boldsymbol{\Sigma}^0\right)|\boldsymbol{x}\right] = g\left(\boldsymbol{d}_j\right)$$

*where $g$ is a* **symmetric** *function of the state vector.*

- **Implication:** $g$ is a *vector symmetric function* (see Briand 2009)

# Why is it useful?

1. **Curse of dimensionality:** The number of basis functions necessary to approximate the reduced-form is *independent* of the number of products (Pakes (1994), Altonji and Matzkin (2005)).
2. **Example:** Single dimension $d_{jt} = \{x_{1t} - x_{jt}, x_{2t} - x_{jt}, \ldots, x_{J_t,t} - x_{jt}\}$
   - First-order approximation of $g(d)$:

$$g(d_{jt}) \approx \sum_{j'} \gamma^1_{j'} d_{jt,j'} = \gamma^1 \left( \sum_{j'} d_{jt,j'} \right)$$

   - Second-order approximation of $g(d)$:

$$\begin{aligned} g(d_{jt}) &\approx \sum_{j'} \gamma^1_{j'} d_{jt,j'} + \sum_{j'} \gamma^2_{j'} (d_{jt,j'})^2 + \gamma^3 \left( \sum_{j'} d_{jt,j'} \right)^2 \\ &= \gamma^1 \left( \sum_{j'} d_{jt,j'} \right) + \gamma^2 \left( \sum_{j'} (d_{jt,j'})^2 \right) + \gamma^3 \left( \sum_{j'} d_{jt,j'} \right)^2 \end{aligned}$$

# Closing the loop: What is a relevant IV?

- Let $A_j(\boldsymbol{x}_t)$ be an $L$ vector of basis functions summarizing the empirical distribution of characteristic differences: $\{\boldsymbol{d}_{jt,k}\}_{k=0,\ldots,J_t}$.
- **Differentiation IV:** These functions are moments describing the relative isolation of each product in characteristic space.
- **Donald, Imbens, and Newey (2003):** Using basis functions directly as IVs, is asymptotically equivalent to approximating the optimal IV.
  - Recommended practice is to use low-order basis functions (Donald, Imbens, and Newey 2008).

# Practical Suggestions: Polynomial Basis

- **Note:** In general, the first-order basis is weak because it does not vary across products within markets (i.e. sum).
  - In practice you might still want to include the first-order basis terms when you have a lot of entry/exit of products (perhaps interacting with other distance measures).
  - Could be useful also to identify random coefficient on the intercept (rough intuition).
- Single dimension measures of differentiation

$$\text{Quadratic: } A_j(x_t) = \sum_{j'} \left( d^k_{jt,j'} \right)^2$$

Note: $\sqrt{z_{jt,k}}$ is the Euclidian distance between product $j$ and its rivals in market $t$ along dimension $k$.

- Adding interaction terms:

$$\text{Covariance: } A_j(x_t) = \sum_{j'} d^k_{jt,j'} \times d^l_{jt,j'}$$

# Practical Suggestions: Histogram Basis

- **Note:** This approach is advisable only in very large samples, and when the goal is to estimate a very flexible distribution of RCs.
- Single dimension measure of differentiation = Number of rivals in discrete bins

$$A_j(x_t) = \left\{ \sum_{j'} 1\left( d^k_{jt,j'} < \kappa_l \right) \right\}_{l=1,\ldots,L}$$

- Multi-dimension measure of differentiation:

$$A_j(x_t) = \left\{ \sum_{j'} 1\left( d^k_{jt,j'} < \kappa_l \right) 1\left( d^{k'}_{jt,j'} < \kappa_{l'} \right) \right\}_{l=1,\ldots,L,l'=1,\ldots,L}$$

# Practical Suggestions: Local Basis

- **Note:** In most parametric models, the inverse demand is function of characteristics of **close-by** rivals. Therefore, in the previous histogram basis, we should be focussing on "local" rivals.
- Single dimension measure of differentiation = Number of nearby rivals along each dimension

$$A_j(x_t) = \sum_{j'} 1\left(|d^k_{jt,j'}| < \kappa_k\right), \text{ e.g. } \kappa_k = sd(x_{jt,k})$$

- Multi-dimension measure of differentiation:

$$A_j(x_t) = \sum_{j'} 1\left(|d^k_{jt,j'}| < \kappa_k\right) \times d_{jt,l}, \text{ e.g. } \kappa_k = sd(x_{jt,k})$$

- When $x_{jt,k}$ is discrete, this basis function boils down to the familiar Nested-logit IVs.
  - Number of competitors and characteristics of rivals within segment. See Bresnahan, Stern, and Trajtenberg (1997).

# Practical Suggestions: Demographics

- In many settings, product characteristics are fixed across markets, but the distribution of consumer types vary (e.g. Nevo 2001).
- To fix ideas, focus on a single non-linear characteristics $x_j^{(2)}$
- Consumer valuation for $x_j^{(2)}$ is

$$\beta_{it} = z_{it}\pi + \nu_i$$

  where $\nu_i \sim N(0, \sigma_x^2)$.
- **Assumption:** The distribution of demographics across markets is known, and can be decomposed as follows: $z_{it} = \mu_t + \text{sd}_t e_{it}$, where $e_{it} \sim F(\cdot)$ and $F(\cdot)$ is common across markets.
  - Example: BLP95 assume that the income distribution is log-normal with market-specific mean/variance.

# Practical Suggestions: Demographics

- Demand function:

$$\sigma_{jt}(\delta_t, x^{(2)}|\pi, \sigma_x) =$$

$$= \int\int \frac{\exp\left(\delta_{jt} + z_{it}\pi x_j^{(2)} + \nu_i x_j^{(2)}\right)}{1 + \sum_{j'} \exp\left(\delta_{j't} + z_{it}\pi x_{j'}^{(2)} + \nu_i x_{j'}^{(2)}\right)} dF_t(z_{it})\phi(\nu_i; \Sigma)$$

$$= \int\int \frac{\exp\left(\delta_{jt} + \pi e_{it}\sigma_t x_j^{(2)} + \pi\mu_t x^{(2)} + \nu_i x_j^{(2)}\right)}{1 + \sum_{j'} \exp\left(\delta_{j't} + \pi e_{it}\sigma_t x_{j'}^{(2)} + \pi\mu_t x_{j'}^{(2)} + \nu_i x_{j'}^{(2)}\right)} dF(e_{it})\phi(\nu_i; \Sigma)$$

$$= \sigma_j(\delta_j, x^{(2)}, \underbrace{\sigma_t x^{(2)}, \mu_t x^{(2)}}_{\text{new state variables}} |\pi, \sigma_x)$$

- **Implication:** Demand is a symmetric function of characteristic differences and demographic moments: $\mathcal{D}(\omega_t, d^{(2)}, \sigma_t d^{(2)}, \mu_t d^{(2)}|\theta)$.
- The previous result therefore applies to the reduced-form of this transformed model:

$$E\left[\sigma_{jt}^{-1}(s_t, x^{(2)}|\pi, \sigma_x)|x_t, \mu_t, \sigma_t\right] = g(d_t, \mu_t d^{(2)}, \sigma_t d^{(2)})$$

# Practical Suggestions: Demographics

- Differentiation IVs with demographics:

$$A_t(x_t, \mu_t, \sigma_t) = \sum_{j'} 1\left(|d^k_{jt,j'}| < \kappa_k\right) \times \mu_t$$

$$A_t(x_t, \mu_t, \sigma_t) = \sum_{j'} 1\left(|d^k_{jt,j'}| < \kappa_k\right) \times \sigma_t$$

$$A_t(x_t, \mu_t, \sigma_t) = \sum_{j'} 1\left(|d^k_{jt,j'}| < \kappa_k\right) \times \sigma_t \times d^l_{jt,j'}$$

- When the distribution of demographics can be "standardized" across markets, this characterization is exact.
- More generally, Differentiation IVs should be interacted with rich moments of the distribution of consumer characteristics.
- **Example:** Miravete, Seim, and Thurk (2017)
  - Combine nested-logit 'type' instruments, with moments of the distribution of demographics across stores.

# Experiment 1: Independent Random Coefficients

- Random coefficient model:

$$u_{ijt} = \delta_{jt} + \sum_{k=1}^{K} v_{ik} x_{jt,k}^{(2)} + \varepsilon_{ijt}, \quad v_i \sim N(0, \sigma_x^2 \boldsymbol{I}).$$

- Data:
  - Panel structure: 100 markets $\times$ 15 products
  - Characteristics: $(\xi_{jt}, \boldsymbol{x}_{jt}) \sim N(0, \mathbf{I})$.
  - Dimension: $|\boldsymbol{x}_{jt}| = K + 1$
  - Monte-Carlo replications $=$ 1,000
- Differentiation IVs $(K + 1)$:
  - **Quadratic:** $A_j(\boldsymbol{x}_t) = \sum_{j'=1}^{J_t} \left(d_{jt,j'}^k\right)^2, \forall k = 1, \ldots, K$

# Simulation Results: Quadratic Differentiation IVs

| | $K_2 = 1$ | | $K_2 = 2$ | | $K_3 = 3$ | | $K_3 = 4$ | |
|---|---|---|---|---|---|---|---|---|
| | bias | rmse | bias | rmse | bias | rmse | bias | rmse |
| $\log \sigma_1$ | 0.00 | 0.03 | -0.00 | 0.03 | -0.00 | 0.03 | -0.00 | 0.04 |
| $\log \sigma_2$ | | | -0.00 | 0.03 | 0.00 | 0.03 | -0.00 | 0.04 |
| $\log \sigma_3$ | | | | | -0.00 | 0.03 | -0.00 | 0.03 |
| $\log \sigma_4$ | | | | | | | -0.00 | 0.04 |
| | | | | | | | | |
| $\sigma_1$ | 0.00 | 0.12 | 0.00 | 0.13 | -0.00 | 0.13 | -0.00 | 0.14 |
| $\sigma_2$ | | | -0.00 | 0.13 | 0.00 | 0.13 | -0.00 | 0.14 |
| $\sigma_3$ | | | | | 0.00 | 0.13 | -0.00 | 0.14 |
| $\sigma_4$ | | | | | | | -0.00 | 0.15 |
| | | | | | | | | |
| 1(Local) | 0.00 | | 0.00 | | 0.00 | | 0.00 | |
| Rank-test | 1202.10 | | 564.03 | | 330.40 | | 206.42 | |
| pv | 0.00 | | 0.00 | | 0.00 | | 0.00 | |
| IIA-test | 359.41 | | 363.22 | | 321.73 | | 276.13 | |
| pv | 0.00 | | 0.00 | | 0.00 | | 0.00 | |

# Experiment 2: Correlated Random Coefficients

- Consumer heterogeneity:

$$\beta_i^{(2)} \sim N(\beta^{(2)}, \boldsymbol{\Sigma})$$

- 4 dimensions $\Rightarrow$ 10 non-linear parameters (choleski)
- Panel structure:

$$100 \text{ markets} \times 50 \text{ products}$$

- **Differentiation IVs:** Second-order polynomials (with interactions):

$$A_j(x_t) = \sum_{j'=1}^{J_t} \left( d_{jt,j'}^k \times d_{jt,j'}^l \right)$$

for all characteristics $k <= l$.

# Simulation Results: Correlated Random-Coefficients

| | $\Sigma_{\cdot,1}$ | $\Sigma_{\cdot,2}$ | $\Sigma_{\cdot,3}$ | $\Sigma_{\cdot,4}$ |
|---|---|---|---|---|
| $\Sigma_{1,\cdot}$ | 0.003 | 0.003 | -0.003 | 0.010 |
| $\Sigma_{2,\cdot}$ | 0.003 | 0.000 | 0.004 | -0.000 |
| $\Sigma_{3,\cdot}$ | -0.003 | 0.004 | -0.009 | 0.006 |
| $\Sigma_{4,\cdot}$ | 0.010 | -0.000 | 0.006 | 0.010 |
| $\Sigma_{1,\cdot}$ | 0.228 | 0.132 | 0.156 | 0.156 |
| $\Sigma_{2,\cdot}$ | 0.132 | 0.232 | 0.145 | 0.143 |
| $\Sigma_{3,\cdot}$ | 0.156 | 0.145 | 0.217 | 0.154 |
| $\Sigma_{4,\cdot}$ | 0.156 | 0.143 | 0.154 | 0.217 |
| IIA test (F) | 157.637 | | | |
| Cragg-Donald stat. | 474.053 | | | |
| Nb endo. | 10.000 | | | |
| Nb IVs | 15.000 | | | |

# How to account for endogenous prices?

- "Quality-ladder" example:

$$u_{ijt} = \delta_{jt} - \alpha_i p_{jt} + \varepsilon_{ijt}$$

where $\alpha_i = \sigma_p y_i^{-1}$, and $\log(y_i) \sim \mathrm{N}(\mu_y, \sigma_y)$ (known).

- Excluded price instruments: $\boldsymbol{w}_t = \{w_{jt}\}_{j=1,\dots,J_t}$

- Reduced-form:

$$E\left[\sigma_j^{-1}\left(\boldsymbol{s}_t, \boldsymbol{x}_t, \boldsymbol{p}_t | \sigma_p^0\right) | \boldsymbol{x}_t, \boldsymbol{w}_t\right] \neq g(\boldsymbol{d}_j^x, \boldsymbol{d}_j^p),$$

where the inequality is due to the simultaneity of prices and $\xi_{jt}$ (BLP, 1995).

# How to incorporate endogenous prices?

- **Heuristic solution:** Distribute the expectation for price *inside* of the inverse-demand (BLP, 1999):

$$
\begin{aligned}
E\left[\sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \Sigma) | \boldsymbol{x}_t, \boldsymbol{w}_t\right] &\approx E\left[\sigma_j^{-1}(\boldsymbol{s}_t, \hat{\boldsymbol{p}}_t, \boldsymbol{x}_t^{(2)}; \Sigma) | \boldsymbol{x}_t, \hat{\boldsymbol{p}}_t\right] \\
&= g(\boldsymbol{d}_{jt}^{x}, \boldsymbol{d}_{jt}^{\hat{p}})
\end{aligned}
$$

where $\hat{d}_{jt,k}^{p} = E(p_{kt} | \boldsymbol{w}_{kt}) - E(p_{jt} | \boldsymbol{w}_{jt})$.

# Experiment 3: Differentiation IVs with Endogenous Prices

## Example with cost shifter

1. Exogenous price index (OLS):

$$\hat{p}_{jt} = \hat{\pi}_0 + \hat{\pi}_1 x_{jt} + \hat{\pi}_2 \omega_{jt}$$

2. Differentiation IV: Quadratic

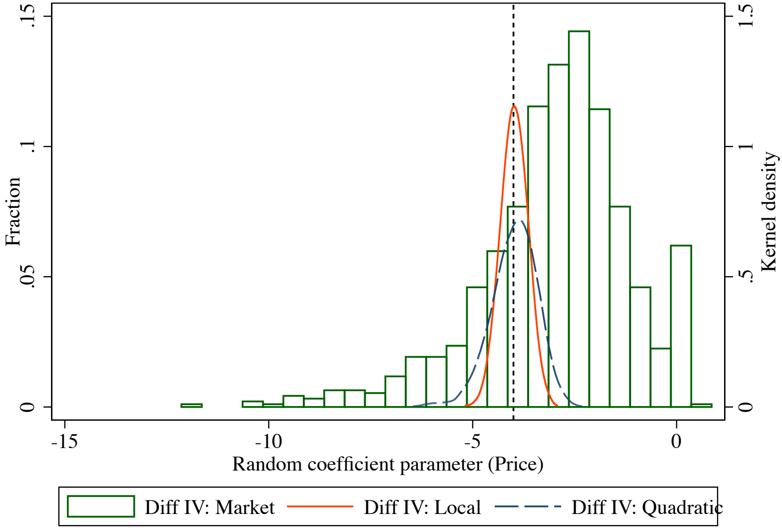
$$\sum_{j'} \left( d^{\hat{p}}_{jt,j'} \right)^2 \text{ and } \sum_{j'} \left( d^{\hat{p}}_{jt,j'} \right)^2 \cdot \boldsymbol{d}_{jt,j'}$$

   where $\boldsymbol{d}_{jt,j'} = (d^{x}_{jt,j'}, d^{\hat{p}}_{jt,j'})$.

3. Differentiation IV: Local

$$\sum_{j'} \left( |d^{\hat{p}}_{jt,j'}| < \text{sd}(\hat{p}_{jt}) \right) \text{ and } \sum_{j'} \left( |d^{\hat{p}}_{jt,j'}| < \text{sd}(\hat{p}_{jt}) \right) \cdot \boldsymbol{d}_{jt,j'}$$

# Distribution of $\hat{\sigma}_p$ with weak and strong IVs




Dash vertical line = True parameter value

# GMM estimates with endogenous prices

| | True | Diff. IV = Local | | | Diff. IV = Quadratic | | | Diff. IV = Sum | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | bias | se | rmse | bias | se | rmse | bias | se | rmse |
| $\sigma_p$ | -4.00 | 0.02 | 0.27 | 0.28 | 0.02 | 0.53 | 0.55 | 1.03 | 158.25 | 2.10 |
| $\beta_0$ | 50.00 | -0.26 | 3.92 | 3.92 | -0.28 | 7.36 | 7.45 | -9.82 | 26.41 | 20.65 |
| $\beta_x$ | 2.00 | -0.02 | 0.46 | 0.45 | -0.02 | 0.47 | 0.47 | 0.34 | 1.11 | 0.83 |
| $\beta_p$ | -0.20 | 0.01 | 0.37 | 0.37 | 0.01 | 0.31 | 0.32 | -0.67 | 201.29 | 1.38 |

# GMM estimates with endogenous prices

| | Diff. IV = Local | Quadratic | Diff. IV = Sum |
|---|---|---|---|
| Frequency conv. | 1 | 1 | 0.94 |
| IIA-test | 109.48 | 53.90 | 1.88 |
| p-value | 0 | 0 | 0.34 |
| 1st-stage F-test: Price | 191.80 | 442.10 | 138.94 |
| 1st-stage F-test: Jacobian | 214.60 | 58.40 | 27.85 |
| Cond. 1st-stage F-test: Price | 252.23 | 479.96 | 7.92 |
| Cond. 1st-stage F-test: Jacobian | 280.31 | 82.44 | 6.19 |
| Cragg-Donald statistics | 170.19 | 54.45 | 4.09 |
| Stock-Yogo size CV (10%) | 16.87 | 13.43 | 13.43 |
| Nb. endogenous variables | 2 | 2 | 2 |
| Nb. IVs | 4 | 3 | 3 |

**Note:** The *Conditional 1st-stage F-test* statistic is the Weak IV test proposed by Angrist and Pischke for multiple endogenous variables.

# Experiment 4: Natural Experiments

- **Hotelling example:** Exogenous entry of a new product ($x' = 5$)

$$u_{ijmt} = \delta_{jmt} - \lambda|\nu_i - x_{jmt}| + \epsilon_{ijmt}$$

- Three-way panel: product $j$, market $m$, and time ($t = 0, 1$).
- Treatment variable:

$$D_{jm} = 1\left(|x_{jm} - 5| < \text{Cutoff}\right)$$

- Reduced-form: Difference-in-difference regression

$$\sigma_j^{-1}(s_t, x_t, p_t|\theta^0) = \mu_{jm} + \tau_t + \gamma D_{jm} \times 1(t = 1) + \xi_{jmt}$$

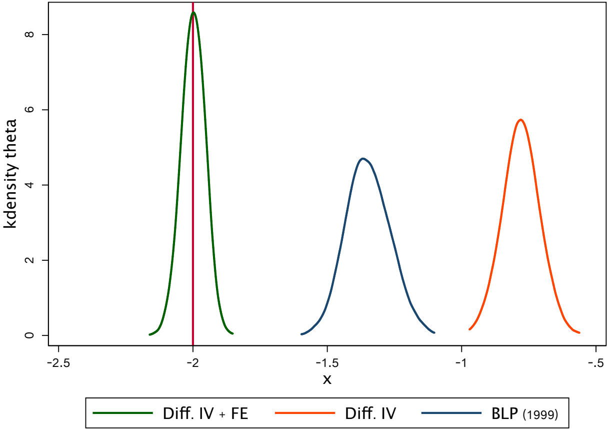
- **GMM:** DiD IVs
  - Linear characteristics: $x_{jmt}^{(1)}$ = Market/Product FE + After Dummy
  - Differentiation IV: $z_{jmt} = D_{jm} \times 1(t = 1)$
  - $\hat{\theta}^{gmm}$ is identified from the DiD variation in $z_{jmt}$.

# Natural Experiment: Hotelling Example

DGP: $\delta_{jmt} = \bar{\xi}_{jm} + \Delta\xi_{jmt}$, where $E(\bar{\xi}_{jm}|\boldsymbol{x}_m) \neq 0$

- "Diff-in-Diff" specification:

$$\boldsymbol{z}_{jmt} = \{\text{Product Dummy}_{jm}, 1(t=1), 1(|x_{jm} - 5| < 1)1(t=1)\}$$



# Review of Nonlinear GMM

- The GMM problem is defined as:

$$\min_{\beta,\Sigma} \quad = \quad \rho(\theta)' Z \hat{W}^{-1} Z' \rho(\theta)$$
$$\text{s.t.} \qquad \rho_j(s_t, x_t|\theta) = \sigma_j^{-1}(s_t, x_t^{(2)}|\Sigma) - x_{jt}\beta$$

- The FOC of the problem is (or moment conditions):

$$\frac{\partial \rho(\theta)'}{\partial \theta} Z \hat{W}^{-1} Z' \rho(\theta) = 0$$

where $\frac{\partial \rho(\theta)}{\partial \theta}$ is a $n \times m$ matrix containing the **Jacobian** of the residual function.

- In our context:

$$\frac{\partial \rho_j(s_t, x_t|\theta)}{\partial \theta} = \begin{cases} -x_{jt,k} & \text{If } \theta_k = \beta_k \\ \frac{\partial \sigma_j^{-1}(s_t, x_t \theta_k)}{\partial \Sigma_k} & \text{Else.} \end{cases}$$

The derivative of the inverse demand can be computed using the implicit function theorem (see Nevo (2001)).

# Review of Nonlinear GMM

- Let $J_{jt}(\theta) = \frac{\partial \rho_j(s_t, x_t|\theta)}{\partial \theta}$ denotes the matrix of Jacobian.
- Notice that the moment conditions imposed by GMM correspond to the moment conditions associated with a **linear approximation** of the model.
- The moment conditions at $\theta^{gmm}$ is:

$$J(\theta)' Z \hat{W}^{-1} Z' \rho(\theta) = 0$$

- This is the moment condition of the following linear IV regression:

$$\begin{aligned} \rho_{jt}(\theta) &= J_{jt}(\theta) b + \text{Error} \\ &= \sum_k (\theta_k - \theta_k^0) \frac{\partial \rho_j(s_t, x_t|\theta)}{\partial \theta} + \text{Error} \end{aligned}$$

  where $b = 0$.
- Therefore, $b^{gmm} = E(\theta^{gmm} - \theta^0) = 0$ (i.e. GMM is consistent).

# Gauss-Newon Regression

- The linear representation can be used to construct a Gauss-Newton algorithm to estimate $\{\hat{\theta}\}$.

$$\rho_{jt}(\theta) \quad = \quad J_{jt}(\theta)b + \text{Error}$$

- Iteration $k \geq 1$:
  1. Invert demand: $\sigma_{jt}^{-1}(s_t, x_t|\theta^{k-1})$ and $J_{jt}(s_t, x_t|\theta^{k-1})$
  2. Estimate $\{\hat{b}^k\}$ by linear GMM.
  3. If $|\hat{b}^k| < \varepsilon$ stop. Else, update $\theta^k = \theta^{k-1} + \hat{b}^k$, and repeat steps (1)-(3).
- With strong instruments, this procedure typically requires less than 5 iterations.
  - Weak IVs lead to severe numerical problems (e.g. Knittle and Metaxoglou (2014), Dube et al. (2012))
  - Why? The central-limit theorem does no hold, and the quadratic approximation is bad.

# Using the Gauss-Newton Regression for Inference

- The Gauss-Newton regression is also useful to conduct inference
- Let $\hat{\theta}$ and $\hat{W}$ denote GMM estimate and weighting matrix (estimated using Julia or any other non-linear optimization package)
- Since non-linear GMM is equivalent to linear IV at the solution, we can conduct inference on $\hat{\theta}$ using the GNR:

$$\rho_{jt}(\hat{\theta}) = J_{jt}(\hat{\theta})\hat{b} + \text{Error}$$

- **Note:** $\hat{b}$ must be **zero** if $\hat{\theta}$ is the GMM solution (i.e. you must be using the same weighing matrix in Julia or Matlab as in STATA or R to run this regression)
- But, variance-covariance matrix of $\hat{\theta}$ is the same as the variance-covariance matrix of $\hat{\theta}$. Therefore, you can use standard statistical routines to calculate standard errors and conduct hypothesis tests on $\theta$ (e.g. cluster standard errors, equality of parameters, etc.).

# Using the Gauss-Newton Regression for to evaluate the Strength of IVs

- This insight is particularly useful to evaluate the relevance of the instruments
- Some weak identification tests are non trivial to code, and are standard in STATA and R
- **Examples:** IVREG2 in STATA reports the Cragg-Donald and the Sanderson-Windmeijer (SW) first-stage tests.
- **Recommeded procedure:**
  - Use Julia's non-linear optimization packages to solve the GMM problem
  - Compute the Jacobian of the inverse demand
  - Expor the Jacobian and weighting matrix to STATA or R (or even better load R in Julia...)
  - Perform inference and weak IV tests using the Gauss-Newton regression

# Ex-Ante Weak IV test: IIA Hypothesis

- **Warning:** The weak IV test based on the Jacobian function at $\hat{\theta}$ are not consistent when the instruments are too weak. You should therefore not put too much weight on the p-values, and combine your analysis with the IIA test.
- A **strong** instrument for $\Sigma$ is able to reject the wrong model (Stock and Wright, 2000)
- Under $H_0 : \Sigma = 0$, the inverse demand equation is independent of $x_{-j}$:

$$\sigma_{jt}^{-1}(s_t, x_t; \Sigma = 0) = \ln s_{jt}/s_{0t} = x_{jt}\beta + z_{jt}\gamma + \xi_{jt}$$

- Standard test statistics for $H_0 : \gamma = 0$, can be used to test null hypothesis of IIA preferences
- With endogenous prices, this test is equivalent to the J-test at $\Sigma = 0$.
- In practice, you should report both this *ex-ante* weak IV test, and the *ex-post* specification tests.

Anderson, S., A. de Palma, and J.-F. Thisse (1992).
*Discrete Choice Theory of Product Differentiation*.
MIT Press.

Berry, S. (1994).
Estimating discrete choice models of product differentiation.
*Rand Journal of Economics 25*, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1999).
Voluntary export restraints on automobiles: Evaluating a trade policy.
*American Economic Review 89*(3), 400–430.

Berry, S. T., A. Gandhi, and P. Haile (2013).
Connected substitutes and invertibility of demand.
*Econometrica 81*(5), 2087–2111.

Bresnahan, T., S. Stern, and M. Trajtenberg (1997).
Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s.
*The RAND Journal of Economics 28*, s17–s44.

Bresnahan, T. F. (1987).
Competition and collusion in the american automobile industry: The 1955 price war.
*The Journal of Industrial Economics 35*(4, The Empirical Renaissance in Industrial Economics), 457–482.

Hotelling, H. (1929).
Stability in competition.
*Economic Journal 39*(153), 41–57.

Lancaster, K. J. (1966, April).
A new approach to consumer theory.
*American economic Review 74*(2), 132–157.

McFadden, D. (1974).
*Conditional Logit Analysis of Qualitative Choice Behavior*, Chapter 4, pp. 105–142.
Academic Press: New York.

Miller, N. H. and M. Weinberg (2016, July).
Mergers facilitate tacit collusion: An empirical investigation of the miller/coors joint venture.
Working paper, Drexel University.

Miravete, E., K. Seim, and J. Thurk (2017, September).
Market power and the laffer curve.
working paper, UT Austin.

Nevo, A. (2001).
Measuring market power in the ready-to-eat cereal industry.
*Econometrica 69*(2), 307.

Newey, W. K. and J. L. Powell (2003, September).
Instrumental variable estimation of nonparametric models.
*Econometrica 71*(5), 1565–1578.

Petrin, A. (2002).
Quantifying the benefits of new products: The case of the minivan.
*Journal of Political Economy 110*, 705.

Shaked, A. and J. Sutton (1982, January).
Relaxing price competition through product differentiation.
*The Review of Economic Studies 48*(1), 3–13.