Measuring Substitution Patterns in Differentiated Product Industries

Amit Gandhi
University of Pennsylvania

Jean-François Houde
UW-Madison & NBER

April 9, 2019

Motivation

Motivation

- **The Beauty of BLP:** Flexible estimation of substitution patterns with *many* products, aggregate data, and unobserved attributes.
  - Workhorse model to study demand for differentiated products in IO
  - Increasingly used to analyse *sorting* problems in urban, education, insurance, etc.

Motivation

- **The Beauty of BLP:** Flexible estimation of substitution patterns with *many* products, aggregate data, and unobserved attributes.
  - Workhorse model to study demand for differentiated products in IO
  - Increasingly used to analyse *sorting* problems in urban, education, insurance, etc.

- Achieving this flexibility can be difficult in practice...
  - **Precision:** Often rely on external restrictions (e.g. supply, survey, etc.)
  - **Numerical:** Multiple solutions and/or poor convergence properties

What explains the difficulties in practice?

What explains the difficulties in practice?

- Is the variation in data simply too weak?

What explains the difficulties in practice?

- Is the variation in data simply too weak?
- Or is it weakness of the instruments (IVs)?
  - i.e., Are we using the variation in the data in the optimal way?

What explains the difficulties in practice?

- Is the variation in data simply too weak?
- Or is it weakness of the instruments (IVs)?
  - i.e., Are we using the variation in the data in the optimal way?
- **Our paper** argue that many *empiricists'* problems are caused by **weak** IVs

What explains the difficulties in practice?

- Is the variation in data simply too weak?
- Or is it weakness of the instruments (IVs)?
  - i.e., Are we using the variation in the data in the optimal way?
- **Our paper** argue that many *empiricists'* problems are caused by **weak** IVs
- Show how to construct **strong** IVs using a new representation of the reduced-form of the model.

Key Takeaways

- **Differentiation IV:** Capture the relative position of each product in the characteristic space
  - Approximate optimal IV without requiring initial estimates
  - Simple to construct and test

Key Takeaways

- **Differentiation IV:** Capture the relative position of each product in the characteristic space
  - Approximate optimal IV without requiring initial estimates
  - Simple to construct and test
- Powerful in practice:
  - 10+ improvement in precision
  - Fast convergence + Numerically stable
  - Flexible substitution: Multiple dimensions + Correlated heterogeneity

Key Takeaways

- **Differentiation IV:** Capture the relative position of each product in the characteristic space
  - Approximate optimal IV without requiring initial estimates
  - Simple to construct and test
- Powerful in practice:
  - 10+ improvement in precision
  - Fast convergence + Numerically stable
  - Flexible substitution: Multiple dimensions + Correlated heterogeneity
- Related work:
  - **Weak Identification:** BLP (1999), Conlon (2013), Reynaert & Verboven (2013), Metaxoglou and Knittel (2014)

Key Takeaways

- **Differentiation IV:** Capture the relative position of each product in the characteristic space
  - Approximate optimal IV without requiring initial estimates
  - Simple to construct and test
- Powerful in practice:
  - 10+ improvement in precision
  - Fast convergence + Numerically stable
  - Flexible substitution: Multiple dimensions + Correlated heterogeneity
- Related work:
  - **Weak Identification:** BLP (1999), Conlon (2013), Reynaert & Verboven (2013), Metaxoglou and Knittel (2014)
  - **Differentiation IV:** *Nested-Logit* (e.g. Berry (1994), Verboven (1996), Bresnahan et al. (1997)), and *Spatial Differentiation* (e.g. Pinkse and Slade (2001), Davis (2006), Thomadsen (2005), Manuszak (2012), Houde (2012))

Baseline Model: Exogenous Characteristics

- **Data:** Market shares (s_{jt}) and characteristics (x_{jt}) observed in T independent markets.
  - Each market includes J_t products + an outside option ($x_{0t} = 0$).

Baseline Model: Exogenous Characteristics

- **Data:** Market shares (s_{jt}) and characteristics (x_{jt}) observed in T independent markets.
  - Each market includes J_t products + an outside option ($x_{0t} = 0$).
- **Demand:** Linear random-coefficient with T1EV random utility shocks

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

where $\delta_{jt} = \beta_0 + x_{jt}^{(1)}\beta_1 + x_{jt}^{(2)}\beta_2 + \xi_{jt}$.

Baseline Model: Exogenous Characteristics

- **Data:** Market shares (s_{jt}) and characteristics (x_{jt}) observed in T independent markets.
  - Each market includes J_t products + an outside option ($x_{0t} = 0$).
- **Demand:** Linear random-coefficient with T1EV random utility shocks

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

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:

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

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, $A_j(x_t)$:

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

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, $A_j(x_t)$:

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

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

Illustration of the Weak IV Problem

- Two detection tests:

Illustration of the Weak IV Problem

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

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

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

$$H_0 : \gamma = 0$$

Illustration of the Weak IV Problem

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

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

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

  2. Local identification: Cragg-Donald rank test

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

$$\leftrightarrow J_{jt,k}(\theta) = z_t\pi^k + u_{jt,k}$$

where π^k are the "reduced-form" parameters of the model. This test can be implemented in STATA (ivreg2 or ranktest).

Illustration of the Weak IV Problem

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

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

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

$$H_0 : \gamma = 0$$

  2. Local identification: Cragg-Donald rank test

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

$$\leftrightarrow J_{jt,k}(\theta) = z_t\pi^k + u_{jt,k}$$

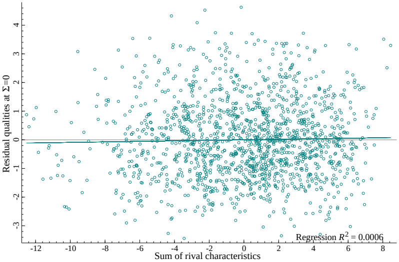
  where π^k are the "reduced-form" parameters of the model. This test can be implemented in STATA (ivreg2 or ranktest).

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

Weak Identification in a Picture: IIA Test

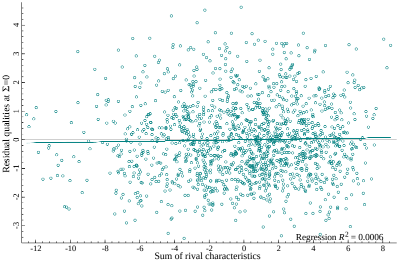
Weak Identification in a Picture: IIA Test

(A) IV: Sum of rivals' characteristics

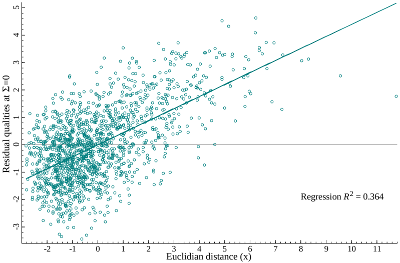


Weak Identification in a Picture: IIA Test

(A) IV: Sum of rivals' characteristics

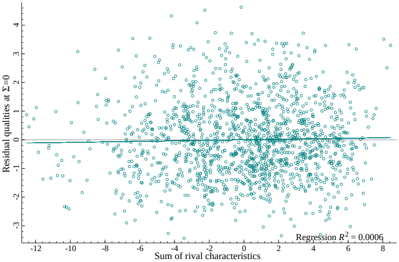


(B) IV: Euclidean distance in x

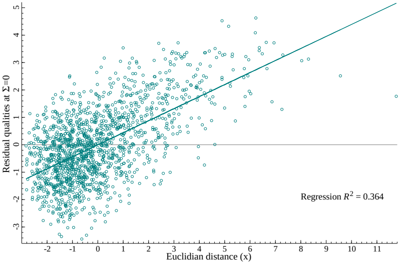


Weak Identification in a Picture: IIA Test

(A) IV: Sum of rivals' characteristics

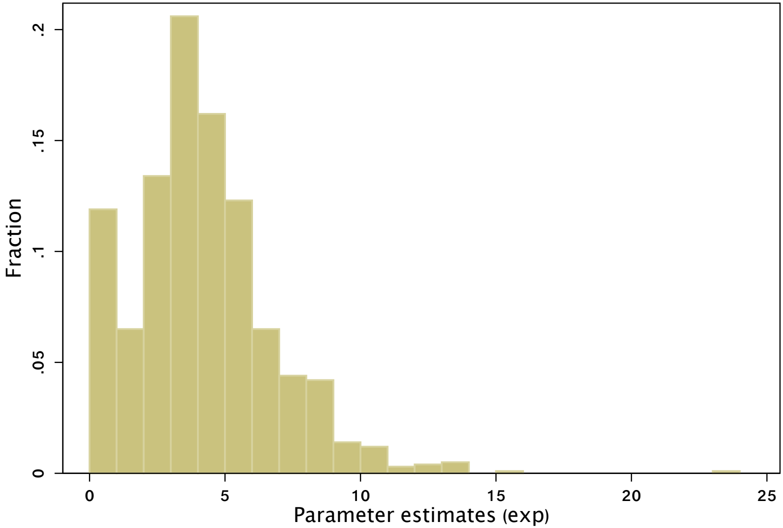


(B) IV: Euclidean distance in x



- **Takeaway:** Independence of ξ_{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
σ_1	0.14	2.64	-0.01	2.49	-0.03	2.19	0.22	2.35
σ_2			0.12	2.42	-0.01	2.27	0.10	2.30
σ_3					0.18	2.38	0.11	2.38
σ_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
σ_1	0.14	2.64	-0.01	2.49	-0.03	2.19	0.22	2.35
σ_2			0.12	2.42	-0.01	2.27	0.10	2.30
σ_3					0.18	2.38	0.11	2.38
σ_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
σ_1	0.14	2.64	-0.01	2.49	-0.03	2.19	0.22	2.35
σ_2			0.12	2.42	-0.01	2.27	0.10	2.30
σ_3					0.18	2.38	0.11	2.38
σ_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	

Identification problem

- **Simultaneous equation:** Reduced-form *vs* structural equation

$$\rho_j(s_t, x_t; \theta) = \underbrace{\sigma_j^{-1}\left(s_t, x_t^{(2)}; \lambda^0\right)}_{\text{Structural equation}} - x_{jt}\beta$$

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

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

Identification problem

- **Simultaneous equation:** Reduced-form *vs* structural equation

$$\rho_j(s_t, x_t; \theta) = \underbrace{\sigma_j^{-1}\left(s_t, x_t^{(2)}; \lambda^0\right)}_{\text{Structural equation}} - x_{jt}\beta$$

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

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

- **Example:** Quasi-linear utility with exogenous prices

$$u_{ijt} = x_{jt}b_i - p_{jt} + \xi_{jt} + \epsilon_{ijt}; b_i = \beta + \lambda\eta_i$$

$$\rightarrow p_{jt} = x_{jt}\beta + \sigma_j^{-1}(s_t, x^{(2)}; \lambda) + \xi_{jt} = \text{Non-linear IV regression}$$

Identification problem

- **Simultaneous equation:** Reduced-form *vs* structural equation

$$\rho_j(s_t, x_t; \theta) = \underbrace{\sigma_j^{-1}\left(s_t, x_t^{(2)}; \lambda^0\right)}_{\text{Structural equation}} - x_{jt}\beta$$

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

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

- **Example:** Quasi-linear utility with exogenous prices

$$u_{ijt} = x_{jt}b_i - p_{jt} + \xi_{jt} + \epsilon_{ijt}; b_i = \beta + \lambda\eta_i$$

$$\rightarrow p_{jt} = x_{jt}\beta + \sigma_j^{-1}(s_t, x^{(2)}; \lambda) + \xi_{jt} = \text{Non-linear IV regression}$$

- **Insight from Berry & Haile:** 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 construct relevant instrument?

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

How to construct relevant instrument?

- Since $\dim(x_t) >> dim(\lambda) = m$, any transformation of $x_t = \{x_{1t}, \ldots, x_{J_t,t}\}$ can be used to construct valid moments.
- **Definition:** An "efficient" instrument is a (basis) function $A^L(x_t)$ of dimension L, that can approximate the reduced-form of the model arbitrarily well:

$$E\left[\sigma_j^{-1}\left(s_t, x_t^{(2)}; \lambda^0\right)|x_t\right] - A_j^L(x_t)\hat{\gamma}_L \rightarrow 0, \text{ as } L \text{ and } n \text{ get large.}$$

  - Where $\hat{\gamma}_L$ are OLS coefficients obtained by projecting σ^{-1} onto $A_j^L(x_t)$.

How to construct relevant instrument?

- Since $\dim(x_t) >> dim(\lambda) = m$, any transformation of $x_t = \{x_{1t}, \ldots, x_{J_t,t}\}$ can be used to construct valid moments.
- **Definition:** An "efficient" instrument is a (basis) function $A^L(x_t)$ of dimension L, that can approximate the reduced-form of the model arbitrarily well:

$$E\left[\sigma_j^{-1}\left(s_t, x_t^{(2)}; \lambda^0\right)|x_t\right] - A_j^L(x_t)\hat{\gamma}_L \to 0, \text{ as } L \text{ and } n \text{ get large.}$$

  - Where $\hat{\gamma}_L$ are OLS coefficients obtained by projecting σ^{-1} onto $A_j^L(x_t)$.
- The same basis functions can be used to construct Chamberlain (1987)'s optimal instruments. See Newey (1990).

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.
- Without further restrictions, we cannot directly use the insights of BH to construct relevant IVs

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

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

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

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

What does the characteristic structure imply about 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:

What does the characteristic structure imply about 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:
  $$\sigma_j\left(\boldsymbol{w}_j, \boldsymbol{w}_{-j}\right) = \sigma_k\left(\boldsymbol{w}_j, \boldsymbol{w}_{-j}\right) \quad \forall k \neq j$$
  - Anonymity:
  $$\sigma\left(\boldsymbol{w}_j, \boldsymbol{w}_{-j}\right) = \sigma\left(\boldsymbol{w}_j, \boldsymbol{w}_{\rho(-j)}\right) \forall \rho$$

What does the characteristic structure imply about 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.

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: $\boldsymbol{\omega}_{j,k} = \left(\tau_k, \boldsymbol{d}_{jt,k}^{(2)}\right)$

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:

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

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

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

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

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

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

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 and markets (Pakes (1994), Altonji and Matzkin (2005)).

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 and markets (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 $A_j\left(\boldsymbol{x}_t\right)$ be an L vector of basis functions summarizing the empirical distribution of characteristic differences: $\{\boldsymbol{d}_{jt,k}\}_{k=0,\ldots,J_t}$.

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.

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).

Suggestion 1: Polynomial Basis

- 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:

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

Suggestion 1: Polynomial Basis

- 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:

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

- **Note:** In general, the first-order basis is weak because it does not vary across products within markets (i.e. sum of rival characteristics).

Suggestion 2: Histogram Basis

- Single dimension measure of differentiation = Number of rivals in discrete bins

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

- Multi-dimension measure of differentiation:

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

Suggestion 2: Histogram Basis

- Single dimension measure of differentiation = Number of rivals in discrete bins

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

- Multi-dimension measure of differentiation:

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

- **Note:** This approach is advisable only in very large samples (+large choice-sets), and when the goal is to estimate a flexible distribution of RCs (e.g. correlation terms)

Suggestion 3: Local Basis

- In most parametric models, the inverse demand is function of characteristics of **close-by** rivals. Therefore, the characteristics of "nearby" rivals should more relevant.

Suggestion 3: Local Basis

- In most parametric models, the inverse demand is function of characteristics of **close-by** rivals. Therefore, the characteristics of "nearby" rivals should more relevant.
- 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 (e.g. Berry (1994), Bresnahan et al. (1997)).

Suggestion 4: 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.

Suggestion 4: Demographics

- Demand function:

$$
\begin{aligned}
&\sigma_{jt}(\delta_t, x^{(2)}|\pi, \sigma_x) = \\
&= \int\int \frac{\exp\left(\delta_{jt} + z_{it}\pi x_j^{(2)} + \nu_i x_j^{(2)}\right)}{1 + \sum_{j'} \exp\left(\delta_{j't} + z_{it}\pi x_{j'}^{(2)} + \nu_i x_{j'}^{(2)}\right)} dF_t(z_{it})\phi(\nu_i; \lambda) \\
&= \int\int \frac{\exp\left(\delta_{jt} + \pi e_{it}\sigma_t x_j^{(2)} + \pi\mu_t x^{(2)} + \nu_i x_j^{(2)}\right)}{1 + \sum_{j'} \exp\left(\delta_{j't} + \pi e_{it}\sigma_t x_{j'}^{(2)} + \pi\mu_t x_{j'}^{(2)} + \nu_i x_{j'}^{(2)}\right)} dF(e_{it})\phi(\nu_i; \lambda) \\
&= \sigma_j(\delta_j, x^{(2)}, \underbrace{\sigma_t x^{(2)}, \mu_t x^{(2)}}_{\text{new characteristics}} |\pi, \sigma_x) \\
&= \mathcal{D}(\omega_t, d^{(2)}, \sigma_t d^{(2)}, \mu_t d^{(2)}|\theta)\text{: Symmetric function!}
\end{aligned}
$$

- The reduced-form of this transformed model can therefore be written:

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

Suggestion 4: 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.
  - Differentiation IVs should be interacted with moments of the distribution of consumer characteristics to separately identify the two sources of heterogeneity.
- **Example:** Miravete, Seim, and Thurk (2017)
  - Combine nested-logit 'type' instruments, with moments of the distribution of demographics across stores.

Monte-Carlo Simulations

1. Independent random coefficients
2. Correlated random coefficients
3. Endogenous prices
4. Natural experiments
5. Optimal IV approximation: Comparison with Berry et al. (1999) and Reynaert and Verboven (2013).

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

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
σ_1	0.00	0.12	0.00	0.13	-0.00	0.13	-0.00	0.14
σ_2			-0.00	0.13	0.00	0.13	-0.00	0.14
σ_3					0.00	0.13	-0.00	0.14
σ_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:

$$\boldsymbol{\beta}_i^{(2)} \sim N(\boldsymbol{\beta}^{(2)}, \boldsymbol{\lambda})$$

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

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

Experiment 2: Correlated Random Coefficients

- Consumer heterogeneity:

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

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

100 markets $\times$ 50 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

		$\Sigma_{\cdot,1}$	$\Sigma_{\cdot,2}$	$\Sigma_{\cdot,3}$	$\Sigma_{\cdot,4}$
Bias	$\Sigma_{1,\cdot}$	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
RMSE	$\Sigma_{1,\cdot}$	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			
Rank test		474.053			
Nb endo.		10.000			
Nb IVs		15.000			

$\Sigma_{1,\cdot}$	4			
$\Sigma_{2,\cdot}$	-2	4		
$\Sigma_{3,\cdot}$	2	-2	4	
$\Sigma_{4,\cdot}$	2	-2	2	4

- **Note:** The vector of non-linear parameters correspond to the lower-diagonal elements of the choleski matrix of Σ (10).

How to account for endogenous characteristics?

- Two cases:
  - *Linear characteristics:* Replace x_{jt} with instrument w_{jt} when defining moment conditions (standard solution).
  - *Non-linear characteristics:* More difficult problem...
- Two approaches:
  1. Heuristic approximation to optimal IVs similar to BLP-1995
  2. Natural experiment-type variation (i.e. fixed-effects)

Example 1: Instruments for non-linear attributes

- Payoff function: Quality ladder

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

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

Example 1: Instruments for non-linear attributes

- Payoff function: Quality ladder

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

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

- BLP (1995): Prices and ξ_{jt} are simultaneously determined

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

where $\boldsymbol{w}_t = \{w_{jt}\}_{j=1,\dots,J_t}$ is a vector of excluded price instruments.

Example 1: Instruments for non-linear attributes

- Payoff function: Quality ladder

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

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

- BLP (1995): Prices and ξ_{jt} are simultaneously determined

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

where $\boldsymbol{w}_t = \{w_{jt}\}_{j=1,\dots,J_t}$ is a vector of excluded price instruments.

- **Curse of dimensionality:** Except in 'very' special cases (e.g. single-product Bertrand), the conditional distribution of prices is not a symmetric function of $(\boldsymbol{x}_t, \boldsymbol{w}_t)$.

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

How to account for heterogenous price coefficient?

- **Heuristic solution:** Distribute the expectation for price *inside* of the inverse-demand function (Berry et al. 1999):

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

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

- $\hat{p}_{jt} = E(p_{jt}|\boldsymbol{w}_{kt})$ is the 'first-stage' predicted price.

Experiment 3: Differentiation IVs with Endogenous Prices

Example with cost shifter

1. Exogenous price index (OLS):

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

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

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

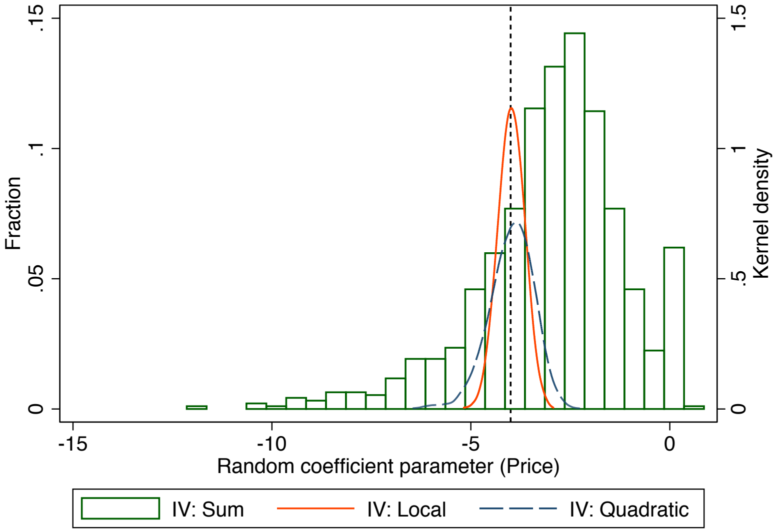
$$\sum_{j'} \left(d^{\hat{p}}_{jt,j'} \right)^2 \text{ and } \sum_{j'} \left(d^{\hat{p}}_{jt,j'} \right)^2 \cdot \boldsymbol{d}_{jt,j'}$$

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

	(1) True	(2) Diff. IV = Local			(3) Diff. IV = Quadratic			(4) Diff. IV = Sum		
		bias	se	rmse	bias	se	rmse	bias	se	rmse
λ_p	-4.00	0.02	0.27	0.28	0.02	0.53	0.55	1.03	158.25	2.10
β_p	-0.20	0.01	0.37	0.37	0.01	0.31	0.32	-0.67	201.29	1.38
β_0	50.00	-0.26	3.92	3.92	-0.28	7.36	7.45	-9.82	26.41	20.65
β_x	2.00	-0.02	0.46	0.45	-0.02	0.47	0.47	0.34	1.11	0.83

GMM estimates with endogenous prices

	(1) IV = Local	(2) IV=Quadratic	(3) 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

- The *Conditional 1st-stage F-test* statistic is the Weak IV test proposed by Angrist and Pischke for multiple endogenous variables.
- The IIA test is testing the exclusion restriction, $H_0 : \gamma = 0$ from the following linear IV regression:

$$\ln s_{jt}/s_{0t} = x_{jt}\beta + \alpha p_{jt} + \gamma IV_{jt}^{diff} + u_{jt}$$

where (β, α, γ) are estimated by GMM using the cost-shifter (ω_{jt}) as excluded instrument.

Example 2: Natural Experiments

- An alternative solution is to exploit natural experiments that vary the choice-set over time or across markets.
- Example: Three-way panel, product j, market m, and time $(t = 0, 1)$.

$$\text{Simultaneity problem: } E[\xi_{jmt} | \mathbf{x_{mt}}] \neq \mathbf{0}$$

$$\text{Decomposition: } \xi_{jmt} = \mu_{jm} + \tau_t + \Delta\xi_{mt}$$

Example 2: Natural Experiments

- An alternative solution is to exploit natural experiments that vary the choice-set over time or across markets.
- Example: Three-way panel, product j, market m, and time $(t = 0, 1)$.

$$\text{Simultaneity problem: } E[\xi_{jmt} | \mathbf{x_{mt}}] \neq \mathbf{0}$$

$$\text{Decomposition: } \xi_{jmt} = \mu_{jm} + \tau_t + \Delta\xi_{mt}$$

- **Assumption:** Quasi-experimental design

$$E[\Delta\xi_{jmt} | \bar{\xi}_m, \tau_t, \mathbf{x_{mt}}] = \mathbf{0}$$

Experiment 4: Random Entry in Hotelling

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

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

Experiment 4: Random Entry in Hotelling

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

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

- Treatment variable:

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

Experiment 4: Random Entry in Hotelling

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

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

- Treatment variable:

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

- Reduced-form: Difference-in-difference regression

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

Experiment 4: Random Entry in Hotelling

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

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

- Treatment variable:

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

- Reduced-form: Difference-in-difference regression

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

- **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}.

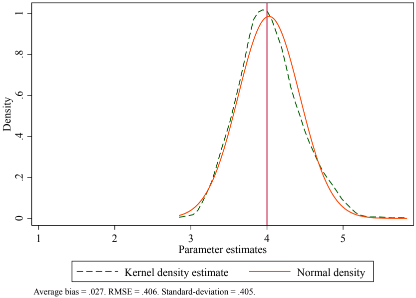
Natural Experiment: Hotelling Example

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

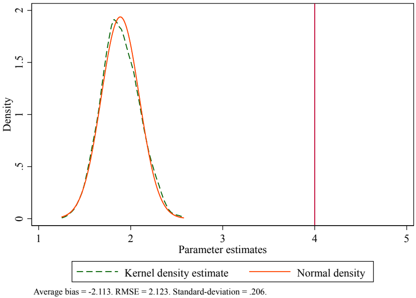
Natural Experiment: Hotelling Example

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

Difference-in-Difference Moments



Differentiation IVs w/o FEs



- "Diff-in-Diff" specification:

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

Optimal IV Approximation

- Abstracting from heteroscedasticity concerns, the "Optimal IV" takes the following form:

$$A_j^*(\boldsymbol{x}_t) = E\left[\frac{\partial \rho_j(\boldsymbol{s}_t, \boldsymbol{x}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big| \boldsymbol{x}_t\right] = \left\{-\boldsymbol{x}_{jt}, E\left[\frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \Big| \boldsymbol{x}_t\right]\right\}$$

- Instead of using non-parametric regressions to approximate $A_j^*(\boldsymbol{x}_t)$, Berry et al. (1999) propose the following heuristic:

$$\tilde{A}_j(\boldsymbol{x}_t|\boldsymbol{\theta}) = \frac{\partial \sigma_j^{-1}(\boldsymbol{s}_t, \boldsymbol{p}_t, \boldsymbol{x}_t^{(2)}; \boldsymbol{\theta})}{\partial \boldsymbol{\lambda}}\Bigg|_{p_{jt}=\hat{p}_{jt}, \xi_{jt}=0, \forall j,t}$$

  where $\hat{p}_{jt} \approx E(p_{jt}|\boldsymbol{x}_t, \boldsymbol{w}_t)$ is a "reduced-form" model for prices independent of ξ_{jt}.

- This leads to a two-step estimator:
  - Obtain initial estimate $\hat{\theta}^1$ using instrument vector $A_j(\boldsymbol{x}_t)$
  - Compute $\tilde{A}_j\left(\boldsymbol{x}_t|\hat{\boldsymbol{\theta}}^1\right)$, and re-estimate the model (just-identified).

Optimal IV Approximation

- Reynaert and Verboven (2013) show that this procedure improves substantially the weak IV problems.
- **Alternative approach:** Exploit the property that the optimal IV is a symmetric function of the vector of characteristics differences.
  - Use **Differentiation IVs** to obtain θ^1
  - Approximate the optimal IV directly by projecting the Jacobian on $A_j(\boldsymbol{x}_t)$ (Newey 1990)
- **Questions:**
  - How important is it to use consistent first-stage estimates to construct a valid Optimal IV approximation?
  - What is the efficiency gain of using optimal IV heuristic, relative to using differentiation IVs directly?

Optimal IV approximation with alternative initial parameter values

	Normal RC			Hotelling		
	λ^1	bias	rmse	λ^1	bias	rmse
Optimal IV approx.:						
(1)	0.5	0.001	0.027	4	-0.003	0.140
(2)	1.5	0.001	0.026	**2**	**-0.004**	**0.126**
(3)	2	**0.001**	**0.026**	0	-0.079	0.509
(3)	2.5	0.001	0.026	-1	-0.344	1.687
(4)	3	0.002	0.028	-2	-0.282	1.254
Differentiation IV	—	0.001	0.031	—	0.017	0.310

- **Takeaway 1:** With IID RC, inconsistent first-stage does not lead to biased or noisy estimates. The optimal IV approximation is "strong" for all λ^1!
- **Takeaway 2:** With the hotelling model, inconsistent first-stage leads to biased estimates and weak instruments.
- Why? The magnitude of λ does not determine "who competes with who". Only the magnitude of diversions.

Example 2: Correlated Random Coefficients

Choleski matrix	True (1)	Opt. IV: $\boldsymbol{\theta}^1 \sim N(0,1)$ bias (2)	rmse (3)	se (4)	Opt. IV: $\boldsymbol{\theta}^1 \sim N(0,4)$ bias (4)	rmse (5)	se (6)	Diff. IV: Quad. bias (7)	rmse (8)	se (9)
$\log c_{11}$	0.69	0.00	0.22	5.42	0.01	1.22	11.92	-0.00	0.03	0.03
$\log c_{22}$	0.55	-0.01	0.19	2.50	-0.16	2.36	192.70	-0.00	0.04	0.04
$\log c_{33}$	0.49	-0.02	0.15	0.46	-0.44	2.69	++	-0.00	0.04	0.04
$\log c_{44}$	0.46	-0.22	1.83	++	-1.78	5.57	++	-0.00	0.04	0.04
c_{21}	-1.00	0.01	0.47	4.51	0.03	0.77	781.85	0.00	0.06	0.06
c_{31}	1.00	0.00	0.33	0.86	-0.02	0.63	23.48	-0.00	0.07	0.07
c_{32}	-0.58	0.02	0.27	2.69	0.03	0.56	285.80	0.00	0.07	0.08
c_{41}	1.00	0.00	0.23	1.37	0.00	0.58	333.93	0.00	0.07	0.07
c_{42}	-0.58	0.01	0.23	2.69	0.04	0.50	484.88	0.00	0.08	0.08
c_{43}	0.41	0.00	0.23	1.59	0.03	0.52	++	0.00	0.08	0.08

Example 3: Efficiency gains

Quality ladder model

		True	Diff. IV = Local bias	Diff. IV = Local se	Diff. IV = Local rmse	Diff. IV = Quadratic bias	Diff. IV = Quadratic se	Diff. IV = Quadratic rmse	Diff. IV = Sum bias	Diff. IV = Sum se	Diff. IV = Sum rmse
1st-stage	λ_p	-4	0.02	0.27	0.28	0.02	0.53	0.55	1.01	2.66	2.09
	β_0	50	-0.26	3.92	3.92	-0.28	7.36	7.45	-9.63	26.48	20.46
	β_x	2	-0.02	0.46	0.45	-0.02	0.47	0.47	0.34	1.11	0.83
	β_p	-0.2	0.01	0.37	0.37	0.01	0.31	0.32	-0.66	1.76	1.37
2nd-stage	λ_p	-4	0.00	0.24	0.23	0.00	0.24	0.23	0.01	0.26	0.31
	β_0	50	-0.07	3.99	3.84	-0.06	3.72	3.65	0.05	4.32	4.61
	β_x	2	-0.01	0.48	0.47	-0.01	0.41	0.41	0.03	0.52	0.51
	β_p	-0.2	0.01	0.36	0.36	0.00	0.31	0.32	-0.03	0.40	0.40

Conclusion

- What did we do:
  - Show how that the characteristic model can be used to construct **relevant** instruments to identify substitution patterns
  - And, eliminate the weak IV problem that is present in applied work
  - *Differentiation IV's*: Capture the relative position of each product in the characteristic space.

Conclusion

- What did we do:
  - Show how that the characteristic model can be used to construct **relevant** instruments to identify substitution patterns
  - And, eliminate the weak IV problem that is present in applied work
  - *Differentiation IV's*: Capture the relative position of each product in the characteristic space.
- Extensions:
  - Optimal IV approximation (Reynaert and Verboven (2013))
  - Natural experiments
  - Demographic variation
  - Weak IV tests

Conclusion

- What did we do:
  - Show how that the characteristic model can be used to construct **relevant** instruments to identify substitution patterns
  - And, eliminate the weak IV problem that is present in applied work
  - *Differentiation IV's*: Capture the relative position of each product in the characteristic space.
- Extensions:
  - Optimal IV approximation (Reynaert and Verboven (2013))
  - Natural experiments
  - Demographic variation
  - Weak IV tests
- What's next?
  - Higher-order basis: Lasso
  - Conduct tests
  - Non-parametric estimation

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.

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.

Chamberlain, G. (1987).
Asymptotic efficiency in estimation with conditional moment restrictions.
Journal of Econometrics 34(305—334).

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

Newey, W. K. (1990).
Efficient instrumental variables estimation of nonlinear models.
Econometrica 58(809-837).

Reynaert, M. and F. Verboven (2013).
Improving the performance of random coefficients demand models: The role of optimal instruments.
Journal of Econometrics 179(1), 83–98.