# Spatial Price Competition: A Semiparametric Approach

Caio Figueiredo

Penn State

2019

# Introduction

- Objective: develop an empirical technique that can be used to discriminate between local and global competition
- Immediate Application: Wholesale Gasoline Market

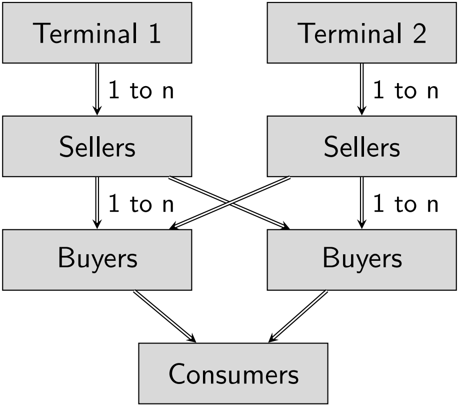
# Definitions

- Local competition denotes a situation in which firms compete directly only with a subset of **closest** firms.
- In contrast, global competition denotes the scenario where a firm price is directly correlated to everyone else's price.
- Note that the notion of **closeness** is not yet defined and, in fact, it's definition is not always straightforward.

# Relevance

- Competition characteristics such as local/global are an important factor for defining a good policy.

# The Market


Terminal 1
Terminal 2
1 to n
1 to n
Sellers
Sellers
1 to n
1 to n
Buyers
Buyers
Consumers

# The Market

- Gasoline is piped by multiple sellers to terminal locations.
- Upstream Market: Sellers offer branded gas to long term binded buyers and unbranded gas to independent buyers
- Downstream Market: Buyers offer gasoline to consumers.
  - We assume that the Downstream market is competitive, therefore the Downstream market price (SPOT price) is taken as given.

# The Market

- We are interest in the inter-terminal unbranded gasoline market
  - **We want to show that this market is locally competitive**
- Since branded gasoline buyers are binded to sellers there is no inter-terminal branded market

# The model

- We have $n$ Sellers:
  - $q = (q_1, q_2, ..., q_n)$ denotes the quantity sold.
  - $y = (y_1, y_2, ..., y_n)$ denotes product characteristics.
  - $\tilde{p} = (\tilde{p}_1, \tilde{p}_2, ..., \tilde{p}_n)$ denotes sellers price.
- and $K$ Buyers:
  - $\tilde{v} = (\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_n)$ denotes buyers reselling price.
  - $\tilde{\Pi}(\tilde{v}, \tilde{p}, \tilde{y}) = \tilde{\pi}_k(v_k, \tilde{p}, \tilde{y})$ denotes buyers profit function.

# The model

- By defining:
  - $V$ a price index.
  - $p = V^{-1}\tilde{p}$.
  - $\bar{v} = V^{-1}\tilde{v}$.
- We can approximate the profit function by:

$$\tilde{\Pi}(\tilde{v}, \tilde{p}, y) \approx V\left\{\tilde{\alpha}_1^T p + \tilde{\alpha}_2 + \frac{V}{2}\left[p^T B^1 p + \bar{v} B^2 \bar{v} + p^T B^3 \bar{v}\right] + \frac{1}{2}\left[p^T B^4 y + \bar{v} B^5 y\right]\right\}.$$

- But since $\bar{v}$ is constant, because the downstream market is competitive, we can rewrite the above as:

$$\Pi(p, y) = a_0 + a^T p + \bar{a}^T y + \frac{1}{2}\left[p^T B^1 p + p^T B^4 y\right]$$

# The model

- We can use Hotelling's Lemma to derive the demand function as:

$$q_i = \frac{\partial \tilde{\Pi}}{\partial \tilde{p}_i} \approx \frac{\partial \Pi}{\partial p_i}\frac{\partial p_i}{\partial \tilde{p}_i} = \partial \frac{\Pi}{p_i} = a_i + \sum_{j=1} [b_{ij}^1 p_j + b_{ij}^4 y_j]$$

- Given the demand function we can derive the maximization problem of the upstream firms:

$$\max_{p_i} (p_i - \gamma^T c_i) \left[ a_i + \sum_{j=1} [b_{ij}^1 p_j + b_{ij}^4 y_j] \right] - F_i$$

- Where $\gamma^T c_i$ is the total variable cost and $F_i$ is the fixed cost.

# The model

- Solving the sellers maximization problem we have:

$$p_i = \frac{1}{-2b_{ii}^1}\left(a_i - b_{ii}^1 \gamma^T c_i + \sum_{j \neq i} b_{ij}^1 p_j + \sum_{j=1} b_{ij}^4 y_j\right) \ \forall i \in 1, \ldots, n$$

# The model

- Now we can aggregate costs factors and product characteristics to have the follow econometrical model:

$$p = A + X\beta + Gp + u$$

- Where:
  - $A$ is a vector of intercepts that is treated as random effects.
  - $X$ is a matrix of control variables
  - $\beta$ is a vector of parameters.
  - $G = (g_{ij})$ is a matrix such that: $g_{ii} = 0 \ \forall i$ and $g_{ij} = g(d_{ij}) \ \forall i \neq j$.
- $g$ **will be estimated using a Semiparametric approach**

# Conventional Method

- If one is willing to impose considerable structure previous equation, it is possible to estimate it by conventional methods.
- Particularly, if we assume that $u_i \sim$i.i.d.$N(0, \sigma^2)$ and that $G = \psi \mathcal{W}$, where $\psi$ is a parameter (it's a scalar) and $\mathcal{W}$ is a weighting matrix (of distances), then the parameters $(\sigma^2, \beta, \psi)$ can be estimated by standard OLS.
- Which is a common approach in the literature

# Semiparametric method

- $g(d)$ can be written as:

$$g(d) = \sum_{l=1}^{\infty} \alpha_l e_l(d)$$

- where $e_l(d)$ is the basis of the functional space (we usually have: $e_l = d^l$)
- and therefore:

$$p_i = a_i + \sum_{l=1}^{\infty} \alpha_l \sum_{j \neq i} e_l(d_{ij}) p_j + \beta^T x_i + u_i$$

# Semiparametric method

- By truncating the number of expansions terms to be estimated, we have:

$$p_i = a_i + sum_{l=1}^{L_n} \alpha_l \sum_{j \neq i} e_l(d_{ij}) p_j + \beta^T x_i + v_i$$

- Where $v_i = u_i + \sum_{l=L_n+1}^{\infty} \alpha_l \sum_{j \neq i} e_l(d_{ij}) p_j$
- Rewriting in vector notation:

$$p = Z\alpha + X\beta + v$$

- Where $Z$ is a whose $(l, i)$ element is $\sum_{j \neq i} e_l(d_{ij}) p_j$.
- **Which can be estimated by OLS**.

# Application visualization

- In the Wholesale Gasoline Market application, $L_n$ is set to 5, using $e_l(d) = d^l$, we have:

$$p_i = a_i + \sum_{j \neq i} (\alpha_1 d_{ij} + \alpha_2 d_{ij}^2 + \alpha_3 d_{ij}^3 + \alpha_4 d_{ij}^4 + \alpha_5 d_{ij}^5) p_j + \beta^T x_i + v_i$$

# Problem

- Rival price ($p_j$) is not independent of $u_i$, let alone $v_i$
- **This endogeneity problem is dealt by the use of Instrumental Variables**
- If $H$ is the size o $X$, we must have at least $H + L_n$ instruments.

# Estimator

- By defining $W$ as the concatenation of $X$ and $Z$ and $\theta = [\alpha^T.\beta^T]^T$, we are able to use a standard IV estimator:

$$\hat{\theta} = (W^T P_B W)^-1 W^T P_B p$$

- Where $P_B$ is the orthogonal projection matrix of the matrix of instruments.
- and hence:

$$\hat{g}(d) = \sum_{l=1}^{L_n} \hat{\alpha}_l e_l(d)$$

# Theorems

- Theorem 1 (Consistency): $\hat{g}(d) - g(d) = o_p(1)$ for *almost all* $d$ and $\hat{\beta} - \beta = o_p(1)$
- Theorem 2 (Asymptotic Normality of $\hat{\beta}$):

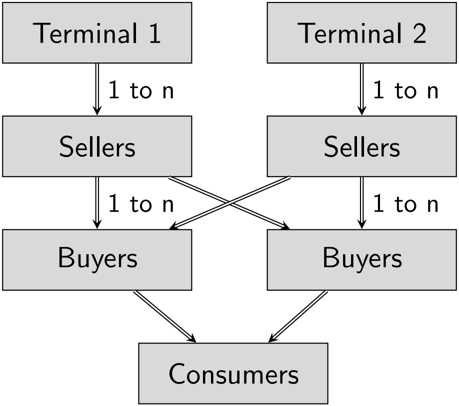
$$(X^T P_B M_{P_B Z} P_B \Omega M_{P_B Z} P_B X)^{1/2} X^T P_B M_{P_B Z} P_B X(\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(0, I)$$

- Theorem 3 (Asymptotic Normality of $\hat{g}$):

$$\hat{\Omega}_g^{-1/2} \{\hat{g}(d) - g(d)\} \xrightarrow{\mathcal{L}} N(0, 1)$$

- The theorems are proved in the papers but the proofs are going to be skipped in this presentation.

# Remmember our application



# Data

- Our data is a cross section of 305 terminals in the lower 48 states.
- Our main variable of interest is the (lowest) unbranded gasoline price per terminal, which is collect by Oil Price Information Service (OPIS):
- *PRICE* - regular unleaded price for the 3rd week in October of 1993

# Data, components of X:

- *SPOT* - gasoline spot prices to capture overall economic conditions in the oil industry
- *STOCK* - the percentage change in stocks to measure of supply/demand imbalance
- *POP* - city population for local demand variables
- *INC* - city household income for local demand variables
- *WAGE* - city wage rates to measure local labor costs
- *MTBE* - indicator for regions that require that gasoline burned in the region contain methyl terciary butyl, which increases production costs
- *NCOMP* - the number of competing sellers at a terminal to capture variations in local market structure
- *BRPRICE* - average branded price for each terminal
- *NBRAND* - the number of branded sellers at that terminal
- *PAD* - dummy variables that distinguish the five petroleum allocation districts

# Data, distances measures:

1. **NNX** - dummy variables that equal one if outlet $j$ is $i$'s nearest neighbor and zero otherwise, where $i$'s nearest neighbor is located in the terminal that is the shortest Euclidean distance from $i$

   **NNP** - the nearest neighbor determined endogenously by prices and transport costs. That is, Terminal B is the nearest neighbor of Terminal A, if B is chepeast alternative to A.
2. **CBX** - dummy variables that equal one if $i$ and $j$ share an exogenous market boundary but are not nearest neighbors, and zero otherwise

   **CBP** - similar to $CBX$ except that market boundaries are endogenously determined

# CBP Illustration

J. PINKSE, M. E. SLADE, AND C. BRETT

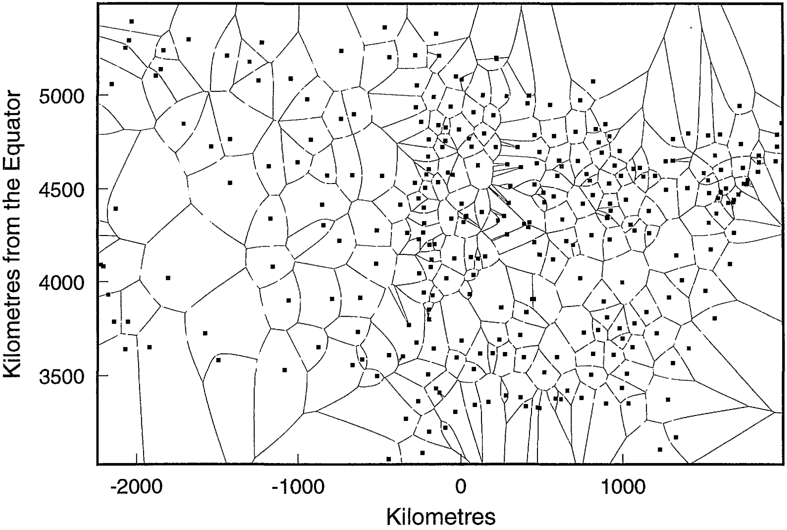



FIGURE 1.—Endogenous market areas.

# Data, distances measures:

3. **CBX2** - dummy variables that equal one if $i$ and $j$ do not share a market boundary in the sense of *CBX*, but each shares a boundary with a third seller
   **CBP2** - similar to *CBX2* except that markets are based on delivery prices instead of Euclidean distances.
4. **EDX** - global measures of closeness, which is the function of the Euclidean distance between locations $i$ and $j$, $\frac{1}{(0.01 \times XX_{ij}+1)}$ where $XX_{ij} = EU_{ij}$
   **EDP** - similar to *EDP* except that $XX_{ij} = EU_{ij} + \frac{(p_j - p_i)}{\tau}$, which means functions are of delivered prices

# OLS results

TABLE II

OLS Price Response Functions

Endogenous Market Boundaries

| Rival-Price Variable | Nearest-Neighbor 1 | Common-Boundary 2 | Second-Order Common-Boundary 3 | Euclidean Distance 4 | Branded Price at Same Terminal 5 | All Rival Prices 6 |
|---|---|---|---|---|---|---|
| *NNPPRICE* | 0.793 | | | | | 0.394 |
| | (0.038) | | | | | (0.042) |
| *CBPPRICE* | | 0.146 | | | | 0.041 |
| | | (0.008) | | | | (0.010) |
| *CB2PRICE* | | | 0.026 | | | 0.006 |
| | | | (0.0025) | | | (0.004) |
| *EDPPRICE* | | | | 0.002 | | 0.0004 |
| | | | | (0.0014) | | (0.0006) |
| *BRPRICE* | | | | | 0.489 | 0.375 |
| | | | | | (0.032) | (0.034) |
| # of competitors | −0.084 | −0.016 | −0.035 | −0.089 | | 0.223 |
| | (0.027) | (0.032) | (0.039) | (0.042) | | (0.032) |
| # of CB competitors | | −8.888 | | | | −2.520 |
| | | (0.475) | | | | (0.813) |
| # of 2nd order competitors | | | −1.226 | | | −0.312 |
| | | | (0.111) | | | (0.260) |
| # of branded competitors | | | | | −0.636 | −0.557 |
| | | | | | (0.047) | (0.054) |
| Spot price | 0.309 | 0.402 | 0.736 | 0.900 | 0.604 | 0.234 |
| | (0.067) | (0.071) | (0.083) | (0.097) | (0.075) | (0.047) |
| | $R^2 = 0.84$ | $R^2 = 0.82$ | $R^2 = 0.73$ | $R^2 = 0.61$ | $R^2 = 0.78$ | $R^2 = 0.93$ |

*Notes:* Supply and demand variables and PAD fixed effects not shown. Standard errors in parentheses.

# Data, the instruments:

- **Both rival prices and distances can be endogenous.**
- We create instruments by multiplying terminal-specific exogenous variable (i.e. population, income, wage, etc..) by an exogeneous weighting matrix. To illustrate, when a specification includes the NNP, instruments are creating by multiplying NNP and POP, INC, etc...:

$$p_i = a_i + \beta x_i + \alpha g(NNP) \times p_j + \lambda NNX \times X + u$$

# IV results

TABLE III

IV PRICE RESPONSE FUNCTIONS

Endogenous Market Boundaries

| Rival-Price Variable | Nearest-Neighbor 1 | Common-Boundary 2 | Second-Order Common-Boundary 3 | Euclidean Distance 4 | Branded Price at Same Terminal 5 | All Rival Prices 6 |
|---|---|---|---|---|---|---|
| *NNPPRICE* | 0.921 (0.181) | | | | | 0.593 (0.145) |
| *CBPPRICE* | | 0.173 (0.017) | | | | −0.071 (0.089) |
| *CBP2PRICE* | | | 0.071 (0.009) | | | 0.031 (0.028) |
| *EDPPRICE* | | | | 0.003 (0.002) | | 0.002 (0.0012) |
| *BRPRICE* | | | | | 0.575 (0.155) | 0.483 (0.115) |
| # of competitors | −0.082 (0.027) | −0.075 (0.375) | −0.010 (0.050) | −0.085 (0.043) | | 0.269 (0.084) |
| # of CB competitors | | −10.435 (0.957) | | | | −5.299 (6.624) |
| # of 2nd order competitors | | | −3.088 (0.359) | | | −1.703 (1.548) |
| # of branded competitors | | | | | −0.752 (0.215) | −0.721 (0.176) |
| Spot price | 0.212 (0.151) | 0.309 (0.088) | 0.454 (0.130) | 0.896 (0.096) | 0.547 (0.127) | 0.243 (0.087) |
| | $R^2 = 0.84$ $J\ Stat = 2.2$ $df = 3$ | $R^2 = 0.82$ $J\ Stat = 0.5$ $df = 2$ | $R^2 = 0.61$ $J\ Stat = 0.02$ $df = 2$ | $R^2 = 0.43$ $J\ Stat = 10.0$ $df = 3$ | $R^2 = 0.78$ $J\ Stat = 0.1$ $df = 3$ | $R^2 = 0.89$ $J\ Stat = 1.2$ $df = 8$ |

# Finally, Semiparametric model

- $g(d)$ is approximated by a 5-order expasions
- The increase in the number of parameters demand an increase in the number of instruments, which is achieved by creating by multiplying the exogenous variable by $e_l(d_{ij})$

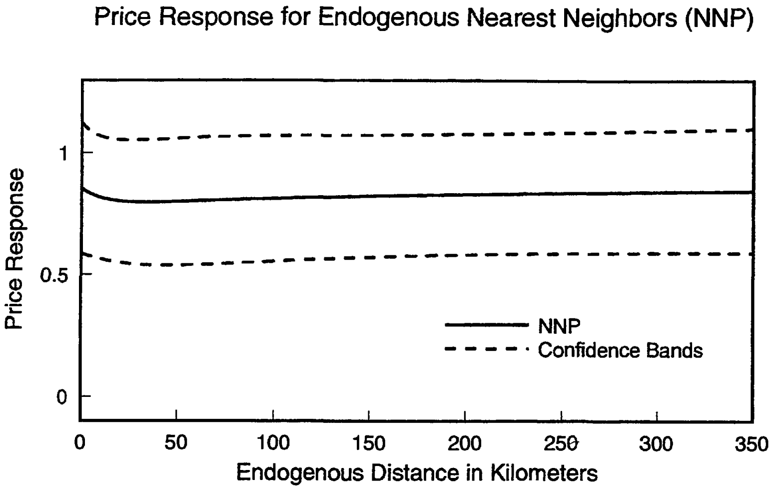
# Semiparametric model

- In all semiparametric model $g(d)$ is the result of a compound of the discrete measures NNP, CPB and CPB2 with the continuous measure EDX, such that:

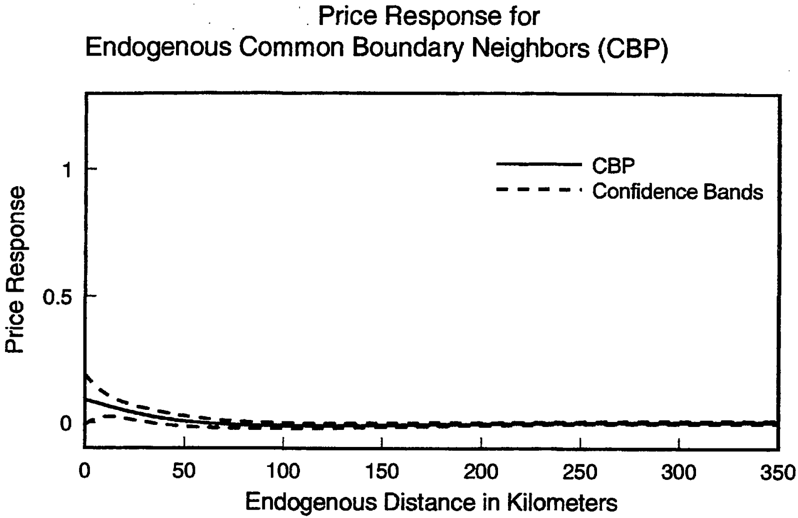
$$g(d) = \sum_{t=\{0,1\}} \mathbf{1}(d_D = t) g^t(d_C)$$

- We expect $g^0(d) = 0$.
- We are interest in $g^1(d)$.

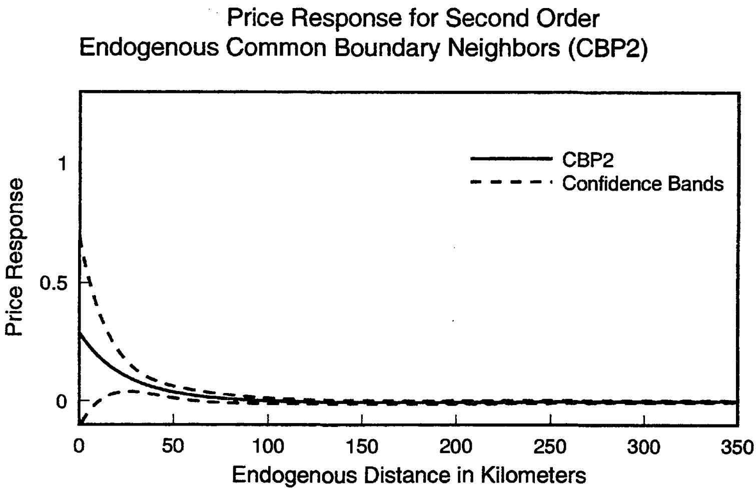
# Semiparametric results



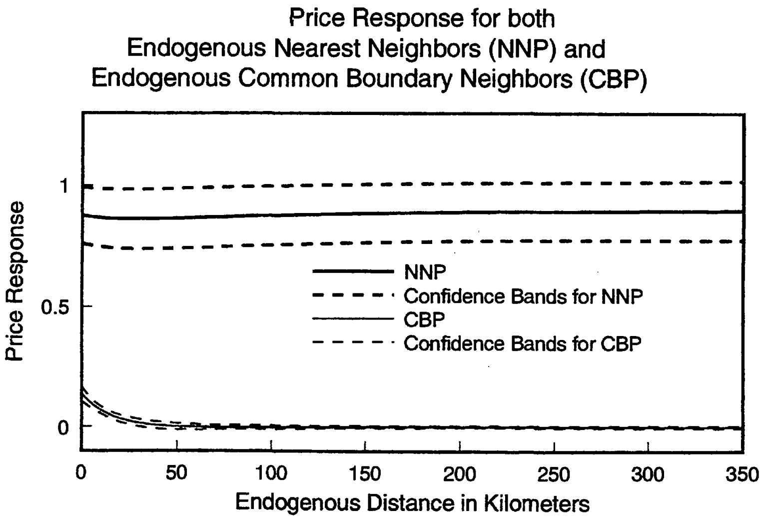
# Semiparametric results



# Semiparametric results



# Semiparametric results



# Semiparametric results

TABLE IV

OLS, IV, AND SEMIPARAMETRIC ESTIMATES
Price Response Functions with Endogenous Market Boundaries

| | OLS | IV | Semiparametric IV |
|---|---|---|---|
| Nearest neighbor price | 0.793 | 0.921 | (See Figure 2 |
| | (0.038) | (0.181) | |
| # of competitors | −0.084 | −0.082 | −0.095 |
| | (0.027) | (0.027) | (0.023) |
| Spot price | 0.309 | 0.212 | 0.235 |
| | (0.067) | (0.151) | (0.107) |
| % $\Delta$ stocks | −0.078 | −0.045 | −0.040 |
| | (0.101) | (0.112) | (0.067) |
| log(population) | −0.128 | −0.105 | −0.177 |
| | (0.079) | (0.087) | (0.048) |
| Average income | 0.092 | 0.085 | 0.063 |
| | (0.030) | (0.032) | (0.031) |
| Wage | −0.119 | −0.082 | −0.076 |
| | (0.052) | (0.074) | (0.059) |
| *MTBE* (gasoline additive) | 2.674 | 2.815 | 3.163 |
| | (1.008) | (1.046) | (1.861) |
| | $R^2 = 0.84$ | $R^2 = 0.84$ | |
| | | $J\ Stat = 2.2$ | $J\ Stat = 8.4$ |
| | | $df = 3$ | $df = 10$ |

# Conclusion

- The paper propose a method of estimating proximity that places minimal structure on the problem.
- Therefore multiple notion of distances can be tested and compared.
- Applying the method on the Wholesale Gasoline Market showed a market with highly localized competition.
- In fact, we seen that sellers compete only with the nearest terminal and ignore others even if it's a commomborder terminal.
- The same method could be applied in a number of other empirical research areas.