Price Dispersion and Market Segmentation: Evidence from the EU Bottled Water

Joris Hoste [1]

[1] University of Cambridge

IIOC 2024

May 4, 2024

Motivation

- Spatial price dispersion is commonplace, especially between countries. Scope for market integration?

Motivation

- Spatial price dispersion is commonplace, especially between countries. Scope for market integration?
- The interpretation and desirability of spatial price dispersion crucially **depends on its origins**:
  1. If driven by cost factors, price dispersion is considered *efficient*.
  2. If driven by trade frictions or price discrimination, the desirability of price dispersion *depends* on whether it facilitates entry or induces spatial misallocation.

Motivation

- Spatial price dispersion is commonplace, especially between countries. Scope for market integration?

- The interpretation and desirability of spatial price dispersion crucially **depends on its origins**:
  1. If driven by cost factors, price dispersion is considered *efficient*.
  2. If driven by trade frictions or price discrimination, the desirability of price dispersion *depends* on whether it facilitates entry or induces spatial misallocation.

- Identifying geographic market segmentation is **empirically challenging.** Need data on:
  - *Destination-specific marginal costs* to account for spatial price discrimination
  - *Global value chain* to separate trade frictions from differences in input prices and returns to scale

This paper - Approach

This paper overcomes these constraints by focusing on the **EU bottled water industry**. Three reasons:

This paper - Approach

This paper overcomes these constraints by focusing on the **EU bottled water industry**. Three reasons:

1. Empirical approach to get backed-out **destination-specific marginal costs**:
   - Household-level scanner data + Empirical model of demand and supply ⇒ *Marginal costs*
   - Household-level info on their residence ⇒ *Destination-specific*

This paper - Approach

This paper overcomes these constraints by focusing on the **EU bottled water industry**. Three reasons:

1. Empirical approach to get backed-out **destination-specific marginal costs**:
   - Household-level scanner data + Empirical model of demand and supply $\Rightarrow$ *Marginal costs*
   - Household-level info on their residence $\Rightarrow$ *Destination-specific*

2. Institutional context allows to deal with **global value chains**:
   - *Regulation*: Disclosure of source is mandatory by law
   - *Technology*: Bottled at the source, shipped to and distributed in the destination market

This paper - Approach

This paper overcomes these constraints by focusing on the **EU bottled water industry**. Three reasons:

1. Empirical approach to get backed-out **destination-specific marginal costs**:
   - Household-level scanner data + Empirical model of demand and supply ⇒ *Marginal costs*
   - Household-level info on their residence ⇒ *Destination-specific*
2. Institutional context allows to deal with **global value chains**:
   - *Regulation*: Disclosure of source is mandatory by law
   - *Technology*: Bottled at the source, shipped to and distributed in the destination market
3. EU-context is intriguing:
   - A *de jure* integrated market (e.g. Treaty of Rome, ...)
   - Given documented price dispersion, unclear whether it is also *de facto* integrated.

This paper - Findings

I discuss three sets of results:

Literature

This paper - Findings

I discuss three sets of results:

1. I start by providing two **reduced-form results**:
   - Spatial price dispersion is substantial
   - Border RDD shows that after-tax consumer prices are higher abroad

▸ Literature

This paper - Findings

I discuss three sets of results:

1. I start by providing two **reduced-form results**:
   - Spatial price dispersion is substantial
   - Border RDD shows that after-tax consumer prices are higher abroad

2. **Structural model** of demand and supply in the bottled water industry
   - Standard model with (1) discrete choice demand and (2) price setting along a vertical chain.
   - Explain spatial price dispersion through (1) price discriminiation, (2) destination-specific costs and (3) trade frictions
   - Model-implied trade friction of 0.09 EUR/L variable ($\sim$ 20% tariff-equivalent)

Literature

This paper - Findings

I discuss three sets of results:

1. I start by providing two **reduced-form results**:
   - Spatial price dispersion is substantial
   - Border RDD shows that after-tax consumer prices are higher abroad

2. **Structural model** of demand and supply in the bottled water industry
   - Standard model with (1) discrete choice demand and (2) price setting along a vertical chain.
   - Explain spatial price dispersion through (1) price discriminiation, (2) destination-specific costs and (3) trade frictions
   - Model-implied trade friction of 0.09 EUR/L variable ($\sim$ 20% tariff-equivalent)

3. A preliminary **counterfactual excercise** yields
   - Trade frictions increase cross-country price dispersion by 5%
   - Consumer surplus decreases by 0.036 EUR/L ($\approx$ 10% tax given average price of 0.37EUR/L)

Literature

Outline

- Dataset
- Motivational evidence
- Structural model + estimation
- Counterfactuals (Preliminary)
- Conclusion

DATASET

Data overview

I combine three data sources:

Data overview

I combine three data sources:

1. **Household-level scanner data** in non-alcoholic beverages from GfK and Kantar
    - SPATIAL: **EUR**: BEL, GER, FRA, NLD; **Non-EUR**: DEN, PLN, SWE, UK across 154 NUTS2-regions
    - TIME: Quarterly from 2010-2019
    - PRODUCTS: A combination of brand-source-flavored-bottle size bought in a given store

Data overview

I combine three data sources:

1. **Household-level scanner data** in non-alcoholic beverages from GfK and Kantar
   - SPATIAL: **EUR**: BEL, GER, FRA, NLD; **Non-EUR**: DEN, PLN, SWE, UK across 154 NUTS2-regions
   - TIME: Quarterly from 2010-2019
   - PRODUCTS: A combination of brand-source-flavored-bottle size bought in a given store

2. Based on Directive 2009/54/EC, I hand-collect 200 **production locations** of bottled water

Summary statistics | Hedonic regression | Coverage | Locations

Data overview

I combine three data sources:

1. **Household-level scanner data** in non-alcoholic beverages from GfK and Kantar
   - SPATIAL: **EUR**: BEL, GER, FRA, NLD; **Non-EUR**: DEN, PLN, SWE, UK across 154 NUTS2-regions
   - TIME: Quarterly from 2010-2019
   - PRODUCTS: A combination of brand-source-flavored-bottle size bought in a given store

2. Based on Directive 2009/54/EC, I hand-collect 200 **production locations** of bottled water

3. Other data sources:
   - ZIPcode level **travel distance and trucking times** from Localyse.eu and **diesel prices** from European Commission
   - Indirect **consumption taxes** (i.e. VAT, excise and packaging) from the European Commission
   - **Labor unit costs** in retail constructed from the EU-SILC database

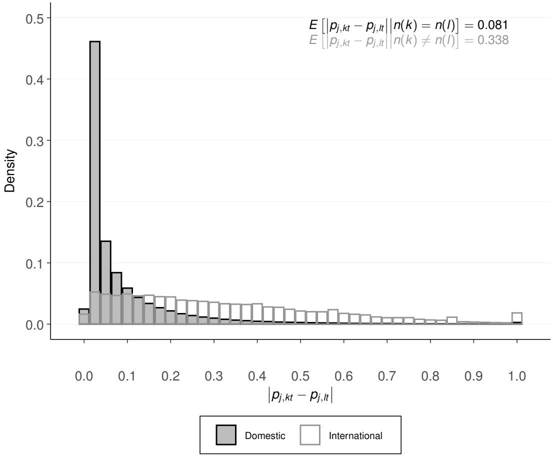
▸ Summary statistics ▸ Hedonic regression ▸ Coverage ▸ Locations

Motivational evidence

After-tax price differences are large

Figure 1: LOP deviations

(a) Between vs. within country LOP



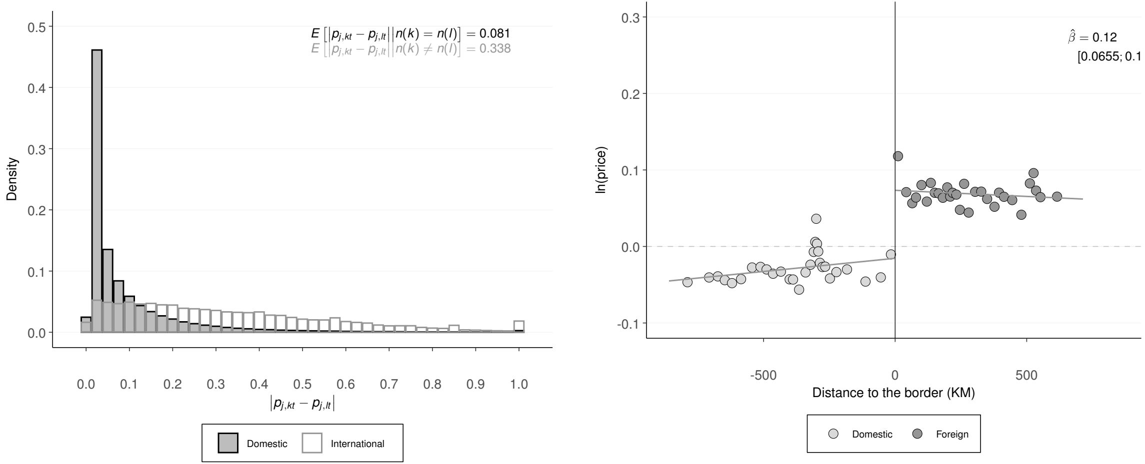
▸ Table ▸ Table - Cross-border shopping

After-tax price differences are large

Figure 1: LOP deviations

(a) Between vs. within country LOP

(b) Foreign vs. domestic prices (GER-FRA border)



▸ Table ▸ Table - Cross-border shopping

STRUCTURAL MODEL

Structural model - Overview

The structural model consists of three building blocks:

▸ Purchase frequency ▸ Cross-border shopping ▸ Demand estimates

Structural model - Overview

The structural model consists of three building blocks:

- Consumers make a **static discrete choice** about which water to buy. (Logit: $\alpha = -4.77$)

▸ Purchase frequency ▸ Cross-border shopping ▸ Demand estimates

Structural model - Overview

The structural model consists of three building blocks:

- Consumers make a **static discrete choice** about which water to buy. (Logit: $\alpha = -4.77$)

- Downstream retailers and upstream manufacturers interact vertically. Three setups:
  - SEQUENTIAL OLIGOPOLY: manufacturers and retailers set prices as oligopolists
  - DOWNSTREAM OLIGOPOLY: retailers set prices as oligopolists
  - UPSTREAM OLIGOPOLY: manufacturers set prices as oligopolists

Purchase frequency
Cross-border shopping
Demand estimates

Structural model - Overview

The structural model consists of three building blocks:

- Consumers make a **static discrete choice** about which water to buy. (Logit: $\alpha = -4.77$)
- Downstream retailers and upstream manufacturers interact vertically. Three setups:
  - SEQUENTIAL OLIGOPOLY: manufacturers and retailers set prices as oligopolists
  - DOWNSTREAM OLIGOPOLY: retailers set prices as oligopolists
  - UPSTREAM OLIGOPOLY: manufacturers set prices as oligopolists
- Beyond production costs, **destination-specific marginal costs** depend on three components:
  - Local distribution incurred by retailers
  - Transport costs
  - Cross-border trade frictions

Purchase frequency   Cross-border shopping   Demand estimates

Empirical framework - Estimating trade frictions

Compute destination-specific marginal costs from downstream and upstream of FOCs:

$$\underbrace{\boldsymbol{p}_{lt}^{r}}_{\text{Retail price}} = -\underbrace{\left(\boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}_{lt}^{r}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^{r};\boldsymbol{\Theta}^{d})}_{\text{Retail markup}} + \underbrace{\boldsymbol{p}_{lt}^{w} + \boldsymbol{c}_{lt}^{r}}_{\text{Marginal cost - retail}}$$

$$= -\underbrace{\left(\boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}_{lt}^{r}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^{r};\boldsymbol{\Theta}^{d})}_{\text{Retail markup}} - \underbrace{\left(\mathbf{PT}_{lt} \cdot \boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}_{lt}^{w}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^{r};\boldsymbol{\Theta}^{d})}_{\text{Manufacturer markup}} + \underbrace{\boldsymbol{c}_{lt}^{m} + \boldsymbol{c}_{lt}^{r}}_{\text{Marginal cost}}$$

$$\boldsymbol{c}_{lt}^{r} + \boldsymbol{c}_{lt}^{m} = \boldsymbol{p}_{lt}^{r} + \left(\boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}_{lt}^{r}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^{r};\boldsymbol{\Theta}^{d}) + \left(\mathbf{PT}_{lt} \cdot \boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}_{lt}^{w}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^{r};\boldsymbol{\Theta}^{d})$$

Empirical framework - Estimating trade frictions

Compute destination-specific marginal costs from downstream and upstream of FOCs:

$$\underbrace{\boldsymbol{p}^r_{lt}}_{\text{Retail price}} = -\underbrace{\left(\boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}^r_{lt}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^r; \boldsymbol{\Theta}^d)}_{\text{Retail markup}} + \underbrace{\boldsymbol{p}^w_{lt} + \boldsymbol{c}^r_{lt}}_{\text{Marginal cost - retail}}$$

$$= -\underbrace{\left(\boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}^r_{lt}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^r; \boldsymbol{\Theta}^d)}_{\text{Retail markup}} - \underbrace{\left(\mathbf{PT}_{lt} \cdot \boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}^w_{lt}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^r; \boldsymbol{\Theta}^d)}_{\text{Manufacturer markup}} + \underbrace{\boldsymbol{c}^m_{lt} + \boldsymbol{c}^r_{lt}}_{\text{Marginal cost}}$$

$$\boldsymbol{c}^r_{lt} + \boldsymbol{c}^m_{lt} = \boldsymbol{p}^r_{lt} + \left(\boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}^r_{lt}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^r; \boldsymbol{\Theta}^d) + \left(\mathbf{PT}_{lt} \cdot \boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}^w_{lt}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^r; \boldsymbol{\Theta}^d)$$

Now parametrize destination-specific marginal costs:

$$c^r_{j,lt} + c^m_{j,lt} = \underbrace{\gamma w^r_{l,t} + \lambda_{c(j),t}}_{c^r_{j,lt}} + \underbrace{\beta t_t^{s(j)l} + \beta_B \text{Border}^{s(j)l} + \beta_C \text{Cur}^{s(j)l} + \omega^m_{j,t}}_{c^m_{j,lt}} + \eta_{j,lt}$$

where $t_{s(j)l,t} = \text{Distance}_{s(j)l} \cdot \text{Diesel}_{s(j),t}$.

Downstream mkt. struct. Upstream mkt. struct.

Empirical framework - Trade costs

Table 1: Trade costs estimates

	(1)	(2)	(3)	(4)
Border$^{s(j),l}$	-	0.0901***	0.088***	0.0918***
	-	(0.007)	(0.007)	(0.007)
$t_t^{s(j),l}$	0.00901**	0.00307	0.002	0.00261
	(0.003)	(0.002)	(0.002)	(0.002)
w_{lt}	0.0156***	0.00818***	0.00828***	0.00828***
	(0.002)	(0.001)	(0.001)	(0.001)
$E\left[\cdot\vert\text{Export}_j = 1\right]$	0.33	0.38	0.38	0.33
τ_B	-	0.24	0.23	0.28
$\mu^r_{j,lt}$	✓		✓	✓
$\mu^w_{j,lt}$	✓	✓		✓
$\omega_{j,t}$	✓	✓	✓	✓
$\lambda_{c(j),t}$	✓	✓	✓	✓
R^2	0.92	0.93	0.93	0.93
N	645,227	645,227	645,227	645,227

Notes: Standard errors at the destination region. Significance at the 0.1*, 0.05**, 0.01*** levels.

COUNTERFACTUALS

Couterfactuals

Table 2: Counterfactual excercises

Counterfactual	τ	μ	$\mathbb{E}\left[\lvert p_{j,lt}-p_{j,kt}\rvert n(k)\neq n(j)\right]$ Level	Change	ΔCS
Integrated economy	0	0	34%	-	-
Segmented - No Market power	$\hat{\tau}$	0	39%	+ 5%	- 0.036 EUR/L
Segmented - No price discrimination	$\hat{\tau}$	$\mu_{j,t}(\hat{\tau})$		Loading ...	
Segmented - Price discrimination	$\hat{\tau}$	$\mu_{j,lt}(\hat{\tau})$		Loading ...	

Conclusion

Conclusion

- Large between-country price differences remain across European countries, casting doubt on market integration in the EU.
- This paper leverages features of the bottled water industry to make progress.
- Preliminary results point to a trade friction of roughly 0.09 EUR/L equivalent to a 10% tax on bottled water.

BACK-UP SLIDES

Related literature - ▸ Back

This paper connects with three strands of literature:

- **Price dispersion and trade frictions**: Goldberg & Verboven (2001), Asplund & Friberg (2001), Crucini et al. (2005), Shiue & Keller (2007), Gopinath et al. (2011), Cavallo et al. (2014), Atkin & Donaldson (2015), Donaldson (2018), Fontaine et al. (2020), Beck et al. (2020) and Chatterjee (2023)
  - *This paper* separate level of trade frictions from cost and price discrimination
- **Trade flows and trade frictions**: McCallum (1995), Anderson & Wincoop (2003), Coşar et al. (2015), Head & Mayer (2019), Santamaría et al. (2023)
  - *This paper* identifies trade frictions under weaker assumptions on market structure and technology
- **Trade and IO**: Goldberg (1995), Goldberg & Verboven (2001), Berry et al. (1999), Loecker (2011), Loecker et al. (2016) and Kalouptsidi (2018)
  - *This paper* estimates the level of trade frictions and the effect on spatial price dispersion and welfare

Data - Consumption data -

Back

Table 3: Sample overview

Variable	Overall	BEL	GER	DEN	FRA	NLD	PLN	SWE	UK
Regions	154	11	38	5	22	12	17	8	41
All sources	200	68	76	8	60	34	40	13	34
Local sources	-	8	60	5	50	3	33	4	23
Firms	127	32	41	8	22	23	42	12	20
Brands	267	59	94	23	53	40	55	20	49
Products	767	226	182	73	187	117	130	69	130
Stores	106	18	15	20	17	25	24	11	13
Households - All	704	352	730	361	561	754	404	299	984
Households - Water	436	286	539	125	470	399	349	93	445
Transactions ('1,000')	12,380	805	3,868	86	2,582	956	1,419	126	2,538
Share water - uncond.	0.23	0.25	0.29	0.11	0.32	0.11	0.36	0.09	0.12
Share water - cond.	0.34	0.32	0.40	0.30	0.39	0.21	0.42	0.30	0.27
Inside good share	0.64	0.79	0.72	0.34	0.83	0.51	0.86	0.31	0.44
Frequency of purchase	0.77	0.87	0.84	0.60	0.89	0.71	0.89	0.57	0.67
Unit price (incl.)	0.37	0.37	0.20	0.61	0.28	0.45	0.20	0.83	0.38
Unit price (excl.)	0.30	0.35	0.16	0.45	0.26	0.35	0.15	0.73	0.30

Notes: Unit price (incl.) is the average unit price inclusive of taxes in EUR and Unit price (excl.) is the average unit price exclusive of taxes in EUR.

Which product characteristics matter?

Back

Table 4: Hedonic price regression

$p_{j,lt}$	Raw sample (1)	Raw sample (2)	Raw sample (3)	Raw sample (4)	Cleaned sample (5)	Cleaned sample (6)	Cleaned sample (7)	Cleaned sample (8)
$\mathbb{1}$ (Sparkling)$_j$	−0.0429	−0.0182	−0.0216	−0.00811	0.00629	0.00714	0.0269	0.0125
	(0.027)	(0.015)	(0.027)	(0.015)	(0.033)	(0.019)	(0.035)	(0.020)
$\mathbb{1}$ (Flavored)$_j$	0.528***	0.484***	0.465***	0.39***	0.541***	0.508***	0.493***	0.401***
	(0.033)	(0.024)	(0.036)	(0.025)	(0.048)	(0.033)	(0.049)	(0.033)
$\mathbb{1}$ (Glass bottle)$_j$	−0.103*	−0.125***	−0.0945	−0.148***	−0.273*	−0.138	−0.229	−0.158
	(0.059)	(0.037)	(0.058)	(0.037)	(0.149)	(0.114)	(0.152)	(0.110)
$\mathbb{1}$ (Other package)$_j$	0.411***	0.138***	0.386***	0.071	0.323***	0.0994	0.342***	0.0265
	(0.053)	(0.049)	(0.066)	(0.053)	(0.105)	(0.071)	(0.119)	(0.091)
$\mathbb{1}$ ((750ml, 1500ml))$_j$	−0.249***	−0.369***	−0.239***	−0.356***	−0.231***	−0.334***	−0.243***	−0.348***
	(0.037)	(0.020)	(0.037)	(0.021)	(0.049)	(0.031)	(0.049)	(0.032)
$\mathbb{1}$ ($\geq$ 1500ml)$_j$	−0.832***	−0.778***	−0.847***	−0.794***	−0.735***	−0.735***	−0.767***	−0.769***
	(0.033)	(0.018)	(0.035)	(0.018)	(0.043)	(0.025)	(0.045)	(0.027)
$\mathbb{1}$ (Private label)$_j$	−0.53***		−0.507***		−0.595***		−0.559***	
	(0.028)		(0.029)		(0.037)		(0.037)	
$\mathbb{1}$ (Foreign)$_j$			0.287***	0.178***			0.261***	0.218***
			(0.043)	(0.052)			(0.051)	(0.062)
Region-Time FEs	✓	✓	✓	✓	✓	✓	✓	✓
Brand FEs		✓		✓		✓		✓
Adj.R^2	0.62	0.84	0.62	0.85	0.69	0.87	0.70	0.88
No. obs	917,894	920,722	742,693	742,693	535,497	536,247	439,772	439,772

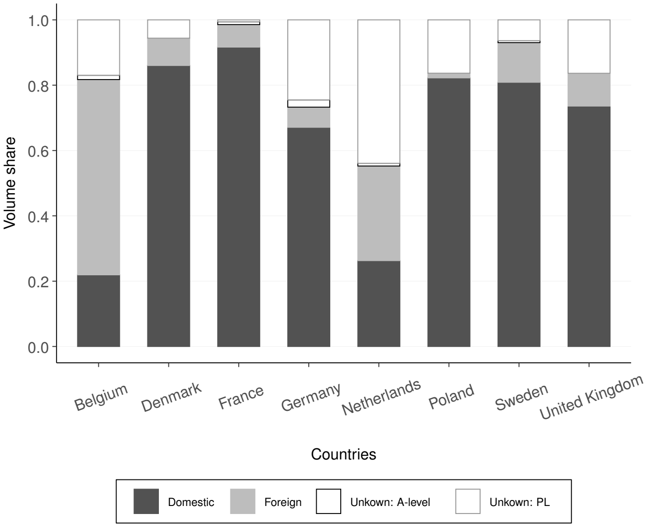
Notes: Standard errors at the variety level. Reported significance levels are at the $p < 0.1^{*}$, $p < 0.05^{**}$ and $p < 0.01^{***}$ levels.

Data - Production locations (≈ 200 sources) - ▸ Back

Figure 2: Production locations

Data - Production location accuracy - ▸ Back

Figure 3: Production locations - Accuracy



Foreign prices rise discontinuously at the border - Back

I estimate a border RDD as follows:

Foreign prices rise discontinuously at the border - ▸ Back

I estimate a border RDD as follows:

- I construct the **sample** as follows:
    1. Get the set of countries with a shared border
    2. Get the set of products that are produced in one country and sold in the other
    3. Rank observations in terms of their great circle distance to the border

Foreign prices rise discontinuously at the border - ▸ Back

I estimate a border RDD as follows:

- I construct the **sample** as follows:
  1. Get the set of countries with a shared border
  2. Get the set of products that are produced in one country and sold in the other
  3. Rank observations in terms of their great circle distance to the border

- I consider the following **RDD-estimator**:

$$\ln\left(p_{j,t}^{s(j)l}\right) = \beta \text{Border}^{s(j)l} + f^{n}\left(\text{Dis}^{s(j)l}; \gamma_{\mathbf{0}}\right) + f^{n}\left(\text{Dis}^{s(j)l}; \gamma_{\mathbf{1}}\right) + \lambda_{j,t} + \varepsilon_{j,t}^{s(j)l}$$

where $\text{Border}^{s(j)l} = \begin{cases} 0 & \text{if} \quad \text{domestic,} \\ 1 & \text{if} \quad \text{foreign.} \end{cases}$ and $\lambda_{j,t}$ are product-quarter fixed effects.

RDD estimate - Baseline

Back

Table 5: Border Regression Discontinuity Design: Results

	1^{st}-order				2^{th} order			
$p_{j,lt}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Border}^{s(j)l}$	0.167^{***}	0.172^{***}	0.107^{***}	0.101^{***}	0.179^{***}	0.155^{***}	0.0888^{***}	0.0927^{***}
	[0.13; 0.203]	[0.134; 0.21]	[0.0803; 0.134]	[0.0755; 0.127]	[0.14; 0.218]	[0.119; 0.191]	[0.0657; 0.112]	[0.0683; 0.117]
$\theta_{j,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Polynomial	1	1	1	1	2	2	2	2
Bandwidth	1,000	500	100	54.1	1,000	500	100	82.9
Optimal				✓				✓
No. obs	1,783,315	1,394,743	425,812	253,914	1,783,315	1,394,743	425,812	360,664

Notes: Standard errors are clustered at the product level reported and are robust to the fact that bandwidths that are far away from zero can lead to bad coverage of the confidence intervals (see Calonico et al. (2014)). I report the robust confidence intervals in square brackets and denote significance at the $p < 0.1^{*}$,$p < 0.05^{**}$ and $p < 0.01^{***}$ levels.

RDD estimate - Cross-border

Back

Table 6: Border Regression Discontinuity Design: Results

	1^{st}-order				2^{th} order			
$p_{j,lt}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\mathbb{1}(\text{Border}^{s(j)l} = 1)$	0.157***	0.154***	0.074***	0.0581***	0.158***	0.133***	0.0426**	0.0482**
	[0.124; 0.19]	[0.12; 0.189]	[0.0406; 0.107]	[0.0228; 0.0934]	[0.123; 0.193]	[0.0996; 0.166]	[0.00427; 0.0809]	[0.0111; 0.0852]
$\theta_{j,t}$	✓	✓	✓	✓	✓	✓	✓	✓
Polynomial	1	1	1	1	2	2	2	2
Bandwidth	1,000	500	100	45.2	1,000	500	100	76.9
Optimal				✓				✓
No. obs	2,302,791	1,834,173	536,661	272,617	2,302,791	1,834,173	536,661	430,165

Notes: Standard errors are clustered at the product level reported and are robust to the fact that bandwidths that are far away from zero can lead to bad coverage of the confidence intervals (see Calonico et al. (2014)). I report the robust confidence intervals in square brackets and denote significance at the $p < 0.1^{*}$,$p < 0.05^{**}$ and $p < 0.01^{***}$ levels.

Gravity: Tariff-equivalent trade friction - Setup

Back

Suppose l and k are origin and destination locations (NUTS2), and following Allen et al. (2020) assume

$$(1)\quad Q_t^{lk} = \left(\frac{P_t^{lk}}{P_{kt}}\right)^{-\sigma} \frac{Y_{kt}}{P_{kt}}, \qquad (2)\quad P_t^{lk} = P_t^{ll}\tau_t^{lk}, \qquad (3)\quad Q_{lt} = \sum_k Q_t^{lk}$$

Then, we write the trade flow, X_t^{lk}, as

$$X_t^{lk} = \left(\frac{\tau_t^{lk}}{\Pi_{lt} P_{kt}}\right)^{-\sigma} Q_{lt} Y_{kt}$$

Following Silva & Tenreyro (2006), we operationalize this as:

$$X_t^{lk} = \exp\left(\beta \ln(1 + \text{Dis}^{lk}) + \gamma_B \text{Border}^{lk} + \gamma_C \text{Cur}^{lk} + \lambda_{lt} + \lambda_{kt}\right) + \varepsilon_t^{lk}$$

where $\varepsilon_t^{lk} \equiv X_t^{lk} - \mathbb{E}\left[X_t^{lk} | \text{Dis}^{lk}, \text{Border}^{lk}, \text{Cur}^{lk}, \lambda_{lt}, \lambda_{kt}\right]$.

Gravity: Tariff-equivalent trade friction

Table 7: Gravity estimation - ▸ Gravity specification

X_t^{lk}	(1)	(2)	(3)
$\ln(1+\text{Dis}^{lk})$	-0.605^{***}	-0.519^{***}	-0.518^{***}
	(0.078)	(0.077)	(0.077)
$\text{Bor}^{lk}=1$		-1.2^{***}	-1.19^{***}
		(0.236)	(0.237)
$\text{Cur}^{lk}=1$			-3.09^{***}
			(1.178)
$e^{\hat{\beta}}-1$	-	-69.8%	-69.6%
$e^{\frac{\hat{\beta}}{\varepsilon_{EK}}}-1$	-	-13.5%	-13.4%
$e^{\frac{\hat{\beta}}{\varepsilon_{BLP}}}-1$	-	-43.2%	-43.0%
$e^{\hat{\gamma}}-1$	-	-	-95.4%
$e^{\frac{\hat{\gamma}}{\varepsilon_{EK}}}-1$	-	-	-31.1%
$e^{\frac{\hat{\gamma}}{\varepsilon_{BLP}}}-1$	-	-	-76.7%
$\lambda_{l,t}$	✓	✓	✓
$\lambda_{k,t}$	✓	✓	✓
No. obs	73,323	73,323	73,323

Notes: Estimated using PPML. $\varepsilon_{EK}=8.28$ taken from Eaton and Kortum (2002) and $\varepsilon_{BLP}=2.12$ taken from Boehm, Levchenko and Pandalai-Nayar (2023). Two-way clustered standard errors at origin and destination level. Significance at the 0.1^{*}, 0.05^{**}, 0.01^{***} levels.

Frequency of purchases

Back

Table 8: Frequency of purchase

Variable	Overall	BEL	GER	DEN	FRA	NLD	PLN	SWE	UK
Regions	154	11	38	5	22	12	17	8	41
All sources	200	68	76	8	60	34	40	13	34
Local sources	-	8	60	5	50	3	33	4	23
Firms	127	32	41	8	22	23	42	12	20
Brands	267	59	94	23	53	40	55	20	49
Products	767	226	182	73	187	117	130	69	130
Stores	106	18	15	20	17	25	24	11	13
Households - All	704	352	730	361	561	754	404	299	984
Households - Water	436	286	539	125	470	399	349	93	445
Transactions ('1,000')	12,380	805	3,868	86	2,582	956	1,419	126	2,538
Share water - uncond.	0.23	0.25	0.29	0.11	0.32	0.11	0.36	0.09	0.12
Share water - cond.	0.34	0.32	0.40	0.30	0.39	0.21	0.42	0.30	0.27
Inside good share	0.64	0.79	0.72	0.34	0.83	0.51	0.86	0.31	0.44
Frequency of purchase	0.77	0.87	0.84	0.60	0.89	0.71	0.89	0.57	0.67
Unit price (incl.)	0.37	0.37	0.20	0.61	0.28	0.45	0.20	0.83	0.38
Unit price (excl.)	0.30	0.35	0.16	0.45	0.26	0.35	0.15	0.73	0.30

Notes: Unit price (incl.) is the average unit price inclusive of taxes in EUR and Unit price (excl.) is the average unit price exclusive of taxes in EUR.

Cross-border shopping

Back

Table 9: Cross-border shopping

Country	All	BEL	GER	DEN	FRA	NLD	PLN	SWE	UK
Transactions (Count)									
·Domestic	15,027,795	859,605	5,120,630	90,889	3,012,993	1,005,842	1,965,480	134,482	2,837,874
·Cross-border	96,685	41,629	10,390	502	15,241	28,743	98	1	81
·Undisclosed	3,838	150	0	2,541	830	7	0	281	29
Transactions									
·Domestic	0.993	0.954	0.998	0.968	0.995	0.972	1.000	0.998	1.000
·Cross-border	0.006	0.046	0.002	0.005	0.005	0.028	0.000	0.000	0.000
·Undisclosed	0.000	0.000	0.000	0.027	0.000	0.000	0.000	0.002	0.000
Liters (liters)									
·Domestic	109,359,137	7,616,084	51,753,208	291,570	27,055,500	3,529,122	9,845,240	312,938	8,955,475
·Cross-border	1,246,252	587,835	187,959	5,249	299,837	164,323	609	0	440
·Undisclosed	18,344	721	0	9,893	7,253	4	0	419	54
Liters									
·Domestic	0.989	0.928	0.996	0.951	0.989	0.956	1.000	0.999	1.000
·Cross-border	0.011	0.072	0.004	0.017	0.011	0.044	0.000	0.000	0.000
·Undisclosed	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.001	0.000
Price (EUR/L)									
·Domestic	0.431	0.380	0.251	0.618	0.285	0.456	0.210	0.849	0.396
·Cross-border	0.298	0.306	0.293	0.665	0.191	0.301	0.158	0.000	0.470
·Undisclosed	0.683	0.636	0.000	0.646	0.333	2.000	0.000	1.182	0.664
Exp. share - NARTD									
·Unconditional	0.098	0.253	0.281	0.085	0.312	0.110	0.363	0.097	0.098

Structural model - Preferences

Back

Consumers $i = 1, \ldots, N_{lt}$ chooses among $j = 1, \ldots, \mathcal{J}_{lt} + 1$ products by solving:

$$\max_{j \in \mathcal{J}_{lt}+1} U_{ij,lt} = \alpha_{i,n(l)} P_{j,lt} + \boldsymbol{\beta}_{n(l)}{}' \boldsymbol{X}_{j,lt} + \xi_{j,lt} + \varepsilon_{ij,lt}$$

where $\varepsilon_{ij,lt} \sim EV(1)$ and

- $\alpha_{i,n(l)} = \alpha_{n(l)} + \alpha_{n(l),y} \mathbb{1}(y_i = y)$
  - $\alpha_{n(l)}$: average price sensitivity
  - $\alpha_{n(l),y}$: income-specific price sensitivity
- $\boldsymbol{X}_{j,lt}$ are product characteristics including:
  - Location-time FEs
  - brand-country FEs
  - retailer-country FEs

Structural model: Preferences

Back

Table 10: Preferences: $\ln\left(\frac{S_{j,lt}}{S_{0,lt}}\right) = \alpha P_{j,lt} + \theta_{b(j)} + \theta_{c(j),t} + \lambda_{lt} + \xi_{j,lt}$

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(P_{j,lt})$	-1.76^{***}	-1.75^{***}	-1.76^{***}	-22.4^{***}	-22.3^{***}	-20.2^{***}
	(0.184)	(0.185)	(0.184)	(1.640)	(1.190)	(1.310)
1^{st} Stage F-stat	-	-	-	4,287.6	5,633.6	6,040.6
$\mathbb{E}\left[\varepsilon_{j,lt}\right]$	-0.35	-0.35	-0.35	-4.47	-4.46	-4.03
$\theta_{b(j)}$	✓			✓		
$\theta_{b(j),n(l)}$		✓	✓		✓	✓
$\lambda_{c,t}$			✓			✓
λ_{lt}	✓	✓	✓	✓	✓	✓
N	952,970	952,970	952,970	786,735	786,735	786,735

Notes: Clustered standard errors at the location level. Significance at the 0.1^{*}, 0.05^{**}, 0.01^{***} levels.

Empirical framework - Downstream market structure

Back

In each market, there is a set of **retailers** $\mathcal{R}_{lt}$ selling varieties $j \in \mathcal{J}^r_{lt}$ that compete in Bertrand-Nash equilibrium and set prices by solving:

$$\max_{\boldsymbol{p}^r_{j,lt}} = \sum_{j \in \mathcal{J}^r_{lt}} \left(p^r_{j,lt} - c^r_{j,lt} - p^w_{j,lt}\right) s_{j,lt}(\boldsymbol{p}^r; \Theta^d) M_{lt}$$

where $p^r_{j,lt}$ are retail prices, $c^r_{j,lt}$ are distribution costs, $p^w_{j,lt}$ are wholesale prices and M_{lt} is the market size.

Use FOCs to decompose retail prices $\boldsymbol{p}^r_{lt}$:

$$\underbrace{\boldsymbol{p}^r_{lt}}_{\text{Retail price}} = -\underbrace{\left(\boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}^r_{lt}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^r; \boldsymbol{\Theta}^d)}_{\text{Retail markup}} + \underbrace{\boldsymbol{p}^w_{lt} + \boldsymbol{c}^r_{lt}}_{\text{Marginal cost - retail}}$$

where $\boldsymbol{\Delta}_{lt}$ is matrix of price partials and $\boldsymbol{\Omega}^r_{lt}$ is the retail ownership matrix.

Empirical framework - Upstream market structure

Back

In each market, there is a set of **manufacturers** $\mathcal{M}_{lt}$ selling varieties $j \in \mathcal{J}_{lt}^{w}$ that compete in Bertrand-Nash equilibrium and set prices by solving:

$$\max_{\boldsymbol{p}_{j,lt}^{w}} = \sum_{j \in \mathcal{J}_{lt}^{w}} \left(p_{j,lt}^{w} - c_{j,lt}^{w}\right) s_{j,lt}(\boldsymbol{p}^{r}; \Theta^{d}) M_{lt}$$

where $p_{j,lt}^{m}$ is the wholesale price, $c_{j,lt}^{w}$ are production costs.

Use FOCs to decompose wholesale prices $\boldsymbol{p}_{lt}^{w}$:

$$\underbrace{\boldsymbol{p}_{lt}^{w}}_{\text{Wholesale price}} = -\underbrace{\left(\mathbf{PT}_{lt} \cdot \boldsymbol{\Delta}_{lt} \odot \boldsymbol{\Omega}_{lt}^{w}\right)^{-1} \cdot \boldsymbol{s}_{lt}(\boldsymbol{p}^{r}; \boldsymbol{\Theta}^{d})}_{\text{Manufacturer markup}} + \underbrace{\boldsymbol{c}_{lt}^{m}}_{\text{Marginal cost - Manufacturer}}$$

where $\boldsymbol{\Delta}_{lt}$ is matrix of price partials, $\boldsymbol{\Omega}_{lt}^{w}$ is the manufacturer ownership matrix and $\mathbf{PT}_{lt}$ is the absolute pass-through matrix.

References I

Allen, T., Arkolakis, C., & Takahashi, Y. (2020). Universal gravity. *Journal of Political Economy*, *128*, 393-433.

Anderson, J. E., & Wincoop, E. V. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, *93*, 170-192.

Asplund, M., & Friberg, R. (2001). The law of one price in scandinavian duty-free stores. *American Economic Review*, *91*, 1072-1083.

Atkin, D., & Donaldson, D. (2015). *Who's getting globalized? the size and implications of intra-national trade costs.*

Beck, G. W., Kotz, H. H., & Zabelina, N. (2020, 11). Price gaps at the border: Evidence from multi-country household scanner data. *Journal of International Economics*, *127*, 1033-1068.

Berry, S., Levinsohn, J., & Pakes, A. (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *The American Economic Review*, *89*, 400-431.

References II

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014, 11). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*, 2295-2326. doi: 10.3982/ecta11757

Cavallo, A., Neiman, B., & Rigobon, R. (2014). Currency unions, product introductions, and the real exchange rate. *The Quarterly Journal of Economics*, *129*, 529-595.

Chatterjee, S. (2023, 8). Market power and spatial competition in rural india. *The Quarterly Journal of Economics*, *138*, 1649-1711. doi: 10.1093/qje/qjad004

Coşar, K., Grieco, P., & Tintelnot, F. (2015, 7). Borders, geography, and oligopoly: Evidence from the wind turbine industry. *The Review of Economics and Statistics*, *97*, 623-637.

Crucini, M. J., Telmer, C. I., & Zachariadis, M. (2005). Understanding european real exchange rates. *American Economic Review*, *95*, 724-738.

Donaldson, D. (2018, 4). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*, 899-934.

References III

Fontaine, F., Martin, J., & Mejean, I. (2020, 5). Price discrimination within and across emu markets: Evidence from french exporters. *Journal of International Economics*, *124*, 1-19.

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets : The case of the us automobile industry. *Econometrica*, *63*, 891-951.

Goldberg, P. K., & Verboven, F. (2001). The evolution of price dispersion in the european car market. *The Review of Economic Studies*, *68*, 811-848.

Gopinath, G., Gourinchas, P. O., Hsieh, C. T., & Li, N. (2011). International prices, costs, and markup differences. *American Economic Review*, *101*, 2450-2486.

Head, K., & Mayer, T. (2019). Brands in motion: How frictions shape multinational production†. *American Economic Review*, *109*, 3073-3124. doi: 10.1257/aer.20161345

Kalouptsidi, M. (2018, 4). Detection and impact of industrial subsidies: The case of chinese shipbuilding. *The Review of Economic Studies*, *85*, 1111-1158.

References IV

Loecker, J. D. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, *79*, 1407-1451.

Loecker, J. D., Goldberg, P. K., Khandelwal, A. K., & Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, *84*, 445-510.

McCallum, J. (1995). National borders matter: Canada - us regional trade patterns. *American Economic Review*, *85*, 615-623.

Santamaría, M., Ventura, J., & Yeşilbayraktar, U. (2023). Exploring european regional trade. *Journal of International Economics*.

Shiue, C. H., & Keller, W. (2007). Markets in china and europe on the eve of the industrial revolution. *The American Economic Review*, *97*, 1189-1216.

Silva, J. S., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, *88*, 641-658.