An empirical analysis of the spatial variability of fuel prices in the United States

Juste Raimbault[1,2,3,4] **and Antonin Bergeaud**[5,6,7]

[1]**LaSTIG, Univ. Gustave Eiffel, IGN-ENSG**
[2]**CASA, UCL**
[3]**UAR 3611 CNRS ISC-PIF**
[4]**UMR CNRS 8504 Géographie-cités**
[5]**HEC Paris**
[6]**CEPR, LSE**
[7]**Innovation Lab, Collège de France**

`juste.raimbault@ign.fr`

Séminaire INSEE - DEE, 17/02/2026

Outline

Paper

Bergeaud, A., & Raimbault, J. (2020). An empirical analysis of the spatial variability of fuel prices in the United States. *Transportation Research Part A: Policy and Practice, 132*, 131-143.

Transportation Research Part A 132 (2020) 131–143

ELSEVIER

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

TRANSPORTATION RESEARCH

An empirical analysis of the spatial variability of fuel prices in the United States

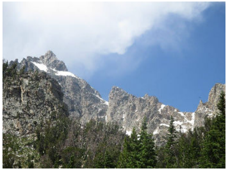
Check for updates

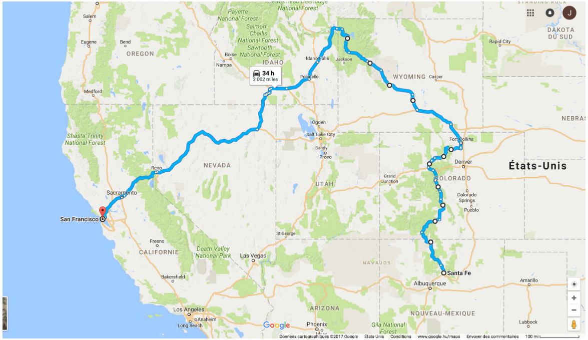
Antonin Bergeaud[a], Juste Raimbault[b,c,d,*]

[a] Bank of France, Paris, France
[b] Center for Advanced Spatial Analysis, UCL, London, UK
[c] UPS CNRS 3611 ISC-PIF, Paris, France
[d] UMR CNRS 8504 Géographie-cités, Paris, France

Accidental fieldwork on the variability of gas prices









Some questions in the fuel price litterature

1. **Spatial Dynamics and Behavioral Response**
   - **Cross-border Arbitrage:** How do price discontinuities (e.g., AZ-CA border) drive “fuel tourism” and spatial tax competition? [Rietveld et al., 2001]
   - **Inter-modality:** How does spatial price heterogeneity shift user behavior between transportation modes? [Combes and Lafourcade, 2005]
2. **Policy Design and Externalities**
   - **Optimal Taxation:** How to design tax systems that internalize negative externalities (pollution, congestion) across diverse territories? [Gregg et al., 2009]
   - **Elasticity Heterogeneity:** How do consumption responses to price shocks vary by household income and geographic location? [Li et al., 2014]
3. **Price Transmission and Frictions**
   - **Pass-through Dynamics:** What is the rate of transmission from crude oil to retail? [Gautier and Saout, 2015]
   - **Asymmetric Adjustment:** Do pump prices react faster to wholesale increases than to decreases? [Deltas, 2008]

The geography of fuel prices

Spatio-temporal variability of retail fuel prices can be linked to various aspects and issues: geographical properties of a particular energy market, characteristics of the transportation system, spatial inequalities, interactions between transportation and territories, . . .

Research Objective: *Empirical analysis of a large and detailed dataset of US fuel retail price, focusing on spatio-temporal variability*

→ Focus on **geographical patterns and structures**

→ **Complementarity** of spatial analysis methods

→ **Interdisciplinary** viewpoint on a multi-dimensionnal system

Our contributions:

- we collect novel micro-data of retail fuel prices with full geographical coverage in the US
- we show the existence of different spatial regimes, superposed to a strong determination by state-level taxes and local socio-economic characteristics
- we introduce a minimal theoretical model explaining the effect of population density on prices

Outline

Dataset Construction

Crowdsourced Big Data as a "new" way to understand better complex socio-technical systems

$\rightarrow$ Construction of a large scale dataset covering most of US fuel stations on two months, by collecting crowdsourced data available on online sources

Requirements for the crawler: Flexibility, performance and anonymity

Architecture: use open source tools developped by [Raimbault, 2019]

- pool of proxy tasks to pipe requests through `tor`
- manager monitors and launches collection tasks
- subtasks crawl and parse target webpages

Dataset Summary

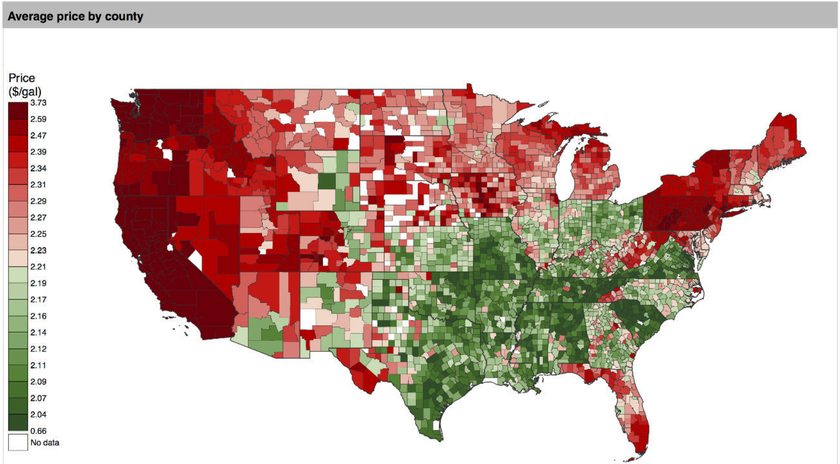
- $41 \cdot 10^6$ unique observations, between January and March 2017
  $\rightarrow$ 5,204,398 gas station - day observations for main purchase mode and regular fuel, used in the analysis
- Socio-economic data from US Census Bureau and other official sources
- Aggregation at the County level for the rest of the analysis

Table: Descriptive statistics on Fuel Price ($ per gallon)

Mean	Std. Dev.	p10	p25	p50	p75	p90
2.28	0.27	2.02	2.09	2.21	2.39	2.65

Data Exploration

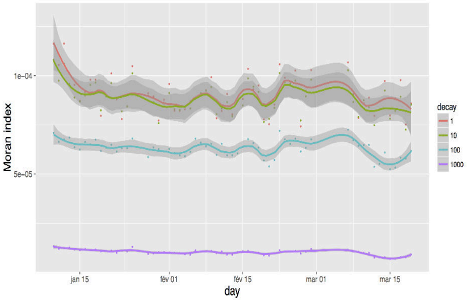
Interactive web-application for spatio-temporal exploration [Kwan, 2000]
https://analytics.huma-num.fr/geographie-cites/fuelprice/

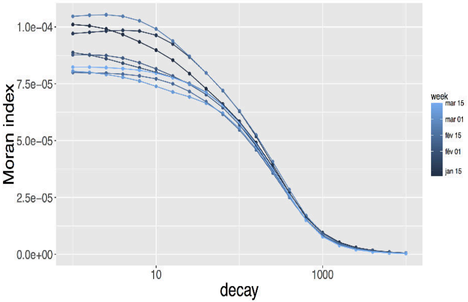


Outline

Spatio-temporal correlations

Variability in space and time of Moran spatial auto-correlation index unveils strong non-stationnarity





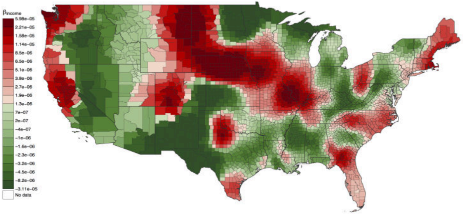
Geographically Weighted Regression

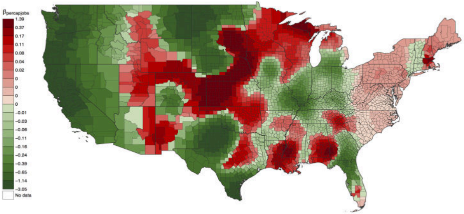
- **Geographically Weighted Regression** (GWR) [Brunsdon et al., 1996] to tackle spatial non-stationarity: capturing a spatially heterogeneous influence of socio-economic variable on prices
- Models based on basic socio-economic variables: income, population, average wage, jobs per capita, jobs
- Multi-modeling using corrected AIC for model selection: find the best model and the best bandwidth controlling for overfitting

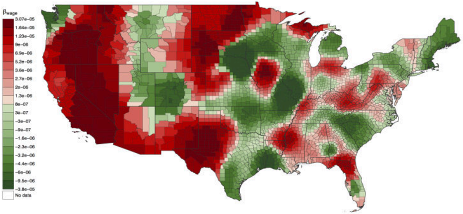
Best model : $price = \beta \cdot (income, wage, percap\, jobs)$ with an optimal bandwidth around 75km (22 neighbors with adaptive bandwidth), interpreted as spatial stationarity scale of price processes in relation to economic agents.

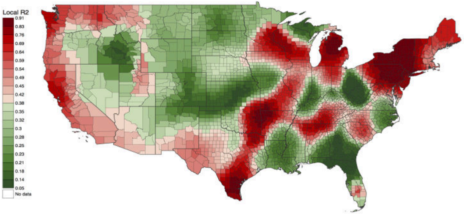
Geographically Weighted Regression: Results

Fitted coefficients and R^2 for the best model









GWR summary

1. **Heterogeneous Spatial Performance**
    - **Key Clusters:** High model fit on West Coast, South Border, North-East, and Chicago-Texas corridor.
    - **Variable Regimes:**
        - *Income:* Inverted influence based on coastal proximity (economic specialization).
        - *Wage/Employment:* Sharp East-West splits, suggesting cultural or structural drivers beyond state policy.
2. **Spatial Stationarity Scale**
    - **Optimal Bandwidth:** approximately log-normal distribution of distances with the adaptive optimal bandwidth
    - Median distance of **77 km** (IQR: 30 km).
    - **Possible interpretation:** defines the range of **coherent market competition** and the spatial stationarity of price-setting processes.

Multi-level Regression

Fixed effect regression at the level of the state : multi-level modeling to capture spatial effect of administrative boundaries

$$log(x_i) = \beta_0 + X_i\beta_1 + \beta_{s(i)} + \varepsilon_i, \quad (1)$$

- Clustering of standard error at the state level motivated by the strong spatial autocorrelation: capture county-level variation controlling for State fixed effect
- Regressing the log of price on a state fixed-effect explains 76% of the variance
- Influence of taxes: regressing the log of oil price on the level of state tax gives a R-squared of 0.33

Multi-level Regression: results

	(1)	(2)	(3)	(4)	(5)	(6)
Density (log)	-0.006^{***}	-0.006^{***}	-0.007^{***}	-0.008^{***}	-0.009^{***}	-0.011^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Cars		-0.151^{***}	-0.150^{***}	-0.187^{***}	-0.179^{***}	-0.165^{***}
		(0.051)	(0.051)	(0.019)	(0.020)	(0.020)
Nb Stations (log)			-0.006^{***}	-0.005^{***}	-0.006^{***}	-0.006^{***}
			(0.002)	(0.001)	(0.001)	(0.001)
Vote				-0.040^{***}	-0.052^{***}	-0.041^{***}
				(0.006)	(0.008)	(0.008)
Poverty					-0.075^{***}	-0.063^{***}
					(0.016)	(0.018)
Manuf					-0.020^{***}	-0.014^{***}
					(0.004)	(0.004)
Bachelor						0.060^{***}
						(0.013)
Unemployment						0.238^{***}
						(0.063)
R Squared	0.767	0.777	0.779	0.816	0.825	0.828
Observations	3059	3059	3059	3046	2993	2993

Multi-level regression summary

- Strong influence of state-level tax
- Prices negatively correlated with population density
- Fuel price also decreases with poverty
- It decreases with the vote for a Republican candidate: suggests a circular link
- Local socio-economic features have explanatory power even when removing State fixed effect

Outline

A minimal theoretical model

Modelling the role of population density in a minimal theoretical model:

- Extension of the Salop model of spatial competition [Salop and Stiglitz, 1977]
- Exogenous population density
- Transportation cost is a function of the firms' output (gas stations)

Model description

- N consumers $n = \{1..N\}$ and J gas station $j = \{1..J\}$, on a unit circle (positions $\theta(n)$ and $\phi(j) = \frac{2\pi j}{J}$)
- Fixed population distribution $\mathscr{H}(\theta)$
- Price $p(j)$ gives a cost $C(n,j) = \eta p(j) |\theta(n) - \phi(j)|$ (η consumption rate)
- Indifference condition between stations j and $j+1$:

$$p(j)\left(1+\eta\left(\theta(n)-\frac{2\pi j}{J}\right)\right) = p(j+1)\left(1-\eta\left(\theta(n)-\frac{2\pi(j+1)}{J}\right)\right)$$

- Position of the indifferent customer:

$$\theta(n) = \frac{1}{\eta}\frac{p(j+1)-p(j)}{p(j+1)+p(j)} + \frac{2\pi j}{J} + \frac{2\pi}{J}\frac{p(j+1)}{p(j+1)+p(j)} \equiv \theta(j,j+1)$$

- Profit of gas station j:

$$\pi(j,p(j),p(j+1),p(j-1)) = (p(j)-P)\int_{\theta(j-1,j)}^{\theta(j,j+1)} d\mathscr{H}(\theta)$$

Model resolution

Maximisation of π for each station gives the First Order Condition giving the prices $\vec{p}$:

$$\int_{\theta(j-1,j)}^{\theta(j,j+1)} d\mathscr{H}(\theta) = \left(\frac{2}{\eta}+\frac{2\pi}{J}\right)(p(j)-P)\left(\frac{p(j+1)h(\theta(j,j+1))}{(p(j)+p(j+1))^2}+\frac{p(j-1)h(\theta(j-1,j))}{(p(j)+p(j-1))^2}\right) \tag{2}$$

- No tractable solution for any distribution $\mathscr{H}$
- Minimise computationally $C[h,J,\eta,P]=\sum_j f_j(\vec{p},h,J,\eta,P)^2$ with a genetic algorithm in R [Scrucca, 2013]

Computational solutions to the minimal model

With parameters $\eta = 1$, $P = 0.8$, $J = 200$ gas stations, prices for 3 population densities: (i) uniform; (ii) linear; (iii) shifted exponential.

$\rightarrow$ consistent with the empirical fact of **decreasing prices with population density**

Outline

Discussion

→ **Complementarity** of spatial analysis (Moran and GWR) and econometrics methods (Multi-level regression), and a **minimal theoretical model**

→ Possible design of **territory-targeted car-regulation policies**, allowing both sustainability and territorial equity

Future work:

→ Microscopic data analysis, calibration of consumer/producer behavior models

→ Longer time series and time-series modeling

→ Parametrisation of a large-scale ABM of the spatialized fuel market: investigation of adaptive policies effects at the local and global level

→ Similar analysis on French open data

Conclusion

→ Insight into **spatio-temporal dynamics** of fuel price and their determinants with **novel "big" micro data**

→ **Complementary of approaches** and conclusions through GWR and multi-level modeling

→ Fundamental role of **spatio-temporal non-stationarity** and existence of typical stationarity scales: back on the importance of **fieldwork**

Reproducibility:
All code and aggregated data available at `https://github.com/JusteRaimbault/EnergyPrice`

References I

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1996).
Geographically weighted regression: a method for exploring spatial nonstationarity.
Geographical analysis, 28(4):281–298.

Combes, P.-P. and Lafourcade, M. (2005).
Transport costs: measures, determinants, and regional policy implications for france.
Journal of Economic Geography, 5(3):319–349.

Deltas, G. (2008).
Retail gasoline price dynamics and local market power.
The Journal of Industrial Economics, 56(3):613–628.

Gautier, E. and Saout, R. L. (2015).
The dynamics of gasoline prices: Evidence from daily french micro data.
Journal of Money, Credit and Banking, 47(6):1063–1089.

References II

Gregg, J. S., Losey, L. M., Andres, R. J., Blasing, T., and Marland, G. (2009).
The temporal and spatial distribution of carbon dioxide emissions from fossil-fuel use in north america.
Journal of Applied Meteorology and Climatology, 48(12):2528–2542.

Kwan, M.-P. (2000).
Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set.
Transportation Research Part C: Emerging Technologies, 8(1-6):185–203.

Li, S., Linn, J., and Muehlegger, E. (2014).
Gasoline taxes and consumer behavior.
American Economic Journal: Economic Policy, 6(4):302–42.

References III

Raimbault, J. (2019).
Exploration of an interdisciplinary scientific landscape.
Scientometrics, 119(2):617–641.

Rietveld, P., Bruinsma, F., and Van Vuuren, D. (2001).
Spatial graduation of fuel taxes; consequences for cross-border and domestic fuelling.
Transportation Research Part A: Policy and Practice, 35(5):433–457.

Salop, S. and Stiglitz, J. (1977).
Bargains and ripoffs: A model of monopolistically competitive price dispersion.
The Review of Economic Studies, 44(3):493–510.

Scrucca, L. (2013).
GA: A package for genetic algorithms in R.
Journal of Statistical Software, 53(4):1–37.

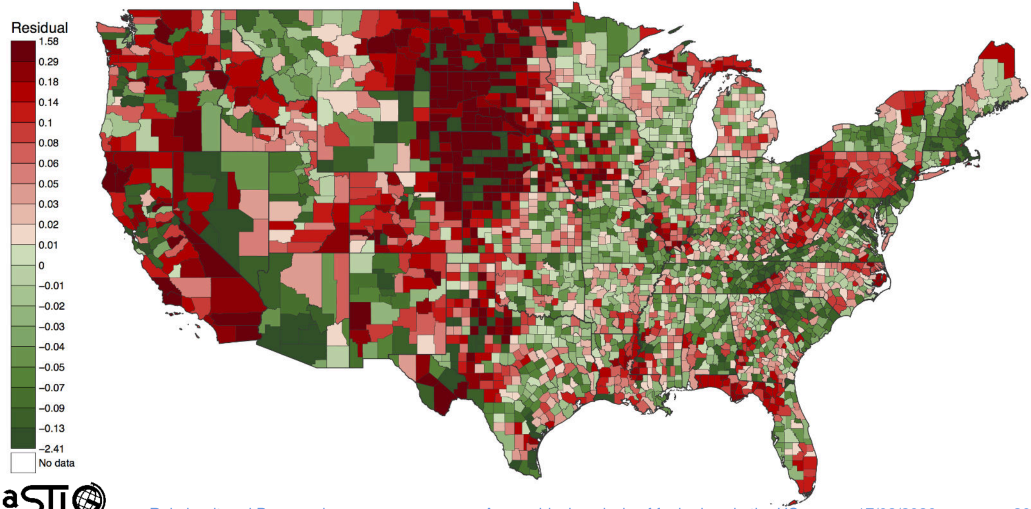
Spatial Heterogeneity

Spatial Autocorrelation as an index of spatial variability, for link i

$$\rho_i = \frac{1}{K} \cdot \sum_{i \neq j} w_{ij} \cdot (c_i - \bar{c})(c_j - \bar{c}) \tag{3}$$

with spatial weights $w_{ij} = \exp\left(\frac{-d_{ij}}{d_0}\right)$

GWR Residuals



Socio-economic variables

Table: Source and description of the county level variables

Variable	Description	Source
Density	Household units per sq miles	Census 2010
Unemployment	Total employed over labor force in 2017	BLS
Nb Stations	Number of gas station	Own calculation
Manuf	Employment share of the manufacturing sector in 2016	CBP - Census
Bachelor	Share of population over 25 with a bachelor degree in 2017	ACS
Cars	Share of people over 16 using their car to commute in 2017	ACS
Poverty	Share of people considered in poverty by SAIPE program	SAIPE - Census
Vote GOP	Share of voters that voted for a republican candidate in the general elections from 2000	MIT election lab

Socio-economic variables

Table: Descriptive statistics

Variable	mean	sd	p10	p25	p50	p75	p90	units
Density	113.4	823.8	2.2	8.4	21.0	51.5	164.2	Households per sq mile
Unemployment (x100)	4.6	1.6	2.9	3.5	4.3	5.3	6.5	Share
Nb Stations	28.0	65.9	1.7	3.3	9.4	24.7	64.9	Number
Manuf (x100)	19.3	17.0	0	6.1	15.4	28.5	43.0	Share
Bachelor (x100)	21.2	9.3	12.1	14.7	19	25.3	33.7	Share
Cars (x100)	89.5	7.3	82.5	87.8	91.3	93.5	95	Share
Poverty (x100)	15.4	6.2	8.7	10.9	14.4	18.4	23.4	Share
Vote GOP (x100)	59.4	13.2	41.8	51.4	60.5	68.7	75.5	Share

High dimensional fixed effects

With $x_{i,s,c,t}$ the fuel price in day t, in gas station i, in state s and in county c, we start by running high dimensional fixed effect regressions

$$
\begin{aligned}
x_{i,s,c,t} &= \beta_s + \varepsilon_{i,s,c,t} && (4)\\
x_{i,s,c,t} &= \beta_c + \varepsilon_{i,s,c,t} && (5)\\
x_{i,s,c,t} &= \beta_i + \varepsilon_{i,s,c,t} && (6)\\
& && (7)
\end{aligned}
$$

Where $\varepsilon_{i,s,c,t}$ contains an idiosyncratic error and a day fixed effect.
$\rightarrow$ It confirms that most of the variance can be explained by a state fixed effect and that integrating more accurate levels has only small effect on the fit of our model as measured by the R-squared.

Model resolution

First order condition for the profit maximisation problem:

$$\int_{\theta(j-1,j)}^{\theta(j,j+1)} d\mathscr{H}(\theta) + (p(j) - P)\left[\frac{d\theta(j,j+1)}{dp(j)} h(\theta(j,j+1)) - \frac{d\theta(j-1,j)}{dp(j)} h(\theta(j-1,j))\right] = 0,$$

with

$$\frac{d\theta(j,j+1)}{d(p(j))} = \frac{-p(j+1)}{(p(j)+p(j+1))^2}\left(\frac{2}{\eta} + \frac{2\pi}{J}\right) < 0$$

and

$$\frac{d\theta(j-1,j)}{d(p(j))} = \frac{p(j-1)}{(p(j-1)+p(j))^2}\left(\frac{2}{\eta} + \frac{2\pi}{J}\right) > 0$$