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

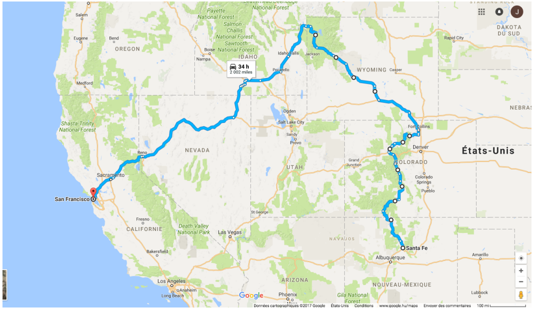
J. Raimbault[1] and A. Bergeaud[2]
[1]**LaSTIG, Univ. Gustave Eiffel, IGN-ENSG**
[2]**HEC Paris**
juste.raimbault@ign.fr

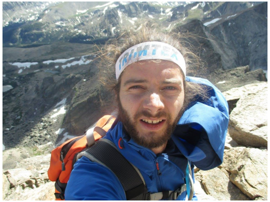
Séminaire INSEE - DEE, 17/02/2026

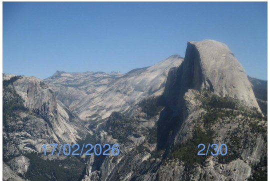
Outline

1. Introduction
2. Dataset construction and visualisation
3. Econometric analysis
4. Theoretical model
5. Discussion

Accidental fieldwork on the variability of gas prices







Geography and Fuel prices

Spatio-temporal variability of Fuel Price can be linked to various aspects: geographical properties of a particular energy market, characteristics of the transportation system, interactions between transportation and territories.

Diverse approaches in the litterature:

- [Rietveld et al., 2001] cross-border variability
- [Gregg et al., 2009] influence on carbon emissions
- [Combes and Lafourcade, 2005] effective transportation costs
- [Gautier and Saout, 2015] from crude oil price to retail price

Exploratory spatial analysis

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 point of view on a fundamentally heterogenous system

Outline

1. Introduction
2. Dataset construction and visualisation
3. Econometric analysis
4. Theoretical model
5. Discussion

Dataset Construction

Crowdsourced Big Data as a new way to unveil structure of complex socio-technical systems

$\rightarrow$ Construction of a large scale dataset covering most of US fuel stations on two month

Requirements : Flexibility, performance and anonymity

Architecture : 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, aggregated at the county level
- Socio-economic data from US Census Bureau

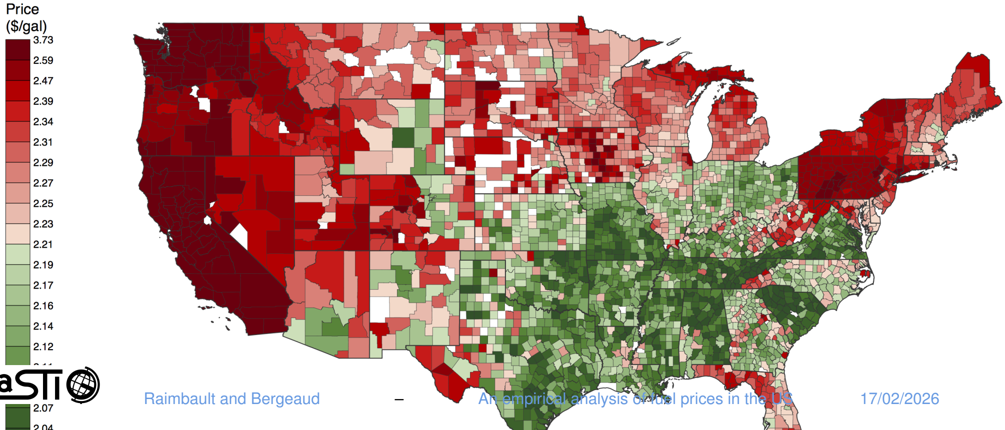
Table: Descriptive statistics on Fuel Price ($ per gallon)

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

Average price by county



Outline

1. Introduction

2. Dataset construction and visualisation

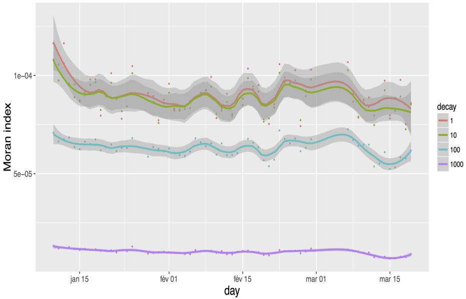
3. Econometric analysis

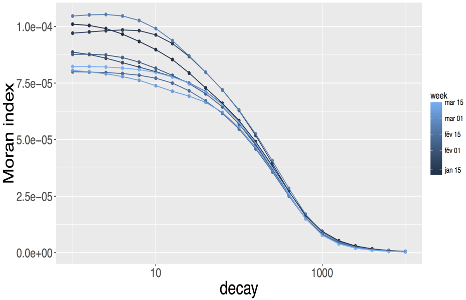
4. Theoretical model

5. Discussion

Spatio-temporal correlations

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





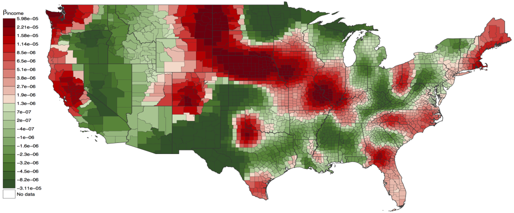
Geographically Weighted Regression

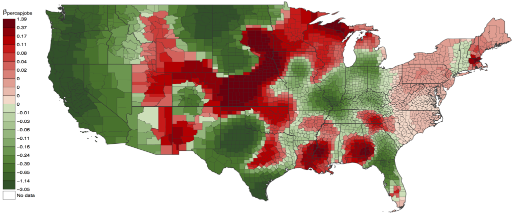
- Geographically Weighted Regression (GWR) [Brunsdon et al., 1996] to tackle spatial non-stationarity: capturing a variable influence in space of explicative variables for prices
- Multi-modeling using corrected AIC for model selection : find the best model and the best bandwidth controlling for overfitting

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





Multi-level Regression

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

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

- 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 74% 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 extent to which a county has voted for a Republican candidate: suggests a circular link
- Overall, local socio-economic features have explanatory power even when removing State fixed effect

Outline

1. Introduction
2. Dataset construction and visualisation
3. Econometric analysis
4. Theoretical model
5. Discussion

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:

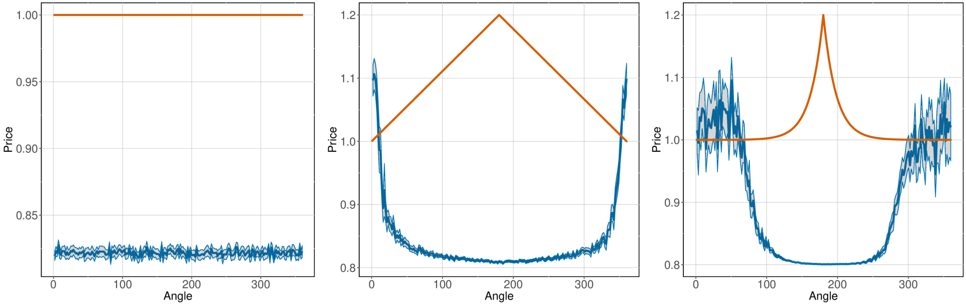
$$\pi(j,p(j),p(j+1),p(j-1)) = (p(j)-P)\int_{\theta(j-1,j)}^{\theta(j,j+1)} d\mathscr{H}(\theta)$$

Model resolution

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

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

- 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


1.00
0.95
0.90
0.85
Price
0
100
200
300
Angle
1.2
1.1
1.0
0.9
0.8
Price
0
100
200
300
Angle
1.2
1.1
1.0
0.9
0.8
Price
0
100
200
300
Angle

Outline

1. Introduction

2. Dataset construction and visualisation

3. Econometric analysis

4. Theoretical model

5. Discussion

Discussion

→ Complementarity of Spatial analysis and econometrics methods
→ Possible design of territory-targeted car-regulation policies, allowing both sustainability and territorial equity

Future work:
→ Microscopic data analysis (requires precise geocoding)
→ Longer time series and time-series modeling
→ Parametrization of a large-scale ABM of the spatialized fuel market : investigation of adaptive policies effects at the local and global level
→ Scales and ontologies to study relations between network and territories

Conclusion

→ A novel insight into spatio-temporal dynamics of fuel price and their determinants thanks to big data harvesting

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

→ Fundamental role of spatio-temporal non-stationarity and existence of typical stationarity scales

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.

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.

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.

References II

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.

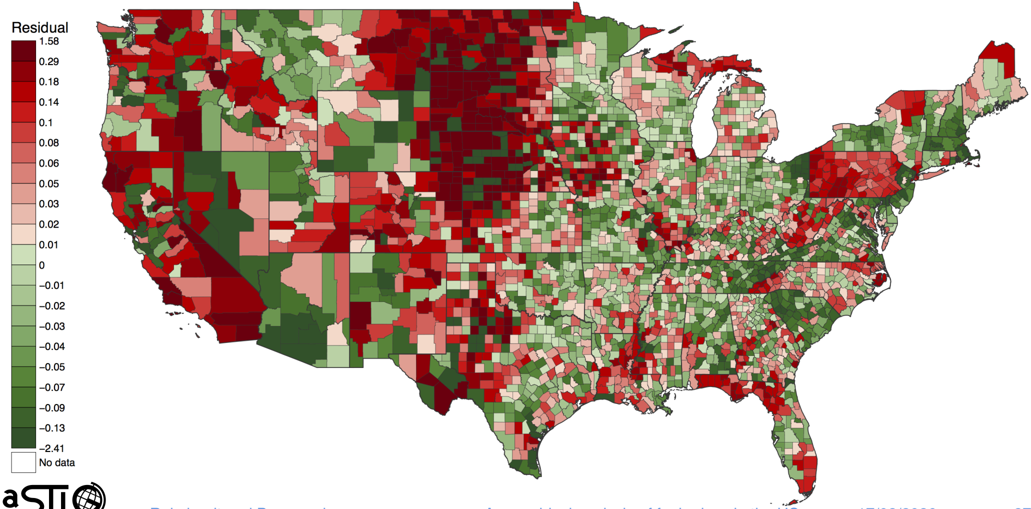
Spatial Heterogeneity

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.