The Cost of Transportation : Spatial Analysis of US Fuel Prices

J. Raimbault[1,2], A. Bergeaud[3]
juste.raimbault@polytechnique.edu

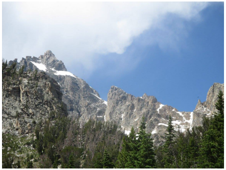
[1]UMR CNRS 8504 Géographie-cités
[2]UMR-T IFSTTAR 9403 LVMT
[3]Paris School of Economics

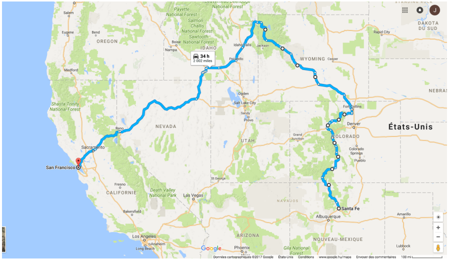
EWGT 2017 - Budapest

Session 10B: Transport Economics and Financing

September 6th, 2017

Road Trippin' : Californication





Geography and Fuel prices

→ In the US you fuel your (necessary big) car everywhere and everyday, and directly feel the strong variability of prices !

Spatio-temporal variability of Fuel Price captures geographical properties of a particular energy market, of the transportation system, of interactions between transportation and territories.

Diverse approaches mostly by economists :

- [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 : *Construction and Exploratory 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

Dataset Construction

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

→ 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 available upon request, "grey area" of semi-open data.

Dataset Summary

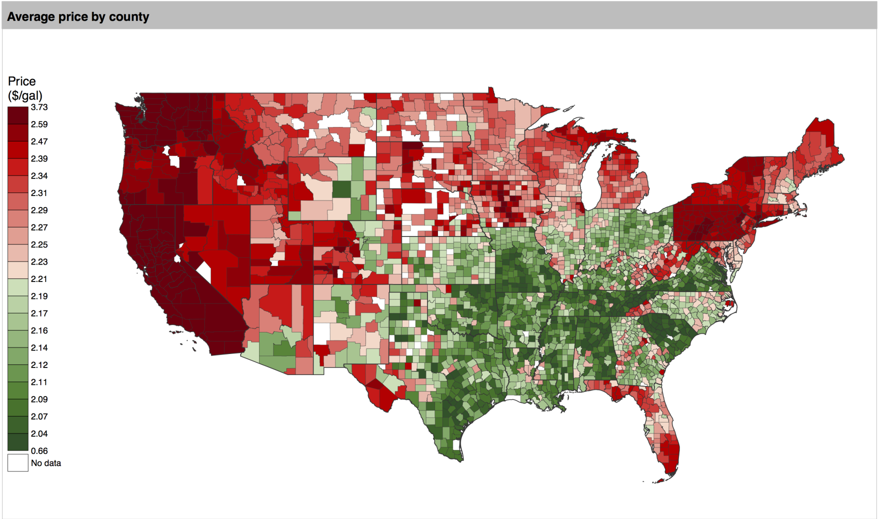
- $41 \cdot 10^6$ unique observations, between January and March 2017 $\rightarrow$ 5,204,398 gas station - day observations for main purchase mode and regular fuel, used in the analysis, aggregated at the county level
- Socio-economic data from US Census Bureau

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

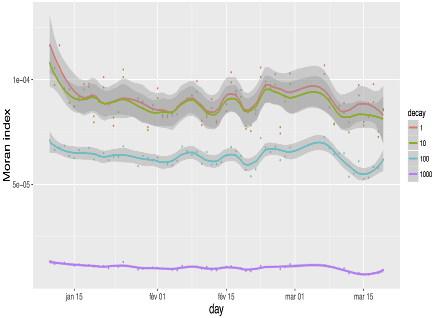
Data Exploration

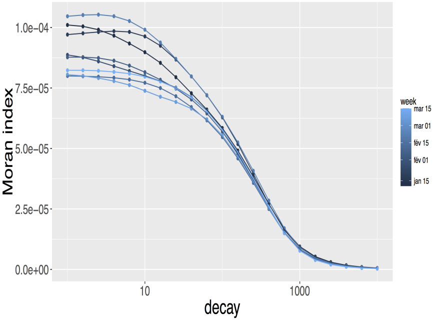
Interactive web-application for spatio-temporal exploration



Spatio-temporal correlations

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





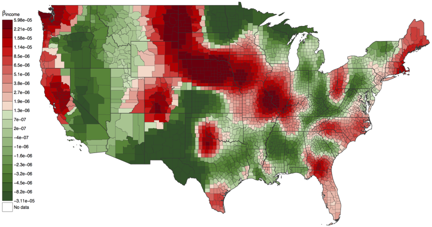
Geographically Weighted Regression

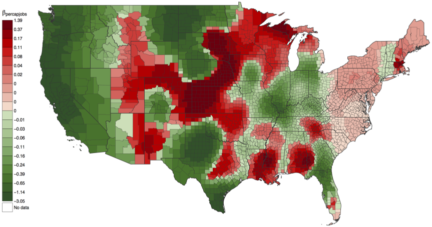
- Geographically Weighted Regression (GWR) [Brunsdon et al., 1996] to tackle spatial non-stationarity : capturing a variable influence in space of explicative variables for prices
- Multi-modeling using corrected AIC for model selection : find the best model and the best bandwidth controlling for overfitting

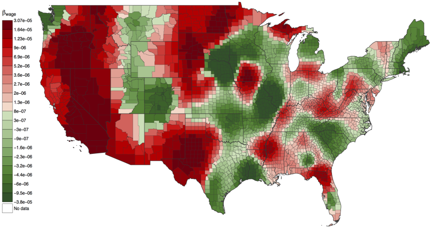
Best model : $price = \beta \cdot (income, wage, percapjobs)$ with an optimal bandwidth around 75km (22 neighbors with adaptive bandwidth), interpreted as spatial stationarity scale of price processes in relation to economic agents.

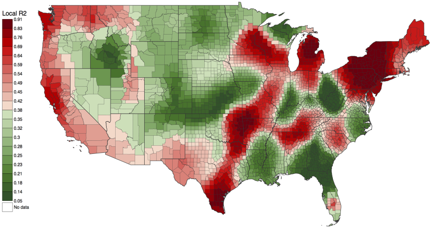
Geographically Weighted Regression : Results

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

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

- 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 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)
Density		0.016***	0.016***	0.016***	0.015***
		(0.002)	(0.001)	(0.001)	(0.001)
Population (log)		-0.007***	-0.040***	-0.041***	-0.039***
		(0.001)	(0.011)	(0.011)	(0.010)
Total Income (log)			0.031***	0.031***	0.027***
			(0.010)	(0.010)	(0.009)
Unemployment			0.001	0.000	0.000
			(0.001)	(0.001)	(0.001)
Poverty			-0.028**	-0.030***	-0.029**
			(0.011)	(0.011)	(0.011)
Percentage Black				0.000***	-0.000
				(0.000)	(0.000)
Vote GOP					-0.072***
					(0.015)
R-squared	0.743	0.767	0.774	0.776	0.781
N	3,066	3,011	3,011	3,011	3,011

Multi-level Regression : summary

- Strong influence of state-level tax
- Dense urban counties have higher fuel price, but price decreases with population
- Fuel price increases with total income, 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 when removing State fixed effect

Implications

Methodological

→ Complementarity of Spatial analysis and econometrics methods : towards integrated approaches to territorial systems

Practical

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

Possible Developments

→ Microscopic data analysis (requires precise geocoding)

→ Longer time series (collection in progress) 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 available at `https://github.com/JusteRaimbault/EnergyPrice`

- Raw dataset available upon request

- Paper preprint available at `http://arxiv.org/abs/1706.07467`

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.

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.

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.

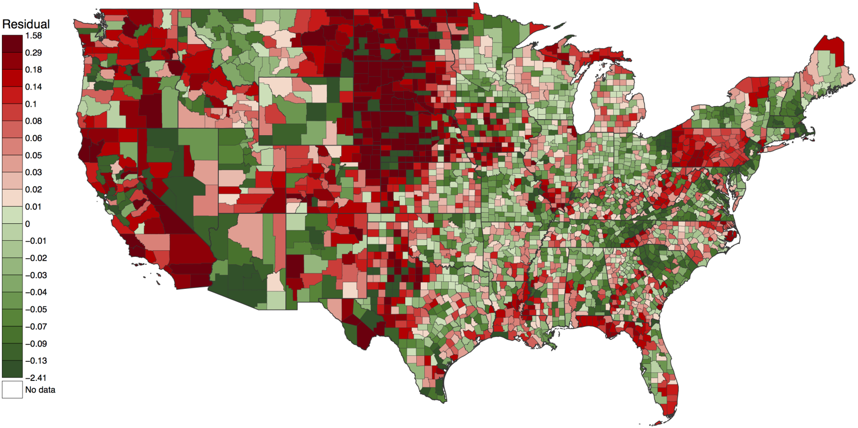
Spatial Heterogeneity

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

$$\rho_i = \frac{1}{K} \cdot \sum_{i \neq j} w_{ij} \cdot (c_i - \bar{c})(c_j - \bar{c}) \tag{2}$$

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

GWR Residuals



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

$$x_{i,s,c,t} = \beta_s + \varepsilon_{i,s,c,t} \quad (3)$$

$$x_{i,s,c,t} = \beta_c + \varepsilon_{i,s,c,t} \quad (4)$$

$$x_{i,s,c,t} = \beta_i + \varepsilon_{i,s,c,t} \quad (5)$$

$$(6)$$

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.