Available online at www.sciencedirect.com

SciVerse ScienceDirect

Transportation Research Procedia 00 (2016) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

20th EURO Working Group on Transportation Meeting, EWGT2017, 4-6 September 2017, Budapest, Hungary

The Cost of Transportation : Spatial Analysis of US Fuel Prices

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

[a] UMR CNRS 8504 Géographie-cités, 13 rue du Four, Paris, France
[b] UMR-T IFSTTAR 9403 LVMT, Cité Descartes, Champs-sur-Marne, France
[c] Banque de France, Paris, France
[d] Paris School of Economics - EHESS, Paris, France

Abstract

The geography of fuel prices has several implications, from a significant impact on accessibility patterns to issues of territorial equity and transportation governance. In this paper, we study its spatio-temporal patterns at a high resolution. We construct a dataset collecting daily oil prices for two months, on a significant proportion of US gaz facilities, using a specifically-designed large scale data crawling technology. The implementation of a web-application for interactive spatio-temporal data exploration guide further statistical investigations, namely that oil price exhibit patterns that are strongly non-stationary in space and time. The behavior of spatial autocorrelation suggests the use of specific spatial econometric methods to study the role of explanatory variables that are either geographical or temporal. We study the influence of socio-economic variables, by using complementary methods: Geographically Weighted Regression to take into account spatial non-stationarity, and Multi-level modeling to condition both at the state and county level. The former yields an optimal spatial range roughly corresponding to stationarity scale, and significant influence of variables such as population density or median income, but is less accurate around administrative borders. On the other hand, multi-level modeling reveals a strong state fixed effect, and also a non-negligible county effect. Through the combination of such methods, we unveil the superposition of a governance process with a local socio-economical spatial process. Results are furthermore consistent across the different dates. We discuss one important application that is the elaboration of locally parametrized car-regulation policies.

Peer-review under responsibility of the Scientific Committee of EWGT2017.

Keywords: Fuel Price, Data Crawling, Spatial Analysis, Geographically Weighted Regression, Multi-level Modeling

1. Introduction

The local cost of fuel is the outcome of heterogeneous processes at multiple scales, from raw oil prices to governance processes inducing taxes and economic interests of various agents including retail companies and consumer. If its variability in time is an evidence, the question of its variability in space has been poorly tackled in the literature, although it may have several empirical applications such as unveiling territorial inequalities, geographical

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: juste.raimbault@polytechnique.edu

2214-241X
Peer-review under responsibility of the Scientific Committee of EWGT2017.

singularities, or serving as a proxy of a socio-economical status, but also practical application such as a higher level of transparency for the consumer or a possible feedback of prices analyses on the process itself (regulation through increased information).

C : (Juste) j'ai pas trop eu le temps de regarder la biblio mais c'est bizarre j'ai l'impression qu'il n'y a quasiment personne qui a fait ce qu'on veut faire, je dois mal chercher ? je mets ci-dessous de trucs relis de prłt ou de loin, il faudra srement faire du tri !

There exists to our knowledge no systematic mapping in space and time of retail fuel prices for a country. First of all the availability of data may have been one significant obstacle, but the nature of the problem studied may also have influence: at the crossing of several disciplines, with Economics having studied theoretical models of price elasticity, New Economic Geography including transportation prices in spatialized models, and Transport geography being not at such a detailed level. Various example of related works can be given. For example, Rietveld et al. (2001) studies the impact of cross-border fuel differences and the implications for gradual spatial taxing in the Netherlands. At the level of countries, Rietveld and van Woudenberg (2005) provides statistical models to explain fuel price differences across European countries. Macharis et al. (2010) models the impact of spatial fuel price variation on patterns of inter-modality, implying that the spatial heterogeneity of fuel prices has a strong impact on user behavior. On an empirical view of the geography of transportation but for an other aspect that are gas emissions, Gregg et al. (2009) studies their spatial distribution, at the scale of the state. The geography of fuel prices also implications on effective costs, as shows Combes and Lafourcade (2005) by determining accurate transportation costs across urban areas for France.

The rest of the paper is organized as follows: we describe in the next section a generic procedure and tool for large scale automatized data collection, and gives summary statistics and structure of the dataset we use. We then proceed to several exploratory and statistical analyses in order to unveil the spatio-temporal structure of fuel price and investigate potential explicative variables. We finally discuss the implications for price processes, in particular comparing the two statistical methods.

2. Dataset

2.1. Collecting large scale heterogeneous data

The availability of new type of data has induced consequent changes in various disciplines from social science (e.g. online social network analysis Tan et al. (2013)) to geography (e.g. new insights into urban mobility or perspectives on "smarter" cities Batty (2013)) and economics (**C :** Anto un insight la dessus en eco ?), without being always smooth at the interface of implied disciplines, as shows the example of the difficult understanding between physics and urban sciences (Dupuy and Benguigui (2015))). The construction of such datasets is however far from straightforward, as their analyses because of the incomplete and noisy nature of data. Specific technical tools have to be implemented, often with a problem-specific tuning making them difficultly generic. We develop such a tool that fills the following constraints typical of large scale data collection: (i) reasonable level of flexibility and genericity; (ii) performance optimized, through parallel collection jobs; (iii) anonymity of collection jobs to avoid any possible bias in the behavior of the data source. The architecture, at a high level, has the following structure:

The application is open and its module are reusable: source code is available on the repository of the project[1].

2.2. Dataset

2.2.1. Fuel price data

C : (Juste) on ne mettra pas le dataset public dans un premier temps, voir ce que les reviewers en disent

Our dataset of fuel prices consists in retail fuel prices, from crowd-sourced information available on diverse internet sources, aggregated with the above-described crawling tool.

C : information density map as supplementary material ?

[1] at `https://github.com/JusteRaimbault/EnergyPrice`

Fig. 1. (a) Example of maps at different dates ; (b) Time series

Fig. 2. Behavior of spatial-autocorrelation in time for different decay parameter values.

The dataset comprises around $41 \cdot 10^6$ unique observations, spanning in time from 2017-01-10 to 2017-03-19, that correspond to 118573 unique retail stations for each the geographical location (city resolution) was obtained. **C :** statistics of observation per station ? Prices correspond to an unique purchase mode (credit card, other modes such as cash being less than 10% in test datasets, they were discarded in the final dataset) and four fuel types: Diesel (18% of observations), Regular (34%), Midgrade (24%) and Premium (24%). The best coverage of stations is for Regular fuel type (**C :** recompute proportion on full dataset), and we choose to work with this one in our exploratory analysis, keeping in mind that further developments with the dataset may include comparative analysis on fuel types.

C : Aggregate on which time span ? best is to check time to converge to a certain coverage of all stations (say 95%) → around one week seems ok (see stations count plot in repo)

C : French Open Data : https://www.prix-carburants.gouv.fr/rubrique/opendata/ → look to what extent our study could be transfered to France

2.2.2. Socio-economic data

3. Results

3.1. Spatio-temporal Patterns of Prices

We propose a first exploratory introduction to the spatio-temporal structure of the dataset, that we consider a crucial stage to guide further analyses, but also to understand their implications in a geographical context.

3.2. Spatial Autocorrelation

C : why using spatial autocorr to unveil spatial heterogeneity ; how its temporal variation can be interpeted. Figure. comments.

The two next subsections investigate potential explicative variables of local fuel prices, using two different statistical techniques.

3.3. Geographically Weighted Regression

The issue of spatial non-stationarity of geographical processes has always been a source of biased aggregated analyses or misinterpretations when applying general conclusions to local cases. To take it into account into statistical models, numerous techniques have been developed, among which the simple but very elegant Geographically Weighted Regression, that estimates non-stationary regressions by weighting observations in space similarly to kernel estimation methods. It was introduced in a seminal paper by Brunsdon et al. (1996) and has been consequently used and matured since.

C : simple GWR → maps with some estimated coefs

3.4. Multi-level Regression

C : multi-level regreg on State/zip ?

4. Discussion

4.1. On the complementarity on Spatial Analysis methods

C : bla bla on conclusions drawn with each method ; how they are complementary → also present it as a methodological contribution. some literature do both, e.g. Chen, D. R., & Truong, K. (2012). Using multilevel modeling and geographically weighted regression to identify spatial variations in the relationship between place-level disadvantages and obesity in Taiwan. Applied Geography, 32(2), 737-745. integrates both in a single approach.

4.2. Towards localized car-regulation policies

C : knowing local price dynamics and their drivers should help designing specific regulations/taxing policies ?

5. Conclusion

Acknowledgements

References

Batty, M., 2013. Big data, smart cities and city planning. Dialogues in Human Geography 3, 274–279.

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

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

Dupuy, G., Benguigui, L.G., 2015. Sciences urbaines: interdisciplinarités passive, naïve, transitive, offensive. Métropoles .

Gregg, J.S., Losey, L.M., Andres, R.J., Blasing, T., 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, 2528–2542.

Macharis, C., Van Hoeck, E., Pekin, E., Van Lier, T., 2010. A decision analysis framework for intermodal transport: Comparing fuel price increases and the internalisation of external costs. Transportation Research Part A: Policy and Practice 44, 550–561.

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

Rietveld, P., van Woudenberg, S., 2005. Why fuel prices differ. Energy Economics 27, 79–92.

Tan, W., Blake, M.B., Saleh, I., Dustdar, S., 2013. Social-network-sourced big data analytics. IEEE Internet Computing 17, 62–69.