OXFORD

Purchase Patterns, Socioeconomic Status, and Political Inclination

Xiaowen Dong, Eaman Jahani, Alfredo J. Morales, Burçin Bozkaya, Bruno Lepri, and Alex 'Sandy' Pentland

Abstract

This paper analyzes millions of credit card transaction records during several months for tens of thousands of individuals from two different countries. The study shows that, purchase patterns are strongly correlated with important societal indices such as socioeconomic status and political inclination. The results suggest the possibility of understanding and predicting the evolution of such societal indices from purchase behavioral patterns, potentially at high temporal and spatial resolutions.

1. Introduction

The emergence of new research fields such as computational social science (Lazer et al. 2009) has promoted data-driven approaches, through the unique lens of Big Data, to study and understand human and social behavior at unprecedented scale. In particular, several studies have proposed the use of large-scale cell phone data to study important economic questions, such as social welfare (Eagle, Macy, and Claxton 2010), poverty mapping (Blumenstock, Cadamuro, and On 2015), and unemployment prediction (Toole et al. 2015). In addition to cell phone data, the recent availability of large-scale financial transaction data

provides even better opportunities to study human behavior and decision making in urban environment, and implications in the organization of society at large.

Human economic behavior has deep connections with sociopolitical processes (Marshall 1890; Cassel 1967). This paper, using large-scale individual credit card transaction data from two countries, one European (Turkey) and one Latin American (Mexico), studies human purchase behavioral patterns in urban environment as well as their relation to sociopolitical factors. In particular, the study focuses on patterns of purchase diversity, that is, how diversely people visit different stores in the city. The study proposes two diversity measures, an "outgoing" diversity from the individual's perspective and an "incoming" diversity from the merchant's perspective. The study tests their correlations with important societal indices, namely, socioeconomic status and political inclination, at the census-district level. The study shows that, in both countries, purchase diversity is strongly correlated with socioeconomic status of the district; furthermore, it finds interesting links between purchase diversity and political inclination of the district, the interpretation of which, however, depends on the political context of the specific country.

These results are among the first to establish links between purchase behavioral patterns and important societal indices, and to test them on large-scale transaction data from more than one country. The results open new possibilities to understand, from only the purchase behavior, the evolution of socioeconomic status and political inclination of districts in a cost-effective way, and potentially at much higher temporal and spatial resolutions compared to census and election results. The study's findings may also foster the prediction of these important measures in a timely fashion. Finally, the proposed "incoming" diversity may help establish the potential links between economic mixing, income segregation, and political polarization, an interesting direction for future studies.

2. Data and Methods

The study analyzes credit card transaction data provided by two major financial institutions in the European and Latin American countries under consideration. The European data contain more than 10 million credit card transactions from more than 100 thousand individuals at more than 100 thousand merchants from April 1 to June 30 2013. Similarly, the Latin American data contain about 60 million transactions from about 3 million customers at about 400,000 stores from April 1 to July 31 2013. Each credit card transaction record comes with metadata such as the time and amount of the transaction, as well as the store location and the customer's home location in terms of census district. In particular, the geographical information makes it possible to investigate the spatial distribution of financial transactions and the economic links between various census districts. The data are analyzed under legal restriction against reidentification, which fully conforms to the privacy laws of the two countries.

This paper focuses on purchase activities in the largest city of the two countries. As preprocessing steps, the study first filters out foreign and online transactions to focus on local and physical activities. Stores are then removed that are located outside the cities. Finally, only customers who made at least 10 transactions in the European data set and 20 in the Latin American data set are considered in the analysis. After these preprocessing steps, about 3 million transactions remain in the European data set and 11 million in the Latin American data set, which are the data used for the analysis in the rest of the paper.

The study analyzes patterns of purchase activities and their relation to socioeconomic and political factors. In particular, it establishes links between purchase diversity of census districts and their socioeconomic status and political inclination. Two measures of purchase diversity are defined:

1. **Outgoing purchase diversity:** This captures district-level exploration, namely, how diversely individuals residing in a given district visit different merchants.
2. **Incoming purchase diversity:** This measures district-level attractiveness, namely, how diversely merchants in a given district attract different individuals.

For each individual, the outgoing purchase diversity is defined as the Shannon entropy of their purchase activities:

$$H(i) = -\sum_{j=1}^{N} p_{ij} \log(p_{ij}), \tag{1}$$

where p_{ij} is the probability that an individual i visits a merchant j within the respective data periods for the two countries and N is the total number of merchants. The diversity of a district is then defined as the average diversity of individuals living in that district. Similarly, the study computes the incoming purchase diversity of a given district by averaging the Shannon entropy of each merchant in that district in terms of attracting different individuals. It is worth noting that this approach is similar to the network-based approach of Eagle, Macy, and Claxton (2010); the difference, however, is that the network in the present case is essentially a bipartite graph where customers and stores form two sets of vertices.

This study examines the relationship between these diversity measures and two important indices at the district level. The first index is the socioeconomic status of each district, which in the case of the European data set is a composite measure between 0 and 100 that quantifies the relative prosperity of the district based on a number of indicators such as income and education level. The higher the index, the more prosperous the district is. For the Latin American data set, the measure is defined similarly but as a categorical variable with five different levels. The European socioeconomic status data are provided by VisioThink, Inc.[1] and the Latin American data are the urban marginalization index provided by the National Population Council (CONAPO) of Mexico.[2]

The second index considered is the political conservatism of each district, which is a measure between 0 and 100 that is computed based on the percentages of votes parties labeled as "liberal" or "conservative" obtained in a recent national election. The higher the index, the more conservative the district tends to be. The data are provided by the Turkish Statistical Institute for the 2011 Turkish general election,[3] and by the National Electoral Institute (INE) of Mexico for the 2012 Mexican presidential election.[4]

3. Results and Discussion

Table 1 shows the Pearson product-moment correlation coefficient between the two measures of purchase diversity and the socioeconomic/political indices of about 600 districts in the European city and more than 1500 districts in the Latin American city.[5] In both cases, purchase diversity measures are positively correlated with the socioeconomic status of the district, with a particularly strong relationship in the case of outgoing diversity ($r = 0.75$ for the European city and $r = 0.61$ for the Latin American city). This indicates that, in general, the higher the diversity, the more prosperous the district tends to be. In addition, for the European city, there exists a fairly strong negative correlation ($r = -0.65$) between outgoing purchase diversity and political conservatism: namely, the higher the diversity, the less conservative the district tends to be. Interestingly, such a negative relationship is reversed in the case of the Latin American city ($r = 0.25$ for outgoing and $r = 0.38$ for incoming), which might be due to the different

1 VisioThink, Inc. (2013), "Turkey Socio-Economic Status Database - version Quarter 4 2012." Data at the level of "mahalle" were used for the analysis.
2 National Population Council (CONAPO) (2015), "Índice de Marginación Urbana 2010." Data at the level of "AGEB" were used for the analysis.
3 Turkish Statistical Institute (2012), "General Election of Representatives Province and District Results." Data at the level of "mahalle" were used for the analysis.
4 National Electoral Institute (INE) (2013), "Atlas de Resultados Electorales Federales 1991–2012." Data at the level of "sección" were mapped to the level of "AGEB" (as in the data provided by CONAPO) for the analysis.
5 For more reliable results of diversity, for the European data set, this study considers districts with at least 10 customers for computing outgoing diversity and those with at least 10 stores for incoming diversity. For the Latin American data set the corresponding thresholds are 20 and 30, respectively.

Table 1. Pearson's *r* between Purchase Diversity Measures (both outgoing and incoming) and Indices about Socioeconomic Status and Political Conservatism

Pearson's *r*	Socioeconomic status	Political conservatism
European: Purchase diversity (outgoing)	0.75***	−0.65***
European: Purchase diversity (incoming)	0.41***	−0.31***
Latin American: Purchase diversity (outgoing)	0.61***	0.25***
Latin American: Purchase diversity (incoming)	0.26***	0.38***

Source: Authors' analysis from credit card transaction data.

Note: *** Indicates significance at the 1 percent level.

Figure 1. (a) Outgoing purchase diversity, (b) incoming purchase diversity, (c) socioeconomic status, and (d) political conservatism, for districts in the Latin American city. The darker the color, the higher the index value associated with a district.

(a)

(b)

(c)

(d)
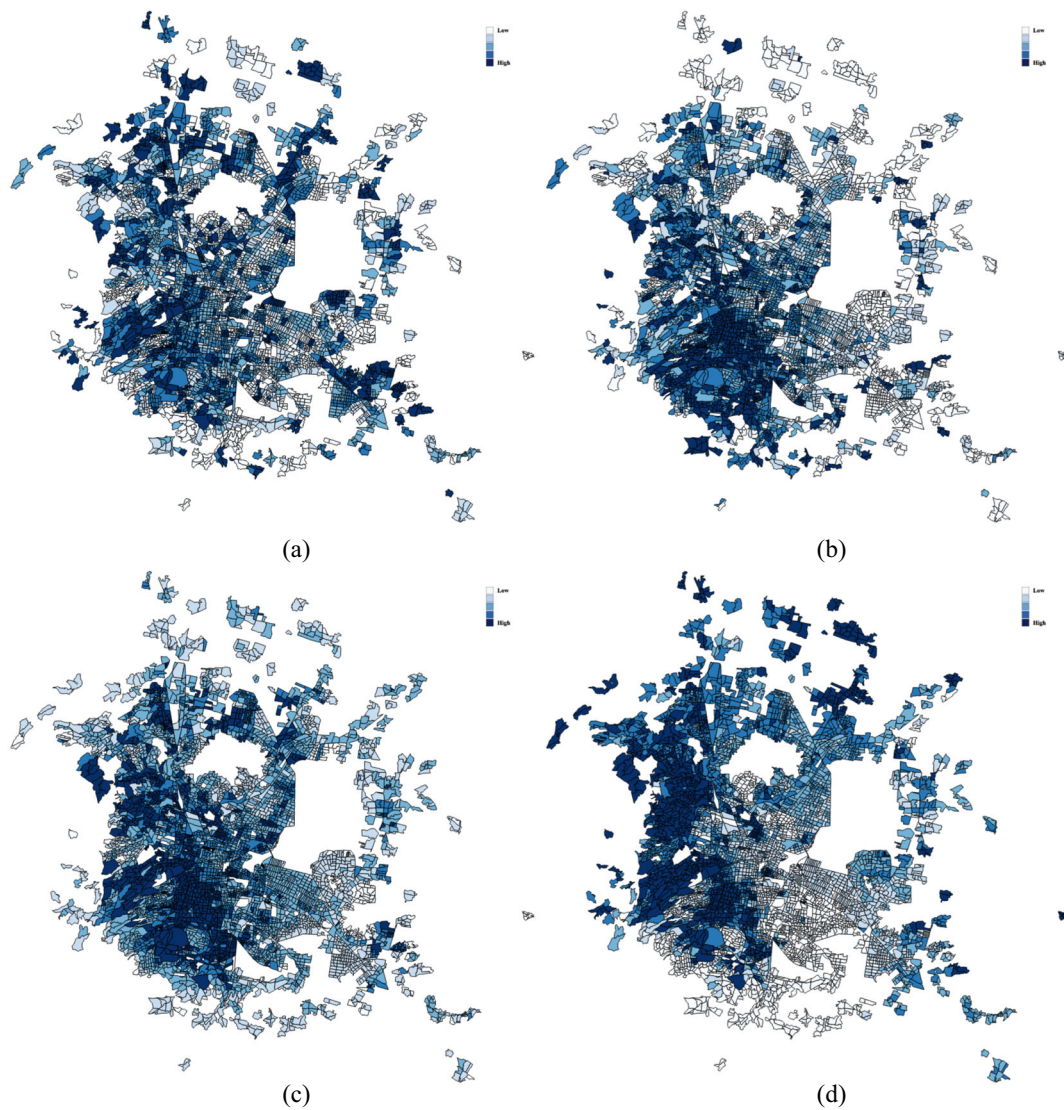
Source: (a–b) Authors' analysis from credit card transaction data. (c) National Population Council (CONAPO) of Mexico. (d) National Electoral Institute (INE) of Mexico.

economic characteristics that a label such as political "conservatism" might have in two different contexts. Figure 1 illustrates the two diversity measures, socioeconomic status, and political conservatism, for districts in the Latin American city.[6] The visual patterns are consistent with the quantitative results in Table 1.

The positive correlation between district-level purchase diversity and socioeconomic status is consistent with the findings in Eagle, Macy, and Claxton (2010) and Smith, Mashhadi, and Capra (2013), where the authors have shown that the diversity of phone communication of local districts is strongly correlated with the socioeconomic index of these districts. This confirms the link between explorative behavior and regional prosperity. Unlike phone communication, however, the present results suggest that exploration could be directly attributed to economic activities, hence explaining such link more explicitly from an economic perspective.

In contrast, the difference between the two countries, in terms of the relationship between purchase diversity and political conservatism, highlights the different meanings conservatism may have in different political contexts. Nevertheless, the presence of these correlations, either positive or negative, suggests that different purchase patterns indicate a separation between the districts in terms of their political views.

Although the correlation-based analysis does not permit causal inference, these results show that purchase patterns at least serve as strong statistical indicators of the socioeconomic status and political views of the districts. This would potentially enable the understanding and prediction of the evolution of such important societal indices, from only purchase behavioral patterns, in a cost-effective and timely fashion.

Finally, both diversity measures, and in particular the incoming purchase diversity, capture the richness of social and economic information that is introduced to each district. The authors believe that such richness of information may have an attentuating effect on any extreme economic and political behavior. It is hypothesized that districts (or cities), which attract economic activities from various regions, tend to be less politically extreme, show high voting diversity, and experience lower economic segregation. The researchers plan to perform such analysis as future work.

References

Blumenstock, J., G. Cadamuro, and R. On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–76.

Cassel, G. 1967. *The Theory of Social Economy*. New York, NY: Augustus M. Kelley.

Eagle, N., M. Macy, and R. Claxton. 2010. "Network Diversity and Economic Development." *Science* 328 (5981): 1029–31.

Lazer, D., A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne. 2009. "Computational Social Science." *Science* 323 (5915): 721–23.

Marshall, A. 1890. *Principles of Economics*. London: Macmillan and Co. Ltd.

National Electoral Institute. 2013. "Atlas de Resultados Electorales Federales 1991–2012." http://siceef.ife.org.mx/pef2012/SICEEF2012.html.

National Population Council (CONAPO). 2015. "Índice de Marginación Urbana 2010." http://www.conapo.gob.mx/es/CONAPO/Indice_de_marginacion_urbana_2010.

Smith, C., A. Mashhadi, and L. Capra. 2013. "Ubiquitous Sensing for Mapping Poverty in Developing Countries." Paper presented at the NetMob, Cambridge, MA, May 1−3, 2013.

Toole, J. L., Y.-R. Lin, E. Muehlegger, D. Shoag, M. C. González, and D. Lazer. 2015. "Tracking Employment Shocks Using Mobile Phone Data." *Journal of the Royal Society Interface* 12, doi:10.1098/rsif.2015.0185.

Turkish Statistical Institute. 2012. "General Election of Representatives Province and District Results." http://www.turkstat.gov.tr/Kitap.do?metod=KitapDetay&KT_ID=12&KITAP_ID=145.

VisioThink, Inc. 2013. "Turkey Socio-Economic Status Database - version Quarter 4 2012." http://www.visiothink.com/en/products/ses-data.

6 The holes in the maps correspond to nonresidential areas such as forests and farms.