Computational Statistics and Data Analysis 58 (2013) 292–307

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Spatial interaction models with individual-level data for explaining labor flows and developing local labor markets

A. Chakraborty [a,*], M.A. Beamonte [c], A.E. Gelfand [b], M.P. Alonso [d], P. Gargallo [c], M. Salvador [c]

[a] Department of Statistics, Texas A & M University, College Station, TX, USA
[b] Department of Statistical Science, Duke University, Durham, NC, USA
[c] Facultad de Economía y Empresa, Universidad de Zaragoza, Spain
[d] Facultad de Letras, Universidad de Lleida, Spain

ARTICLE INFO

Article history:
Received 28 October 2011
Received in revised form 27 May 2012
Accepted 24 August 2012
Available online 5 September 2012

Keywords:
Approximate block diagonalization
Bayesian hierarchical modeling
Conditionally autoregressive models
Markov chain Monte Carlo
Origin–destination models
Spatial partitioning

ABSTRACT

As a result of increased mobility patterns of workers, explaining labor flows and partitioning regions into local labor markets (LLMs) have become important economic issues. For the former, it is useful to understand jointly where individuals live and where they work. For the latter, such markets attempt to delineate regions with a high proportion of workers both living and working. To address these questions, we separate the problem into two stages. First, we introduce a stochastic modeling approach using a hierarchical spatial interaction specification at the individual level, incorporating individual-level covariates, origin (O) and destination (D) covariates, and spatial structure. We fit the model within a Bayesian framework. Such modeling enables posterior inference regarding the importance of these components as well as the O–D matrix of flows. Nested model comparison is available as well. For computational convenience, we start with a minimum market configuration (MMC) upon which our model is overlaid. At the second stage, after model fitting and inference, we turn to LLM creation. We introduce a utility with regard to the performance of an LLM partition and, with posterior samples, we can obtain the posterior distribution of the utility for any given LLM specification which we view as a partition of the MMC. We further provide an explicit algorithm to obtain good partitions according to this utility, employing these posterior distributions. However, the space of potential market partitions is huge and we discuss challenges regarding selection of the number of markets and comparison of partitions using this utility. Our approach is illustrated using a rich dataset for the region of Aragón in Spain. In particular, we analyze the full dataset and also a sample. Future data collection will arise as samples of the working population so assessing population level inference from the sample is useful.

1. Introduction

Population mobility plays a key role both in the performance of an economic system and in the daily life of individuals. Commuting is increasing in absolute volume and in number of destinations. So, it has become a key element in defining organization with regard to economic geography (Ball, 1980; Simpson, 1992; Henley, 1998; Kaufmann, 2000).

Urban factors such as housing availability together with expanded automobile and public transport use, encourage the growth of commuting and result in consequential discrepancies between where individuals work and where they

* Corresponding author.
E-mail addresses: avishekc@stat.tamu.edu (A. Chakraborty), asunbea@unizar.es (M.A. Beamonte), alan@stat.duke.edu (A.E. Gelfand), p.alonso@geosoc.udl.cat (M.P. Alonso), pigarga@unizar.es (P. Gargallo), salvador@unizar.es (M. Salvador).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.08.016

live (Jansen, 1993). Hence, it becomes useful to explain these daily mobility patterns. A frequent approach for dealing with this problem is to analyze the displacement flow matrices using spatial interaction models, see, e.g., the review paper of Roy and Thill (2003). These models seek to describe the processes by which entities located in different locations interact with each other, either for migratory movements, labor displacement or other reasons. At the individual level, they reflect personal cost–benefit decisions associated with the displacement. The most commonly used models include the so-called gravitational models (Alonso, 1978; De Vries et al., 2001, 2002, 2009; Ding and O'Kelly, 2008; Fotheringham, 1983; Hua, 2001; Roy and Thill, 2003; Wilson, 1967) that try to explain flows observed through origin and destination explanatory variables. Early work introduced model fitting through utility maximization, with connections to entropy and likelihood maximization (see, e.g., Wilson (1967) and the review paper of Wilson (1975)). More recent work implements model fitting through least squares (De Vries et al., 2009; Ding and O'Kelly, 2008; Sen and Sööt, 1981) or the use of instrumental variables (De Vries et al., 2002).

In addition, these mobility patterns suggest formation of functional economic areas with strong internal flows of commuting. This phenomenon was originally conceived in terms of big urban agglomerations but is now applied to finer scales (Giuliano and Gillespie, 1997; Lowe, 1998). Furthermore, customary administrative demarcations do not capture the limit of space in the daily life of populations (Gaussier et al., 2003; Van Ham et al., 2001). Rather, the flows of daily mobility have become a primary identifier of territorial aspects and can contribute to better articulation of public policies with regard to land management, transport, housing, economic activity and work (Amedeo, 1969; Cörvers et al., 2009). Unfortunately, administrative areas are often used as a surrogate for labor market areas in terms of statistical, analytical, and policy-making purposes though, again, these areas usually do not reflect functional reality (Ball, 1980; Coombes, 2002; Smart, 1974) and may compromise the effectiveness of resulting policies (Coombes et al., 1986).

The dominant concept in defining functional regions is that of local labor markets (LLMs) (see, for instance Goodman, 1970, Smart, 1974, Coombes et al., 1986, Tolbert and Killian, 1987 and Coombes, 1992). See also Casado Díaz and Coombes (2011) for a full critical review. LLMs identify the areas within which there is a close relationship between labor supply and demand. Qualitatively, such a market is characterized as an area in which a large proportion of the workers both live and work.

Some early approaches based on numerical taxonomy principles and statistical objective were proposed in Brown and Holmes (1971), Masser and Brown (1975), Fischer (1980), Masser and Scheurwater (1980) and Baumann et al. (1983). Brown and Holmes (1971) make a distinction between functional and nodal regions. They apply a Markov chain analysis to the interaction commuting flow matrix that transforms the matrix between basic spatial units (BSUs) into a mean first passage time matrix (MFPT). Those regions are delineated through hierarchical and non-hierarchical clustering techniques, applied to a distance matrix built from the MFPT. Masser and Brown (1975) describe two algorithmic clusterings based on flux. They use separate procedures for aggregating BSUs called Intramax and Intramin. The former defines subsystems to maximize the proportion of total interaction within the BSUs which do not cross the borders. The latter maximizes the proportion of border crossings between units.

A more recent methodology to construct LLMs is the algorithm developed by Coombes et al. (1986) which was accepted by the UK Department of Employment to produce travel-to-work areas (TTWA), based on 1981 Census data (see also Coombes, 1992). Subsequently, this methodology has been adopted, with minor modifications, by many countries including Italy (Sforzi et al., 1991), Spain (Alonso et al., 2008; Casado-Díaz, 2000), New Zealand (Newell and Papps, 2002), Denmark (Andersen, 2002), and Australia (Watts, 2004).

This methodology uses the daily labor *matrix of flows* between a collection of BSUs, usually districts, counties or municipalities. LLMs are created by attempting to maximize the interaction level of each BSU within its LLM, subject to restrictions on the *self-containment* level and on the size of each LLM. A sophisticated version of this methodology is due to Flórez-Revuelta et al. (2008a; 2008b), who view the problem as one of optimization and propose a genetic algorithm to achieve a solution.

A key remark is that none of these algorithmic approaches can obtain a ''best''solution. The number of partitions of a collection of administrative units into potential LLMs for a region is enormous. With any criterion, there will be a very large number of local optima. Any solution will, at best, be one of these. Furthermore, these methodologies lack explicit probability modeling, precluding uncertainty in comparing partitions.

Our contribution is to propose, using individual-level data, Bayesian hierarchical Poisson spatial interaction models for joint origin–destination modeling. These models employ spatial random effects and regression coefficients that allow us to incorporate an origin function, a destination function, a worker attributes component, and spatial structure to explain the variation in the origin–destination flows. To facilitate computation and search, over the partitioned space, the model is applied to a minimum market configuration (MMC), as described at the end of Section 2. The Bayesian approach lets us make a comparison between nested and non-nested models with regard to the explanation of the observed flows. It enables us to make an exact inference about the model parameters given the data we have observed, without relying on asymptotic theory (Banerjee et al., 2004). The model allows learning regarding systematic population mobility patterns. It allows comparison of patterns across varying individual-level characteristics. In fact, we obtain full posterior inference regarding the matrix of flows.

Once the model is chosen, we turn to delineation of local labor markets. In fact, by introducing worker attributes, we can work at the sub-group level, e.g., delineate LLMs for males vs. females or for different age groups. In any event, a specification of a map providing LLMs can be viewed as a partition of the set of units in the MMC. As noted above, the number