STATE OF CALIFORNIA • DEPARTMENT OF TRANSPORTATION
TECHNICAL REPORT DOCUMENTATION PAGE
TR0003 (REV 10/98)

1. REPORT NUMBER	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
CA-16-2851		

4. TITLE AND SUBTITLE
Business Establishment Survival and Transportation Level of Service

5. REPORT DATE
05/15/2016

6. PERFORMING ORGANIZATION CODE

7. AUTHOR
Konstadinos Goulias

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of California Transportation Center (UCTC)
University of California, Santa Barbara
Department of Geography - 1832 Ellison Hall
Santa Barbara, CA 93106-4060

10. WORK UNIT NUMBER
3762

11. CONTRACT OR GRANT NUMBER
65A0528 TO 010

12. SPONSORING AGENCY AND ADDRESS
California Department of Transportation
Division of Research, Innovation and Systems Information (DRISI)
MS-83, PO Box 942873
Sacramento, CA 94273-0001

13. TYPE OF REPORT AND PERIOD COVERED
Final Report

14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

16. ABSTRACT
In this project we seek to fill a gap in empirically supported knowledge linking the survival and economic success of business establishments to local land use and access to the transportation system that serves these establishments. We investigate this relationship for the entire State of California over the last two decades while controlling in a statistically robust way for a variety of factors influencing business life cycle events, such as closures, formation/birth, and relocation. We accomplish this by combining longitudinal business establishment population event data, various transportation access and level of service indicators, and geographical market size from available US Census data. In addition, we narrow our analysis to a specific year and region in order to investigate a broader range of industries and utilize a detailed accessibility dataset for Southern California.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
Transportation, Business Establishment, Level of Service	No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161

19. SECURITY CLASSIFICATION *(of this report)*	20. NUMBER OF PAGES	21. COST OF REPORT CHARGED
Unclassified	66	None.

DISCLAIMER STATEMENT

Business Establishment Survival and Transportation Level of Service

FINAL REPORT

Adam W. Davis, Jae Hyun Lee, Elizabeth McBride,
Srinath Ravulaparthy, and Konstadinos G. Goulias

University of California Santa Barbara
Department of Geography
Geotrans Laboratory

Contract Number: 65A0528.
Principal Investigator: Konstadinos G. Goulias;
Lead Researcher: Adam W. Davis
Support Researchers: Jae Hyun Lee & Elizabeth McBride
Research Affiliate: Dr. Srinath Ravulaparthy

May 15, 2016
Santa Barbara, CA
GEOTRANS REPORT 2016-06-01

Table of Contents

Introduction and Literature

In this project we seek to fill a gap in empirically supported knowledge linking the survival and economic success of business establishments to local land use and access to the transportation system that serves these establishments. We investigate this relationship for the entire State of California over the last two decades while controlling in a statistically robust way for a variety of factors influencing business life cycle events, such as closures, formation/birth, and relocation. We accomplish this by combining longitudinal business establishment population event data, various transportation access and level of service indicators, and geographical market size from available US Census data. In addition, we narrow our analysis to a specific year and region in order to investigate a broader range of industries and utilize a detailed accessibility dataset for Southern California.

The spatial distribution of economic activities has a profound impact on urban organization and development. Business establishments provide services and employment opportunities that influence the locational patterns of households, the behavior of individuals seeking to purchase goods and services, and the revenues of local jurisdictions. The spatial distribution of existing business establishments also affects location decisions of other business establishments and impacts the regional transportation network in the form of accessibility, traffic circulation, and possible congestion. In this context, integrated models of land-use and transportation are used to further analyze the impacts of these changes in regional planning and policy. Increasingly among researchers and practitioners there is a heavy push towards more disaggregate modeling of these integrated systems. This type of modeling is based on the behaviors of relevant market agents such as households, persons, business establishments and land developers that make decisions regarding their locations as well as personal travel and the movement of goods and services (Waddell et al. 2007; Strauch et al. 2005; Miller et al. 2010; Hunt and Abraham 2005).

One of the most critical dimensions of these systems is the spatial distribution of economic activities, which is significantly determined by geographical movements of business establishments, together with business establishment formation and expansion, decline and closure. An approach to study these processes is to track the lifecycle events of business establishments in a region and then develop simulation software that replicates the evolution of these business establishments while associating this evolution with urban planning policies (van Wissen 2000; Hunt et al. 2003; Elegar and Miller 2006; Kumar and Kockelman 2008; Moeckl 2009; de Bok and Oort 2011; Maoh and Kanaroglou 2013). A fundamental aim of these model systems is to accurately describe the triggers underlying the dynamics in the spatial distribution of economic activities in a region. Therefore, to better represent business establishment dynamics the spatial environment and the location of these business establishments and their evolution process must be explicitly accounted for and depicted by these modeling efforts.

A few studies in this context have recognized the importance of the spatial dimension of business establishment represented as demographic processes (de Bok and Oort 2011; Manzato et al. 2010a, 2010b; Maoh and Kanaroglou 2009; Maoh and Kanaroglou 2013) and analyzed the spatial characteristics in the form of accessibility to infrastructure, agglomeration economies and regional effects and their impacts on business establishment decision making behavior. However, these studies address a limited selection of business establishment lifecycle events. Manzato et al. (2010a, 2010b) includes space in modeling only business establishment survival rates, while, Maoh and Kanaroglou (2009, 2013) model the spatial dimension of both business establishment migration and business establishment dissolution. In addition, these studies simplify the spatial representation framework in their modeling efforts, thereby ignoring important aspects of spatial effects (e.g., relative location in the region) on business establishment lifecycle events.

Ravulaparthy rectified some of these issues and thoroughly investigated locational impacts on business establishment lifecycle events by explicitly representing the relative importance of a location in the network across multiple spatial scales through the notion of roadway network centrality indicators. For this purpose, he examined the quality and locational advantage of a business establishment in terms of its closeness to other businesses, intermediacy in the paths of travelers, straightness of the paths along which each business is located and the reach of other businesses by also accounting for the configuration and connectivity of the regional transportation network. These indicators of centrality, were proven to be significant factors in business establishment financial success, the probabilities of relocating, dissolving, and in the creation of new business establishments. This analysis showed centrality to be important even when controlling for the fundamental linkages of business establishment internal and external factors on business establishment relocation, dissolution and formation. Moreover, the influence of external factors to each business establishment depends heavily on the type of business and its locational needs (e.g., proximity to a freeway for a manufacturing firm versus a restaurant). To keep the analysis feasible Ravulaparthy and Goulias (2014) performed this research exclusively with data from Santa Barbara County.

The basic data ingredients for the methods presented in this report are: a) the longitudinal record of all business establishments in California in the NETS database; b) highway and other transportation infrastructure locations for the entire State of California; c) US Census data population counts at the block level; and d) fine-grained accessibility data for the Los Angeles area. Merging and fusing data from different sources presents some challenges, which are solved by using business establishment locations as the fixed points of analysis.

The basic methodological ingredients for this project are adapted from the work of Ravulaparthy and Goulias (2014) in a Santa Barbara case study estimating regression models for the probability of birth, dissolution, and relocation. New methods are presented here to extract land use and competition metrics from the business establishment data and handle the massive amount of data for the entire State of California.

The analytical methods used in this project are described in the summary below.

Table 1 Analytical Methods Used in this Project

Stage / Section	**Methods Employed**	**Output**
Data Processing / Spatial Variable Creation	Kernel Density Estimation, distance raster generation, and raster-to-points extraction.	Long-format (one row per year) dataset of business establishment events, including internal and environmental variables.
Panel Models	Panel binary probit models.	Estimates of the effects of various external and internal effects on the likelihood of firm birth/death/relocation for California retail and manufacturing firms.
Relocation within California	Means comparisons and spline smoothing.	Comparison of relocation origin and destination sites for California retail and manufacturing firms. Visualization of changing relocation preferences over time.
Southern California Case Study	Cross-sectional binary probit models.	Estimates of the effects of various external and internal effects (including detailed transportation network accessibility) on the likelihood of firm birth/death for retail, manufacturing, health care, and professional services firms in Southern California, 2008.

The key contribution of this research study is in thoroughly operationalizing and unifying the internal, regional, and locational factors that affect business establishment survival to more conclusively identify the role transportation system and land use in economic development and business establishment success. We seek to answer these questions:

1. **In what ways does access to transportation infrastructure affect the success, failure, and relocation of businesses?**
2. **In what ways does local land use affect the success, failure, and relocation of businesses? Which types of mixed-use environments are beneficial to business establishment success?**
3. **How do the effects of land use and transportation accessibility vary across different business types?**
4. **How are the factors that predict the formation of new business establishments relate to the factors that predict the dissolution of businesses?**

Data Overview and Processing

This project investigates the success and failure of business establishments as a function of attributes of their local environment. To do this, we bring together a range of data sources, including a comprehensive record of business establishment life history events, major infrastructure locations, and census data. In this section we describe the steps we took to process this data and merge it into a format usable in modeling. Figure 1 shows an overview of our data processing steps.

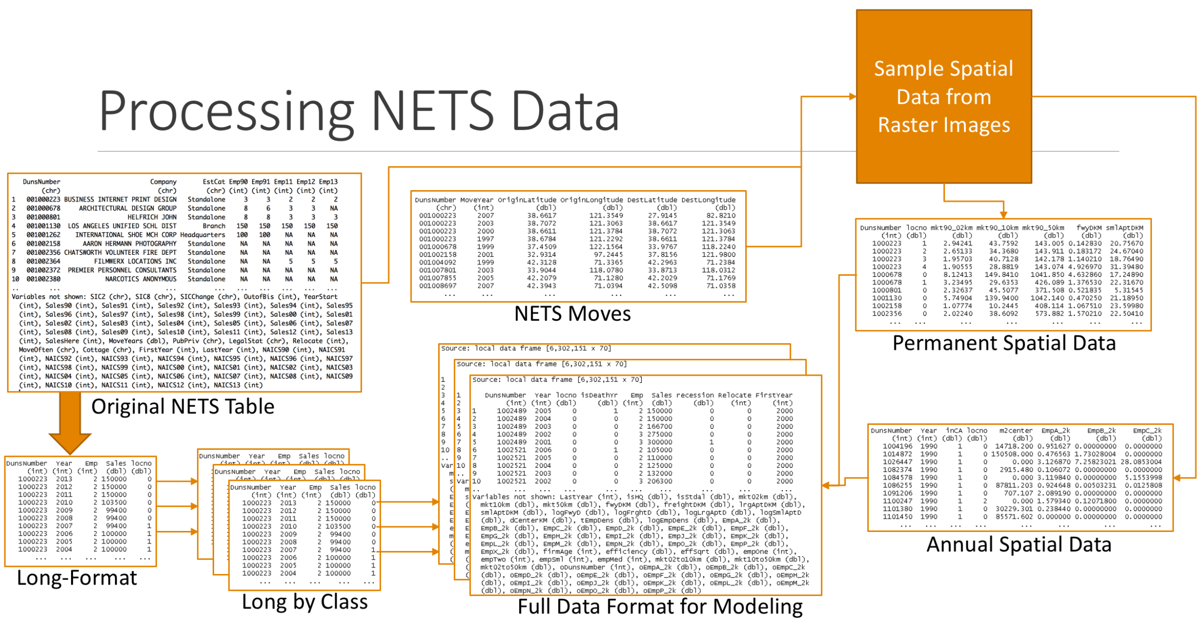

Figure 1 NETS Data Processing Overview

NETS Data Structure

The main data source for this project is the 2013 NETS database. This database contains geo-coded firm-level records for 6.7 million business establishments in California with longitudinal information about their industrial type, location, headquarters and performance over the period of 1990-2013. The NETS database is constructed by taking a series of 'snapshots' based on the Dun and Bradstreet archival national establishment data (Walls 2007). The unit of observation in the NETS database is a business establishment that produces goods or services at a single physical location – for example, a single store. This database tracks every establishment from its formation to its dissolution, through any physical moves it makes, capturing any changes in ownership and business type. NETS records information on location of the establishment, employment, sales and industry type for each year.

The NETS dataset is delivered in a tabular format, with each row containing the permanent characteristics and year-by-year employee counts and sales for a single business establishment with a unique DUNS number. The general overview of the steps we take to process the data is shown in Figure 1 displaying major steps and associated tables. Additional tables contain business categories, providing a 6-digit North American Industrial Classification System (NAICS) code for each year the business existed, and a record of relocation events. To extract the relevant information for this application, we perform two major processing tasks performed primarily with the packages Dplyr and Tidyr in the statistical language R: A) converting the "wide" source table to a "long" format table separated by class (the steps counterclockwise from top left to bottom center in the figure) and B) extracting and updating location data and extracting spatial variables for the datasets (steps clockwise from top left to bottom center in the figure). The outputs of these processes are joined into a set of final tables for each establishment class that contain annual records of internal and environmental characteristics for all the businesses in that class.

Process A: The "wide" format source table contains an individual column for each year's sales, employees, and business classification data (e.g. "Sales90" and "Emps97"). The first step of this process is to stack each set of these columns into a single column that contains annual records for the given business establishment. The long-format table is then filtered to eliminate rows for years in which a given business establishment did not exist. For example, a business establishment that formed in 1993 and dissolved in 2000 would be represented by a single row in the source table, with blank or zero values for employees, sales, and category in each column from 1990 to 1992 and from 2001 onward; in the long-format table, this business would be represented by eight consecutive rows containing information for 1993, 1994, 1995, and so forth. Because the resulting long-format table contains roughly 50 million rows and the models presented in this project are calculated independently for each industrial category, the full long-format table is partitioned by class, producing 16 tables of varying length. Parallel processing was used to accelerate the process of separating businesses by NAICS classification. Table 2 lists all business classes we used, along with the number of unique businesses and total observations for each class.

Table 2 Business Establishment Categories

Category	**2-digit NAICS**	**Establishments (Unique DUNS)**	**Observations (DUNS:Years)**
Agriculture, Forestry, Fishing and Hunting	11	78,990	728,397
Mining	21	6,676	55,160
Utilities	22	7,061	66,747
Construction	23	550,750	3,862,499
Manufacturing	31, 32, 33	343,107	2,916,915
Wholesale Trade	42	431,344	3,005,409
Retail Trade	44, 45	928,089	6,379,413
Transportation and Warehousing	48, 49	169,354	1,111,009
Information	51	212,048	1,337,209
Professional, Scientific, Management, Administrative and Waste Services	54, 55, 56	1,975,382	12,129,165
Health Care	62	466,341	3,700,886
Arts, Entertainment, Recreation, Accommodation and Food Services	71, 72	346,838	2,871,635
Other Services	81	699,929	5,259,591
Finance, Insurance, Real Estate and Rental and Leasing [FIRE]	52, 53	691,557	4,921,333
Public Administration and Armed Forces	92	29,665	247,909
Educational Services	61	84,159	787,201
Undefined	99	9,344	45,816

Process B: This chain addresses business establishment locations and extracts the various attributes of their environments that affect the success or failure of business establishments.

Locational features are the main external attributes used in our models, so it is important to use precise business establishment locations updated for each year to extract this information.

The first step is to match businesses to their correct location for each year and export a list of all unique locations that we can use to extract spatial information. To do this, the final locations in the main NETS table are combined with locations from the NETS moves table to create a record of successive locations (DUNS, location number, latitude, longitude). In each business move record, NETS provides both an origin and a destination location; to provide complete coverage of the dataset, we use origin locations. Destinations locations match the origin location of the next move or the business's final location more than 98% of the time, and errors are generally small in magnitude. Because many businesses relocated multiple times over the study period, we join the move table to the long-format business table to determine whether the business moved in a given year. We then ensure records are grouped by business and sorted by year, then perform a cumulative window-sum of number of moves to determine which location number to use for a given business in a given year (e.g. if a business moved in 1994 and 1999, then its third location will be used starting in 2000). Unique business locations are then exported into an ArcGIS point shapefile and used to extract permanent spatial information about the local business environment. Appropriate locations with annual employment counts are also exported to calculate land use densities. More detail on the spatial data extraction process is provided in the next two subsections.

Merge Process: Finally, the spatial variable tables are joined to the long-format tables to produce a full annual record of the internal and external factors hypothesized to affect business establishments in each category in each year. Once these tables are produced, a number of other variables are calculated for use in the final models, namely a variable indicating whether the United States economy was in recession for at least one quarter in the given year (1990, 1991, 2001, 2008, and 2009), firm age, age squared, and age cubed (this enables models to incorporate a polynomial approximation of the effects of business age on firm success). An approximate measure of business efficiency is calculated by dividing a business's sales by its employees (this variable is divided by 10,000 to bring coefficients more in line with those of our other variables). Size categories were calculated from employee counts to address the nonlinear impact of business size on business success. Specific cutoffs used were chosen by consulting the histograms of business size: one employee, two employees, three to eight employees, nine to twenty employees, and more than twenty employees; in each model, the reference case is Large Establishments (>20 employees). Business establishment formation, dissolution, and relocation years are confirmed by checking the FirstYear and LastYear columns in the original dataset and looking for zero values in employee counts.

For all spatial variable processing steps, it is important to keep all raster datasets snapped to the same grid. This makes the extraction process more efficient and makes it possible to check data quality in a consistent way and produce combined raster surfaces from model outputs.

Activity Density / Land Use Surface Estimation

Local land use is likely to play a major role in the success or failure of business establishments. Business establishment data can be used to compute multiple possible measures of land use and it is likely impossible to represent all aspects of land use with a single variable. Diverse, high density environments foster agglomeration economies that help some businesses thrive, but they also likely drive up real estate rents, pushing out less successful businesses. The presence of related businesses may result in collaboration or intense competition. To measure land use, we use employee density in multiple business categories, which enables us to investigate both the direct effects of density and the effects of collaboration and competition. A summary view of the US Bureau of Economic Analysis Input-Output tables (U.S. Bureau of Economic Analysis) shows that the direct financial relationships between groups of businesses are significant and diverse (see Figure 2); some of these interactions likely require spatial proximity. The variation of land use preferences among different business classes is of particular interest in this report.

For this analysis, we must convert business establishment / employment count points to a map of activity density measured consistently across the entire state. Land use is an areal property so it can be modeled as either a continuous surface (raster) or a set of bounded units (vector polygons). Because business locations were stored as point features, they must be converted into one of these formats to be useable for land use estimation. The most straightforward method would be to choose a single polygon scale of aggregation such as zip codes or block groups and sum up the employees of a given business category located within each polygon, but this process has one main shortcoming in this application: it produces considerable edge effects that decrease its accuracy when values are extracted to business locations. Edge effects become a problem when a business is located near the border between two zones; by simply aggregating to containing polygons, this business would be counted exclusively towards one, even though it should relate almost equally to the land uses of both. In high density areas, the simple aggregation process may underrepresent the density of areas on the edge of a dense business zone and in areas where a small area of lower employment density doesn't represent an actual change in local land use over space. These problems are particularly important because census polygons are designed to equalize population at home locations, not the locations of businesses.

Name	Agriculture, forestry, fishing, and hunting	Mining	Utilities	Construction	Manufacturing	Wholesale trade	Retail trade	Transport. and warehousing	Information	Finance, insurance, real estate	Professional and business services	Education, health, and social assistance	Arts, ent., food services, and accommodation	Other services, except government	Government
Agriculture, forestry, fishing, and hunting	103993	105	0	1399	280940	1547	2244	108	1	35	1799	916	8214	82	2292
Mining	2520	58402	43079	10742	560306	46	49	1580	339	4297	1563	435	1286	592	16944
Utilities	4955	4430	3076	2371	66293	5092	11171	6625	4414	71743	10765	25415	11448	4378	27741
Construction	2884	5713	4749	158	16279	1577	2980	4683	2733	128332	2153	2970	2387	3884	72240
Manufacturing	87134	52838	25686	281090	2042570	45665	59431	198525	105380	61146	163507	190020	155655	63114	355582
Wholesale trade	26447	8437	4813	47406	300870	41132	25790	37865	29959	15763	28149	39472	26013	10642	43856
Retail trade	220	230	378	82613	13762	813	7650	5505	406	7181	2694	1687	7992	5409	595
Transport. and warehousing	14259	13205	18308	20930	160719	66389	71188	119692	19398	29061	53577	25000	16158	6490	61579
Information	588	1225	1438	4530	26382	18091	20467	5818	241694	59498	79696	30260	10221	10078	73521
Finance, insurance, real estate	21044	19499	10768	31460	85593	101708	147946	75213	68530	884455	246119	316319	93082	83370	134123
Professional and business services	6245	45803	19550	46229	445816	206347	166840	65236	149996	440548	566526	251468	141266	45125	249852
Education, health, and social assistance	495	0	148	20	85	1230	9106	75	927	50	817	27265	1975	2562	43850
Arts, ent., food services, and accommodation	666	1197	2656	2444	23416	9312	6376	3398	35975	48182	69327	31235	29291	4222	31454
Other services, except government	1156	734	737	4900	18173	18919	12264	5413	10735	33889	35044	36953	12107	7354	25069
Government	47	6	454	26	6026	10472	6640	14479	2358	15152	9282	6726	6355	1824	8437

Figure 2: USA Input Output Table (source US BEA).

Instead of relying on simple counts, we employ the kernel smoothing process implemented in ArcMap to estimate an activity/land use density surface from business establishment locations. Kernel density functions fit a smooth, curved surface over the input points (in this case businesses). Each point's contribution to the density surface is highest at its location and diminishes with increasing distance from the point, reaching zero at the distance from the point specified by the bandwidth/maximum distance parameter. ArcGIS uses a quartic kernel function to calculate the density. The total volume under each point's kernel density surface is equal to the point's population field value (in this case, the business's number of employees in a given year in NETS). The total density in the output raster is calculated by adding the values of all the kernel surfaces at the center of each raster cell (ESRI 2015a).

By smoothing employee/activity density over space, we seek to produce a more accurate representation of land use that can be used for statewide analysis. In addition to eliminating the issues described above, smoothing addresses the error caused by small inconsistencies in the precision/accuracy of business establishment coordinates provided in NETS, which are more accurate for newer business locations than for ones that have existed since 1990.

To produce a final activity density map, we tested a range of kernel bandwidths (from 200m to 20km) and chose a 2km bandwidth for the final product. This kernel balances the benefits of the detailed but irregular surfaces provided by smaller bandwidths and the smooth but overgeneralized surfaces produced by larger kernels. The choice of kernel bandwidth is highly dependent on the specific application of the density surface. In final analysis, the 2km bandwidth also seemed appropriate because the highest-density part of each point's kernel is contained within a reasonable size area for neighborhood scale analysis. Figure 3 shows the resulting density surface when aggregated to a block group level for display purposes.

This density generation process is repeated for each year for each of the 16 business categories. The point files used to generate kernel density surfaces are then used to extract density values for each business location. To increase processing speed, only locations active in a given year are used to extract densities for that year. The extracted annual density tables are grouped by business class and then joined with other datasets to produce the final table for modeling, as described above.

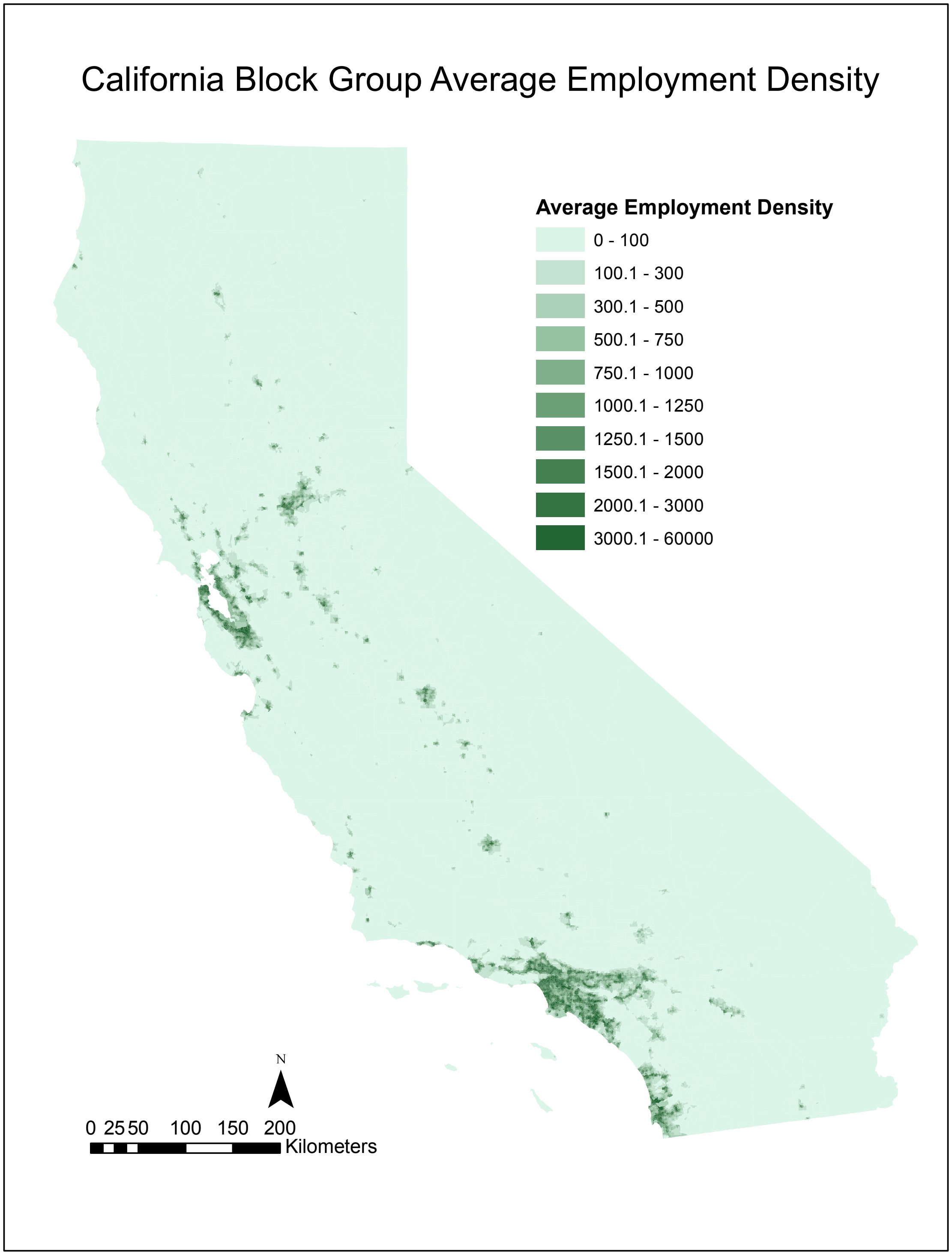

Figure 3 Average Employment Density at the US Census Block-Group Level in 2012

Other Spatial Variables

In addition to the land use information conveyed by employment density, we consider other aspects of the business environment that vary spatially. This section presents the methods used to generate two sets of business environment variables: market area and distance to infrastructure.

Access to customers and employees is extremely important to business function, so it is an essential component of survival models like those presented here. Population counts provide one measure of this important aspect of business, and in this project we use total population within three distance bands (2km, 10km, and 50km) to produce variables that can test for the effects of market area. Final models presented in this report use either total population within 50km (which accounts for overall market area) or total population within 2km and total population between 2 and 50km (the first for local density and the second for overall market area). Market area variables are produced using block-level population totals from the decadal census and interpolated for years in between. The creation of the American Community Survey in the mid-2000s presents a tradeoff between fine spatial resolution (census blocks provide very detailed information about the distribution of people in small areas) and better-than-decadal temporal resolution, but in order to maintain a consistent dataset, census block totals are used for all years.

To create the market area dataset, we download block level total population totals and boundary files for 1990, 2000, 2010 available from the National Historical GIS (Minnesota Population Center, 2011). Block polygons are converted to a raster grid with 50m pixels snapped to the grid used for density and distance datasets; almost all blocks occupy numerous pixels. Block population totals are divided by the number of pixels each block occupies, and the resulting density value (measured in people per raster cell) is assigned to the block raster, producing a statewide map of block-level average population density. To convert this into a count of people within a certain distance, the ArcGIS Focal Statistics tool is used to count total value of all cells within 2km, 10km, and 50km of each cell using a moving window technique (ESRI 2015b). Business establishment locations are then used to extract values from the nine market area raster files (three distance bands for each of three census years). Market area for a business location in a specific year between 1990 and 2010 is generated with a linear interpolation of the values from the two nearest census years (e.g., 2004 population is equal to 0.6 times the 2000 population plus 0.4 times the 2010 population). For 2011-2013, the annual population growth rate from the 2000s is projected forward. To make the resulting values more useful in a model, the 2km radius population totals are divided by 10,000 and the 50km radius and 2 to 50km annulus population totals are divided by 1,000,000.

This method of estimating market areas entails two notable shortcomings: population is not evenly distributed within each Census Block and population totals are unlikely to change over time in a strictly linear fashion. Unevenness of human activity and natural attributes over space can be substantial within larger polygons (and assuming that zone-wide values are fully representative of the entire region they cover is called ecological fallacy), but blocks are very small, so they capture population density quite accurately in areas with many residents. This method underestimates local density in the populated parts of very sparsely populated blocks, but the scale of the underestimation is quite small compared to the range of density statewide (on the order of one hundred people for variables that have ranges of tens of thousands to several million) and few businesses will be affected by the error, so the problem will not affect model results substantially.

Access to transportation infrastructure is an important factor for many businesses. For this project we use Euclidian distance to a number of key facilities, including freeways, small commercial airports, major airports, and freight/intermodal facilities (California Department of Transportation. Though these variables do not provide the level of detail we have in the accessibility datasets used in the Los Angeles case study, they provide useful general measures of the degree of access businesses have to transportation infrastructure. For each of these facilities, a raster dataset is produced with each cell containing the distance to the nearest feature of a given type. Business establishment locations are then used to extract values from these distance raster files; the extraction process uses bilinear interpolation to improve accuracy.

Target Classes

The analysis in this project focuses on four specific business types: retail, manufacturing, professional services, and healthcare. These businesses were chosen because they represent a range of business structures and are expected to prefer different types of environments.

Retail businesses are included in our analysis because retail establishments are generally very clearly delineated – each is an individual store. Retail businesses should respond to geographic variables because different locations provide different degrees of access to customers, real estate costs, and local competition. Retail businesses belong to NAICS categories 44 and 45. California retail businesses exhibited steady, but slow growth over the study period (Figure 4).

Manufacturing businesses were chosen because they are a classic focus of study in economic geography and they are likely to respond to different locational cues than retail firms. Like retail establishments, manufacturers require access to the transportation network – in their case to moving inputs and products rather than to bring customers – and may prefer less-dense environments than retail businesses do. Manufacturing businesses belong to NAICS categories 31, 32, and 33, and the total number of establishments was very stable over time (Figure 5)

Professional Services firms were chosen as a topic of study because they exhibit very different locational preferences to retail and manufacturing firms. Professional services firms, such as individual CPAs, lawyers, and consultants) are often smaller than other firms and many are based in people's homes. We anticipate finding very different locational preferences than the other businesses we are investigating, particularly with regard to the effects of transportation network accessibility, which we expect to be limited. These firms belong to NAICS classes 54, 55, and 56. There was a very substantial growth in professional services firms over the study period (Figure 6).

Health Care establishments were chosen as a topic of study because businesses in this category represent a mix of structures and should exhibit vastly varied locational preferences. Health care firms range in size from individual doctors who contract with hospitals to small clinics to very large hospitals. We expect these businesses to respond most weakly to general locational characteristics, in part because their locations were found to cluster strongly around major hospital locations. These businesses belong to NAICS category 62 and became three times as numerous between 1990 and 2013 (Figure 7)

In each of the figures that follow, total business establishments each year are shown by the blue line; establishment dissolutions are shown by the red line, and establishment formations are shown by the green line.

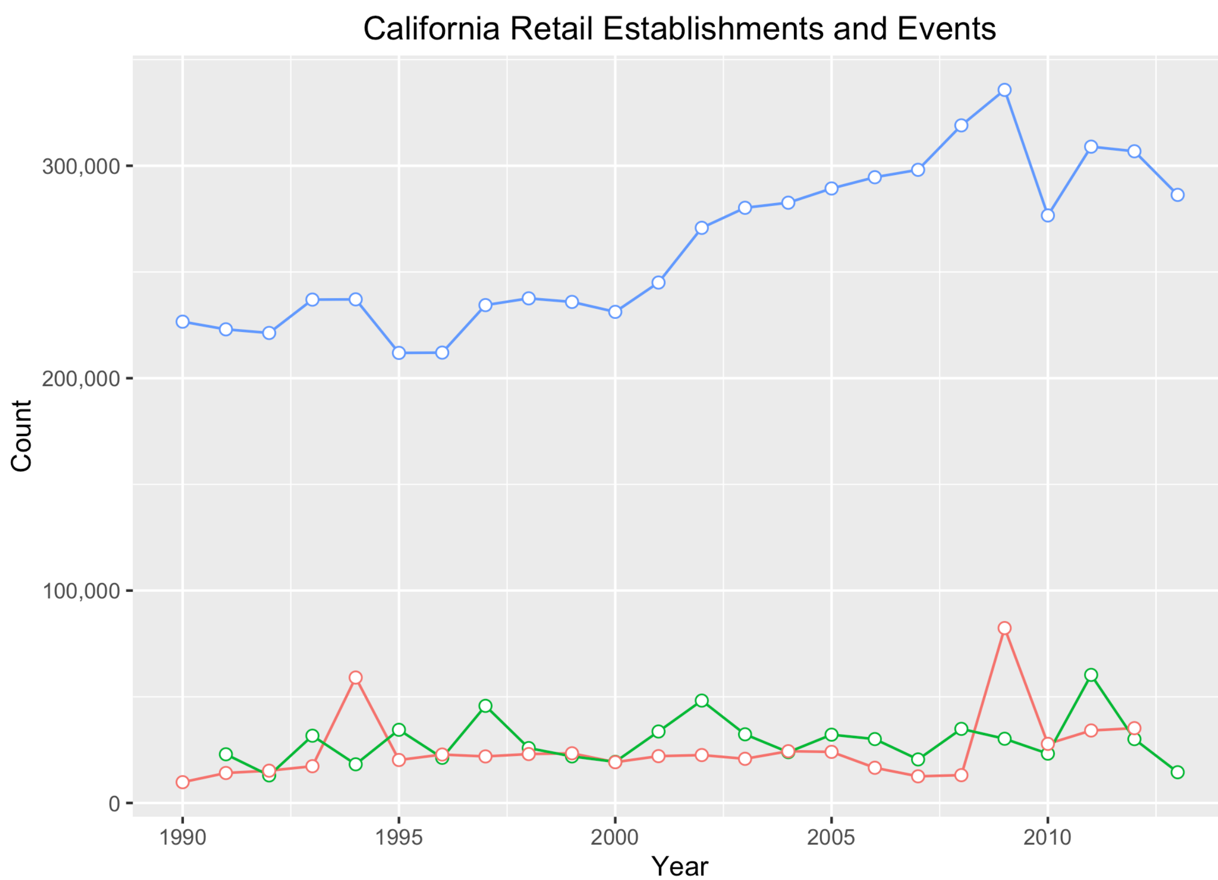

Figure 4 Retail Establishment Events

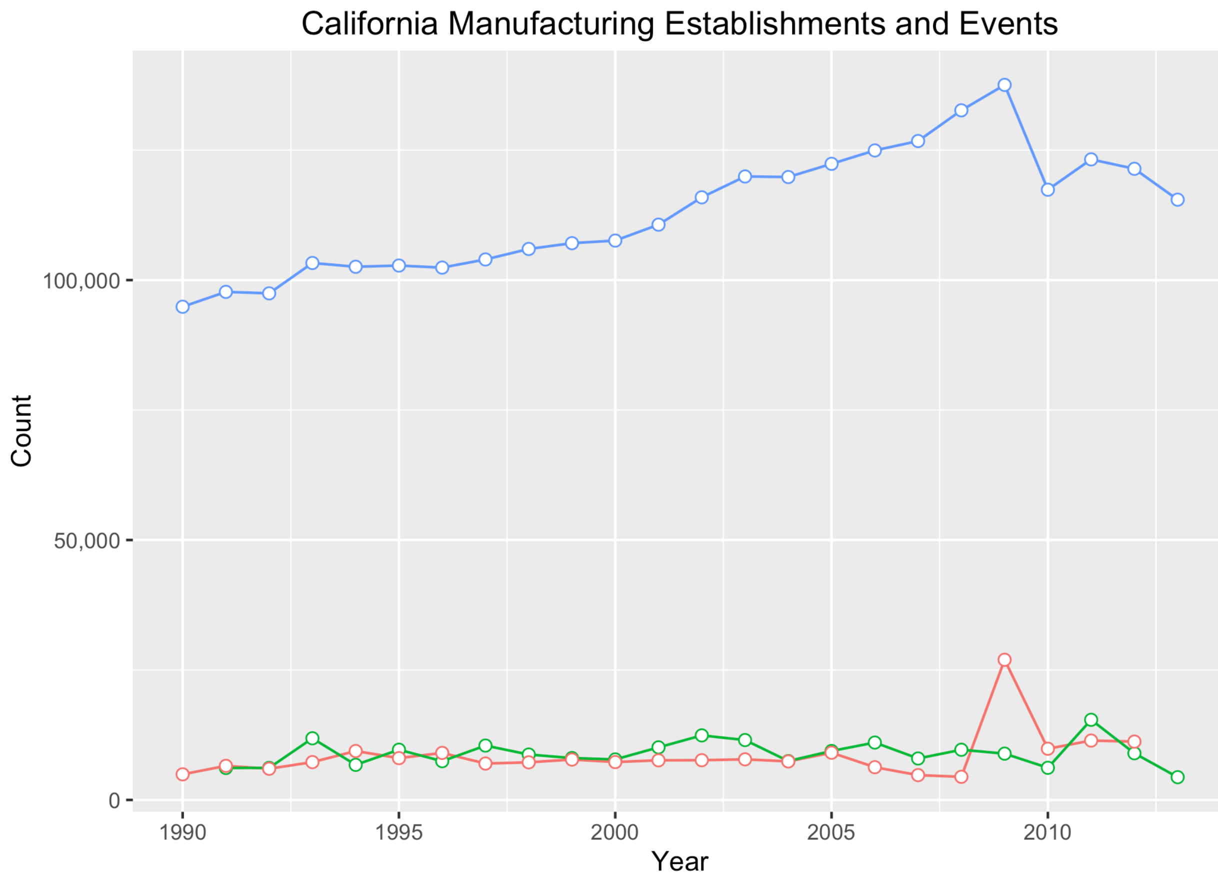

Figure 5 Manufacturing Establishment Events

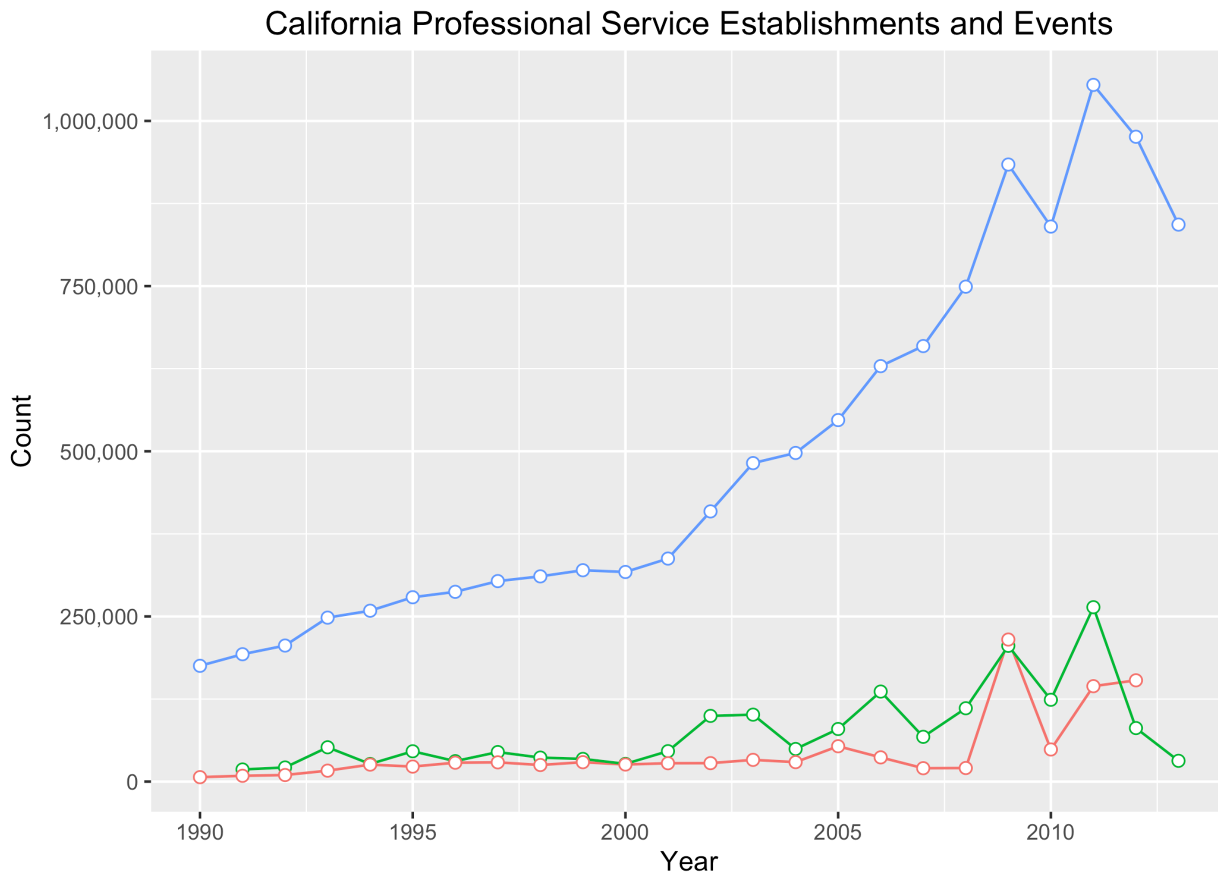

Figure 6 Professional Services Establishment Events

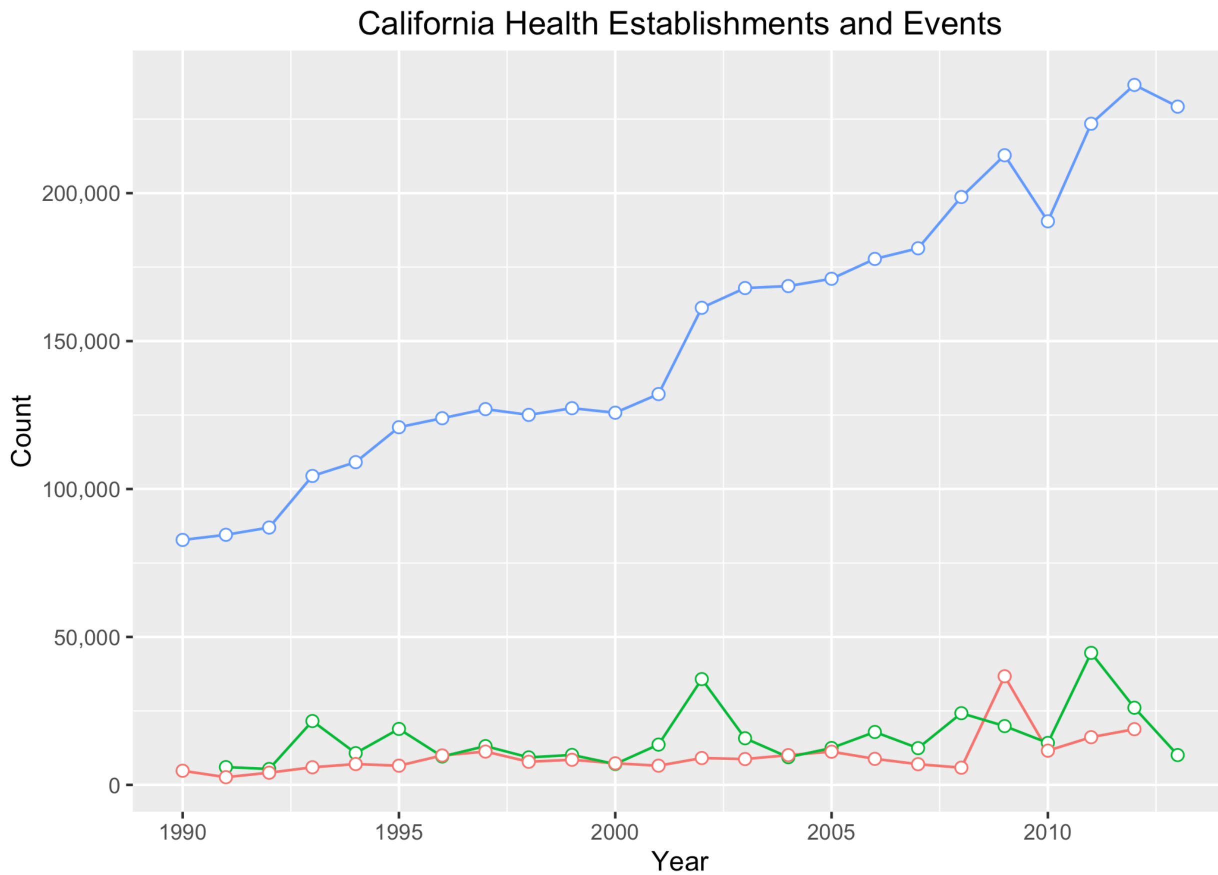

Figure 7 Health Care Establishment Events

Modeling Methods

While the reasons behind the success or failure of each individual firm are too varied and complex to fully capture in a general model, our framework does allow us to investigate the various attributes of business locations that make some places more favorable to businesses than others. Since these locational attributes are often directly subject to state and local planning decisions, they are particularly important to study. By analyzing how access to infrastructure, local land use, and local market areas affect the likelihood of individual businesses surviving or failing, we can gain a clearer picture of how to plan economically sustainable communities.

Probit regression is used to analyze the sensitivity of each event to a wide variety of internal and external variables of each business establishment. The reasoning for using probit regression and estimation details are provided in Greene (2003) and Ravulaparthy (2013) provides a detailed review of the background of pooled estimation in this context. The results are presented in the same format as in linear regression models. In terms of interpretation a positive significant coefficient indicates that the coefficient's associated variable is contributing positively to the occurrence of an event and negative indicates the opposite. A non-significant coefficient indicates no correlation between its associated variable and the event under study. We have two types of panel regression that are a panel regression to account for the repeated observation of the same business establishment and therefore the presence of the same establishment at multiple time points (panel model) and a cross-sectional probit regression for the case study in an area where detailed accessibility indicators are available.

The panel models in this study are estimated with random effects. These are essentially a firm-level error that contains unmeasured causes of variation between businesses that explains part of their success or failure. Random effects may be significant or insignificant depending on other variables included, but it’s important to consider variability between members of the panel as well as between all observations because multiple observations of the same business are not independent.

Our model development process is fairly similar between the panel and cross-sectional models. We start with a model that contains a large set of variables believed to be important, including most industry-specific densities. We then iteratively adjust the model, removing some non-significant variables at each step until we are left with a final model that is both parsimonious and sensible. Internal attributes included at the start are business size (categorical – 1, 2, 3-8, 9-20 employees included as dummy variables, with businesses with more than 21 employees serving as the reference class), natural log of sales, efficiency (sales per employee), and a few measures of firm age. The initial market characteristics included are population within 2km (divided by 10,000) and population between 2 and 50 km (divided by 1,000,000). When the model indicates that local density is not significant, we replace both variables with total population within 50km, again divided by 1 million. Log of kernel density for all 16 business category land uses were used in the initial testing for each class and insignificant variables are removed through an iterative process. Transportation accessibility variables tested include distance to airports, distance to freeway, and distance to intermodal processing locations.

Sales per employee is an imperfect measure of the efficiency of a business because it ignores the differences in labor and input costs experienced by different businesses, but it should be relatively consistent within a given business category. To improve the usefulness of this variable and remove extreme cases, our analysis excludes establishments that bring in more than $400,000 in revenue per employee per year; depending on the random sample taken, this eliminates roughly 1.6% of retail business establishment observations and 2.2% of manufacturing business establishment observations.

The panel models should only include variables for which we have data at multiple time steps (such as the land use densities, population totals, and some firm internal attributes) or that are generally static over time (access to major infrastructure like airports and freeways, as well as some permanent firm attributes), but we have additional data for the Los Angeles 2008 cross-section model. This dataset contains additional transportation variables, so these models initially include maximum auto accessibility, auto network density, and maximum transit accessibility.

Many independent variables are strongly correlated with each other, so care must be taken when choosing variables to include or exclude from the model specification. Correlation is strongest among the land use densities and among the accessibility measures included in the cross sectional models. Various arrangements of variables were tested to lessen the effects of multicolinearity on the model. Model results for density variables were determined to be generally stable as long as only a few of them were kept in the final specification. For the accessibility variables, the solution is to include transit accessibility and road network density and exclude accessibility by car, since it is also strongly correlated with the land use density measures.

When coefficients are similar in size and direction between the two models, then the variable should be interpreted as predicting turnover rates rather than an overall shift in a sector's locational preferences. If a variable has a significant positive coefficient in the formation model and a similarly sized positive coefficient in the dissolution model, then it indicates that businesses in that sort of environment or with that characteristic will generally have shorter lifespans but that those locations are also considered suitable sites for new business formation. If a variable has a negative effect on both formation and dissolution, it indicates lower turnover rates. Variables with a positive effect on firm birth and negative or no effect on death indicate areas of growth for a given category of firm, whereas variables with a positive effect on death and negative or no effect on birth indicate areas of decline. It is also worth noting that in general, formation and dissolution are codetermined with the overall success of a business establishment, which is difficult to measure and is not directly addressed by this model.

Panel Models

For the panel models, we focus on retail and manufacturing firms because these classes are similar in size and were more stable over the study period than were professional services and health care, both of which experienced substantial spikes in growth at the end of the period.

We also expect these businesses to prefer different types of environments, since retailers may prefer dense, mixed-use environments with access to customers and manufacturing firms are likely to perform just as well in more sparsely-developed areas, so long as those are well-served by transportation infrastructure.

As noted above, business creation, success, and failure are codetermined processes that may best be modeled simultaneously, but even these relatively simple models reveal many things about the ways in which business locations support or harm businesses. Birth and death models should be analyzed side-by-side, to make it possible to distinguish between variables that cause change in the overall locations of businesses over time and those that predict higher turnover in certain locations than others. These models include only observations from the years 1995-2012; we limit the study period because establishment ages are not known for certain for those that began before 1990 and because excluding observations that are more than two decades old should improve the relevance of results to today.

In order to save processing time and keep within the memory requirements imposed by our statistics program, we extract random samples from the dataset and estimate models on these samples. Given that our samples represent more than 10% of all observations for retail and over 20% for manufacturing, sampling error should introduce very little sampling error to the results, and indeed repeating model estimation on new samples returned nearly identical results. Highly significant coefficients ($p < 0.01$) remained significant across all sample replications and the only variables to experience substantial changes were the insignificant ones included for comparison. Random effects are significant only in the birth models, possibly because the nonlinear time-trend variables used in the dissolution models masked the autocorrelation of firm events. The inclusion of random effects should not impact coefficient estimates one way or the other. The order in which variables are discussed in text matches the order in which these variables appear in the accompanying tables of coefficients.

Retail

In general, our models for retail indicate that retail businesses fare best in dense, accessible urban environments in close proximity to entertainment and finance firms.

This project focuses primarily on the locational aspects of firm success, but internal attributes must be considered as well. The internal aspects of retail establishment formation and dissolution found in the panel models are generally unsurprising.

- The four dummy variables for employee count that we use to indicate business size all have very significant and, with one exception, positive coefficients in both models (when compared to the reference class – businesses with more than 20 employees); this indicates that small businesses are generally more likely to fail and that new businesses are generally smaller than established firms (businesses are likely to start small).
- As discussed in the section on general methods, sales per employee is an imperfect measure of a firm's efficiency, but it has a significant effect in these models. Less efficient businesses are more likely to fail, since presumably they are less profitable. The negative coefficient in the formation model indicates that low efficiency is a useful predictor of a business's newness, suggesting both that it may take a few years of existence for a new retail businesses to reach its full market potential and that less efficient businesses are winnowed out within a few years, and are therefore disproportionately present among newer establishments.
- Independent businesses experience higher rates of turnover than do those affiliated with a multi-establishment firm. The coefficient on the variable indicating whether an establishment is a standalone entity is positive for both the birth and death model, which means that they tend to have short lifespans.
- We incorporate the effects of time into this model in a number of ways. A simple linear time trend is included, as is an indicator for recession years (as discussed in the data processing section, it takes a value of 1 for the years 2001, 2008, and 2009 and is 0 for all other years). Additionally, the death model includes firm age as well as the square and cube of age, which essentially represents a polynomial estimate of the change in a business's success over its lifespan. The combined effect of these five time variables over the study period is shown in Figure 8, with each curve representing the effect on a business that opened in a different year. The combined effect of the time variables suggests that businesses are most vulnerable to fail in their first few years, after which their odds of failure decrease and eventually stabilize. There may be a significant uptick in failure odds for retail establishments over 20 years old. Recessions significantly increase the odds of business failure and significantly decrease the number of new firms that start.

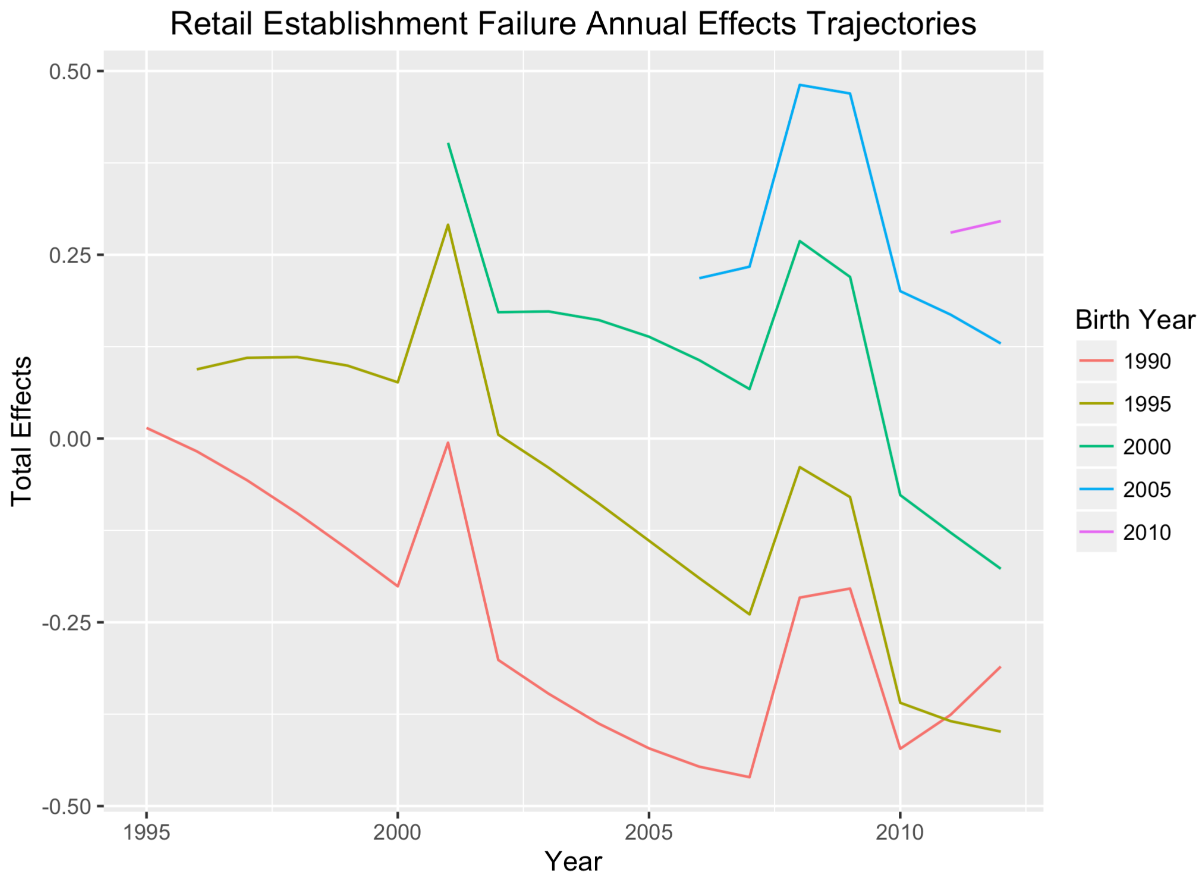

Figure 8 Time Effects on Death of a Retail Business Establishment

The two market area variables that we include are chosen to indicate both a firm's local environment and its overall access to potential employees and customers. By assessing the formation and dissolution models jointly, we find that increasing local density decreases turnover. Disproportionately many businesses choose to locate in regions with high density, presumably because access to large numbers of customers is important for the success of retail firms, but the effect of regional density on failure rates is weaker and not clearly significant. The local population density effect may be related to the land use employment density variables, but it also may indicate that retail businesses benefit from having access to a very local customer base, despite this likely being associated with high rents.

As discussed in the data processing section, we attempt to capture the effects of collaboration/agglomeration and competition by using kernel density surfaces of employment counts in various sectors. These findings here can be taken as indicators of which sorts of mixed use environments are best for retail businesses.

- Retail businesses compete with each other for customers and real estate, leading to high retail formation and dissolution in areas with lots of retail employees.
- Retail also appears to compete to some degree with wholesalers, though the effect on failure rates is much weaker. Retail establishments located near professional services firms also experience higher rates of turnover; since these firms are often located in low-density suburban environments, this may indicate that these areas represent somewhat marginal sites for retail. Contrary to findings for the Los Angeles cross-section, manufacturing does not have a significant effect on retail either way when the entire study period is considered.
- Health care businesses appear to exert an exclusively negative impact on retailers, as health care density predicts higher failure rates and lower birth rates. One potential explanation of this effect is that since health care companies often cluster around major hospitals, their rapid expansion over time may have pushed most other land uses out of these areas.
- Retail firms benefit from areas with more entertainment businesses, since these may attract shoppers and entice them to spent more time in the area. Finance firms and public administration density also appears to benefit retail firms; the local presence of employees with high paying or stable jobs provides retail businesses with a stable customer base beyond what is offered by local residents.

Transportation infrastructure is a key component of business success, and for retail businesses, access to freeways was found to be the most significant predictor of all the variables we tested. Retail establishments farther away from freeways are more likely to fail and new retail businesses are less likely to be located far from freeways. Though our Southern California cross-section models suggests that public transit matters a great deal as well, we did not have access to statewide transit data. Retail businesses also prefer to locate near airports, though this does not have a consistent effect on their success rate.

Table 2 Panel Data Probit Regression for Births (Formation) and Deaths (Dissolution)

RETAIL BUSINESS ESTABLISHMENT	Formation		Dissolution	
Variables	**Coefficient**	**P-Value***	**Coefficient**	**P-Value***
(Intercept)	-1.7945		-1.7781	
*One Employee**	0.5418	<0.0001	0.1731	<0.0001
*Two Employees**	0.5895	<0.0001	0.1917	<0.0001
*3-8 Employees**	0.4470	<0.0001	0.2981	<0.0001
*9-20 Employees**	-0.0783	0.0012	0.0633	0.0006
Sales / Employees	-0.0196	<0.0001	-0.0116	<0.0001
*Standalone Business**	0.3122	<0.0001	0.1527	<0.0001
Year - 1990	-0.0230	<0.0001	0.0124	<0.0001
Recession Year	-0.0279	0.0004	0.2462	<0.0001
Establishment Age (yrs)			0.0285	<0.0001
Age Squared			-0.0091	<0.0001
Age Cubed			0.0003	<0.0001
Population within 2km (per 10,000)	-0.0150	<0.0001	-0.0081	<0.0001
Population between 2 and 50km (per million)	0.0062	<0.0001	0.0010	0.1891
Manufacturing Density	0.0003	0.9299	0.0033	0.2379
Retail Density	0.0274	<0.0001	0.0203	<0.0001
Wholesale Density	0.0354	<0.0001	0.0065	0.0708
Prof. Service Density	0.0272	<0.0001	0.0295	<0.0001
Health Care Density	-0.0214	<0.0001	0.0037	0.1435
Entertainment and Food Service Density	-0.0287	<0.0001	-0.0166	<0.0001
Finance Density	0.0013	0.8023	-0.0078	0.0668
Public Admin. Density	-0.0138	<0.0001	-0.0040	0.0001
ln(distance to freeway)	-0.0155	<0.0001	0.0065	0.0009
ln(distance to commercial airport)	-0.0141	0.0012	0.0012	0.7301
Rho (Panel Variance Component)	0.3993		0.0000	

* P-values are reported here to show the exact Type I error. The lower the value the more significant a regression parameter is.

Manufacturing

These models indicate that manufacturing businesses generally are less likely to survive in mixed-use environments but have similar requirements for access to transportation infrastructure as retailers do.

The internal characteristics of manufacturing businesses mainly have the same effects as those of retail firms. Small-scale manufacturers generally perform worse, as do less efficient ones. In contrast to retail, standalone manufacturers generally experience less turnover. The capital costs involved in opening a new factory are likely larger than those experienced by retailers, which may reduce the flexibility of smaller manufacturers. Larger, multi-establishment manufacturing firms may also be more flexible for other reasons, since they can more easily move operations to other parts of the United States or overseas.

The combined effect of time trend variables on manufacturing firms was generally similar to the results we find for retailers, but the effect of age cubed is much stronger (see Figure 9), which means business failure probabilities are low and stable only briefly before beginning to rise again. A study that specifically addressed obsolescence would be better-suited to investigate the causes of this increasing failure rate, but we hypothesize that it may result from technological change and obsolescence of machinery and the increasing automation of manufacturing processes, since these do not impact retail in the same way.

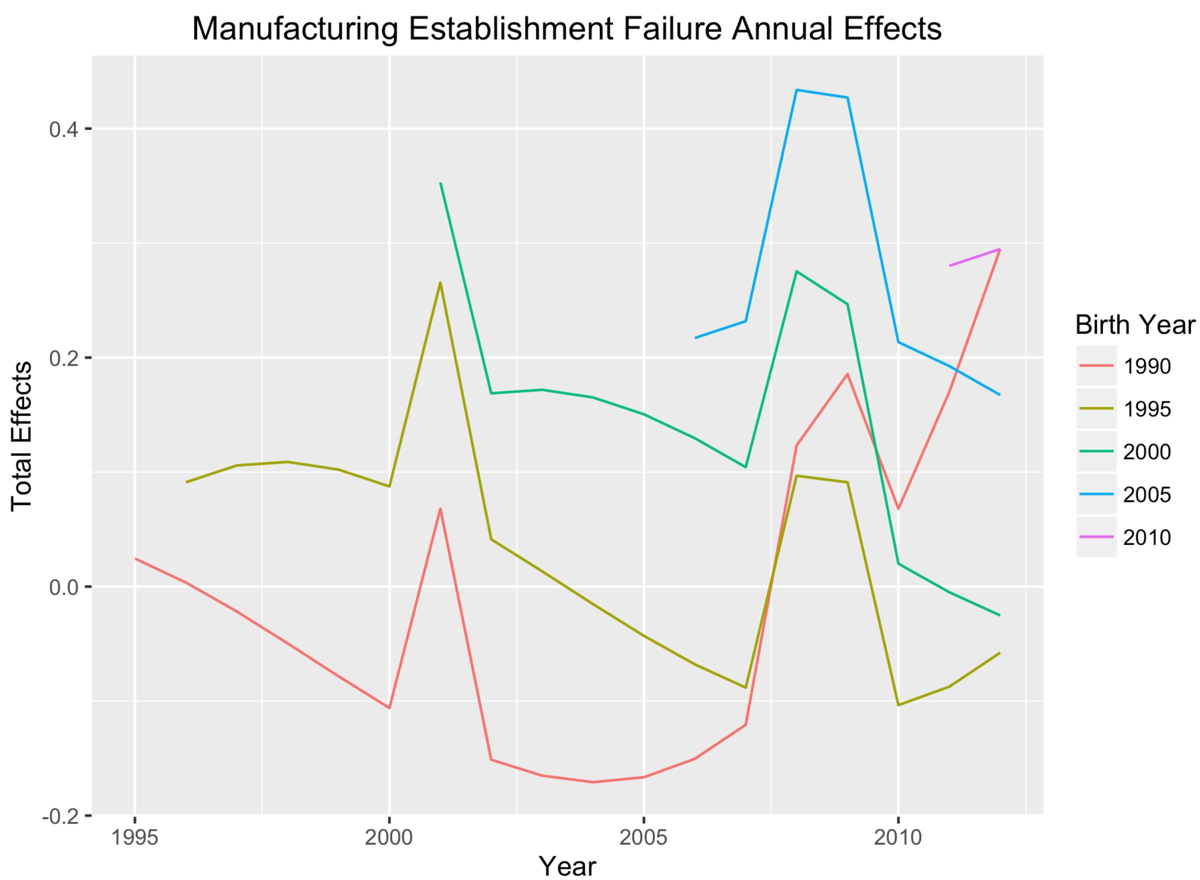

Figure 9 Time Effects on Death of a Manufacturing Business Establishment

The effects of market size on manufacturing firms are also generally similar to those experienced by the retail sector. Firms in dense areas experience much lower turnover, though the effect is stronger on the formation side, indicating new manufacturing companies prefer to avoid extremely dense areas. Manufacturers also show a slight preference for dense regions, presumably because access to large, diverse labor pools is important, though these effects are not consistently significant.

- Whereas local competition increased turnover among retailers, manufacturing firms are less likely to fail when surrounded by other manufacturing firms. This difference may be explained by the absence of competition for customers due to specialization and complementarity, reliance on shared infrastructure, and zoning laws that seek to keep factories separate from other land uses.
- In fact, manufacturing firms appear more likely to fail in any kind of mixed environment, as they perform worse in areas with high density of agriculture, retail, or transportation.

- Information technology density was hypothesized to be beneficial for manufacturing, because research and development operations and manufacturing are likely to collocate, and this is borne out by the model. Over the study period, the high birth rate and low death rate of manufacturing firms situated in areas with information technology workers mean that these firms became more common over the study period and a strengthening relationship between information technology and manufacturing.

Manufacturers generally performed best in areas with good access to freeways and airports (odds of failure increased with distance and new starts were concentrated at lower distance). Proximity to freight processing centers decreased turnover rates for manufacturing firms, though the effect was much stronger for new starts.

Table 3 Panel Data Probit Regression for Births (Formation) and Deaths (Dissolution)

Manufacturing	Formation		Dissolution	
Variables	**Coefficient**	**P-Value***	**Coefficient**	**P-Value***
(Intercept)	-1.7752		-1.5653	
*One Employee**	0.8896	<0.0001	0.1252	<0.0001
*Two Employees**	1.0223	<0.0001	0.1471	<0.0001
*3-8 Employees**	0.6352	<0.0001	0.1354	<0.0001
*9-20 Employees**	0.1664	<0.0001	-0.0041	0.7252
Sales / Employees	-0.0172	<0.0001	-0.0077	<0.0001
*Standalone Business**	-0.0551	<0.0001	-0.1029	<0.0001
Year - 1990	-0.0244	<0.0001	0.0126	<0.0001
Recession Year	-0.0557	<0.0001	0.1989	<0.0001
Establishment Age (yrs)			0.0228	<0.0001
Age Squared			-0.0076	<0.0001
Age Cubed			0.0003	<0.0001
Population within 2km (per 10,000)	-0.0110	<0.0001	-0.0033	0.0148
Population between 2 and 50km (per million)	0.0026	0.0400	-0.0013	0.1450
Agriculture Density	-0.0100	<0.0001	0.0040	0.0048
Utilities Sector Density	-0.0028	0.0255	-0.0020	0.0334
Manufacturing Density	-0.0237	<0.0001	-0.0208	<0.0001
Retail Density	0.0192	0.0001	0.0151	0.0001
Transportation and Warehousing Density	-0.0052	0.1098	0.0079	0.0018
Information Sector Density	0.0056	0.0778	-0.0055	0.0273
Prof. Service Density	0.0451	<0.0001	0.0270	<0.0001
Other Service Density	-0.0212	0.0002	0.0114	0.0094
ln(distance to freeway)	-0.0054	0.0664	0.0053	0.0157
ln(distance to commercial airport)	-0.0078	0.1036	0.0085	0.0156
ln(distance to freight processing)	-0.0269	<0.0001	-0.0059	0.0434
Rho (Panel Variance Component)	0.3688		0.0000	

* P-values are reported here to show the exact Type I error. The lower the value the more significant a regression parameter is.

Leaving California

Relatively few businesses relocate between states, but according to these models, relocation events share major commonalities with firm dissolution events, which suggests there are similar characteristics among the that cause businesses to leave the state and the locations that push businesses towards failure.

Among retail businesses, standalone firms were much more likely to relocate out of state, as were very large (>20 employees) and very small stores (one or two employees). Firms in retail-dense areas were less likely to leave, whereas retail firms in low-density / suburban environments were more likely to leave the state. Long-distance relocation of retail firms is generally rare (since it entails abandoning any long-term customers), the higher rates of relocation among retail firms in unusual locations indicates that there an unmeasured variables that may be playing an important role (e.g., extreme specialty stores and mail-order companies may be less tied to specific locations and thus more willing to leave).

Manufacturing firms that leave the state are disproportionately likely to be large, efficient firms (maybe because they can afford substantial relocation costs). Firms in areas with less manufacturing are more likely to leave the state, which matches the birth/death model result that manufacturers prefer areas that specialize in manufacturing (e.g., technology parks). In contrast to what the death model showed, firms far from freight processing locations are more likely to leave the state, as are manufacturers located far from freeways, though this result is not clearly significant. These results indicate that manufacturing firms are more likely to leave relatively remote areas in California and may be relocating to find more suitable locations.

Table 4 Panel Data Probit Regression for Relocation Out of California

Leaving California	Retail		Manufacturing	
Variables	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**
(Intercept)	-2.3567		-2.3346	
*One Employee**	0.0123	0.7534	-0.1114	<0.0001
*Two Employees**	-0.0020	0.9601	-0.1300	<0.0001
*3-8 Employees**	-0.0646	0.0852	-0.0905	<0.0001
*9-20 Employees**	-0.0902	0.0411	-0.0232	0.3380
Sales / Employees	0.0000	0.9556	0.0001	0.2024
*Standalone Business**	0.1106	<0.0001	-0.2280	<0.0001
Year - 1990	0.0165	<0.0001	0.0090	<0.0001
Recession Year	0.0436	0.0137	-0.0010	0.9557
Establishment Age (yrs)	0.0012	0.9134	0.0449	<0.0001
Age Squared	0.0006	0.5962	-0.0040	0.0004
Age Cubed	0.0000	0.2488	0.0001	0.0018
Population within 2km (per 10,000)	-0.0623	<0.0001	-0.0424	<0.0001
Population between 2 and 50km (per million)	-0.0033	0.1733	-0.0014	0.5670
Agriculture Density			-0.0112	0.0018
Utilities Sector Density			-0.0101	<0.0001
Manufacturing Density	-0.0027	0.7148	-0.0179	0.0055
Retail Density	-0.0413	0.0010	-0.0229	0.0163
Wholesale Density	-0.0082	0.3899		
Transportation and Warehousing Density			-0.0123	0.0466
Information Sector Density			0.0135	0.0289
Prof. Service Density	0.0986	<0.0001	0.0783	<0.0001
Health Care Density	-0.0008	0.9060		
Entertainment and Food Service Density	-0.0010	0.9137		
Other Service Density			0.0079	0.4631
Finance Density	-0.0071	0.5268		
Public Admin. Density	-0.0067	0.0217		
ln(distance to freeway)	0.0212	0.0001	0.0073	0.1866
ln(distance to commercial airport)	-0.0203	0.0277	0.0004	0.9597
ln(distance to freight processing)			0.0340	<0.0001

Relocation within California

Though some businesses relocations involve moves between states, relocations within California are much more common. For these moves, our data provides a more complete picture, because we have equally detailed information about the origin and destination environments. Unfortunately, this symmetry of information cannot be extended to businesses that did not relocate, so these comparisons will be made without the use of an econometric model. To eliminate autocorrelation, this analysis includes only the final move recorded for each business establishment in the dataset. In this section we briefly investigate the ways in which businesses that relocate within the state are different from those that do not, as well as the ways in which origin and destination locations differ from each other.

Table 5 compares the mean characteristics of businesses that moved with those of the entire dataset. The Total Observations row contains the number of business establishment – years recorded in the dataset. Slightly less than 1% of retail business observations include a relocation, and slightly more than 2% of manufacturing business observations do. For all variables, the difference between the move sample mean and the population are highly statistically significant, but this is partly an effect of the large sample size, and these differences are generally fairly small compared to the overall differences between firms. For both business types, the firm internal characteristics of movers generally match the population as a whole, but medium-sized, highly-efficient firms are significantly more likely to move. Standalone retail establishments are somewhat more likely to relocate, but standalone manufacturers are slightly less likely to do so. For both business types, establishments in heavily populated regions (movers have a higher population within 50km than do non-movers) with a good access to transportation infrastructure (the three distance to transportation variables are lower for the movers) are more likely to relocate, but the effects of local land use (population within 2km, total employment density, and same-class density) are mixed. Retail businesses in highly dense environments are less likely to relocate (their mean density values are lower), whereas manufacturers in high density areas are more likely to relocate.

Table 5 Relocation Origins vs All Locations

Within California	Retail			
Variables	**Move Origins**	**All Locations**	**Move Origins**	**All Locations**
Total Observations	60,336	6,302,151	44,978	2,713,860
One Employee	21.1%	22.1%	12.8%	19.9%
Two Employees	26.8%	26.1%	14.2%	18.7%
3-8 Employees	39.9%	39.2%	36.4%	33.1%
9-20 Employees	8.2%	7.5%	18.6%	13.7%
21+ Employees	4.1%	5.0%	18.1%	14.6%
Standalone Business	87.1%	84.5%	84.0%	85.9%
Efficiency	$108,775	$98,320	$119,427	$99,401
Population within 2km	32,076	32,418	29,025	28,532
Population within 50km	5,505,318	5,073,490	6,128,219	5,603,230
Total Employment Density	2,475	2,617	2,953	2,820
Density of Same Class	259	267	580	454
Distance to Freeway	2.2	3.1	1.7	2.6
Distance to Commercial Airport	19.8	22.6	17.9	20.9
Distance to Freight Processing	23.5	27.4	19.7	24.2

Comparing the origins and destinations of business relocation reveal some slight location preferences but also reveal the idiosyncratic nature of relocation decisions. Tables 6 and 7 contain mean values for origins and destinations of moves, as well as the percent of relocating businesses that increased the value of a given variable. Both retail and manufacturing businesses generally move to areas of less dense land use; this result holds on average across all measures of employment density for both business types, but more than 40% of individual businesses move to areas with increased density. This preference for … Population density tells a slightly different story – while businesses generally locate to areas with less-dense population in their immediate vicinity and regional density decreases on average, a majority of businesses move to higher-density regions. This effect likely represents a combination of two types of moves to suburban developments (whether malls or industrial parks): some businesses move from very dense areas to less dense areas in order to decrease property rental costs and a larger number move from very low-density areas in order to increase access to customers and labor. Businesses also appear willing to sacrifice some access to transportation infrastructure when they move, but these results are much closer to 50-50.

Table 6 Retail Relocation Origins vs Destinations

Variables	**Origin**	**Destination**	**% that Increased**
Population within 2km	32,076	28,806	43.4%
Population within 50km	5,505,318	5,366,531	54.6%
Total Employment Density	2,475	2,067	41.3%
Manufacturing Density	262	236	43.7%
Wholesale Density	156	132	43.3%
Retail Density	259	217	41.0%
Transportation and Warehousing Density	55	47	43.6%
Information Sector Density	109	91	42.8%
Professional Services Density	421	356	43.1%
Health Care Density	223	184	42.5%
Entertainment and Food Service Density	210	172	41.7%
Other Service Density	135	110	40.4%
Finance Density	248	206	42.5%
Public Admin. Density	157	116	43.5%
Educational Services Density	109	89	42.7%
Distance to Freeway	2.2	2.4	51.0%
Distance to Commercial Airport	19.8	20.8	52.7%
Distance to Freight Processing	23.5	24.7	52.9%

Table 7 Manufacturing Relocation Origins vs Destinations

Variables	**Origin**	**Destination**	**% that Increased**
Population within 2km	29,025	25,282	42.4%
Population within 50km	6,128,219	5,816,715	50.4%
Total Employment Density	2,953	2,421	41.3%
Manufacturing Density	580	513	44.4%
Wholesale Density	271	220	43.7%
Retail Density	243	196	41.1%
Transportation and Warehousing Density	80	70	44.5%
Information Sector Density	121	98	43.1%
Professional Services Density	489	400	42.5%
Health Care Density	176	144	42.9%
Entertainment and Food Service Density	189	152	41.7%
Other Service Density	133	108	40.7%
Finance Density	264	204	42.6%
Public Admin. Density	135	94	44.4%
Educational Services Density	104	81	42.8%
Distance to Freeway	1.7	2.0	51.9%
Distance to Commercial Airport	17.9	19.5	53.5%
Distance to Freight Processing	19.7	21.8	53.2%

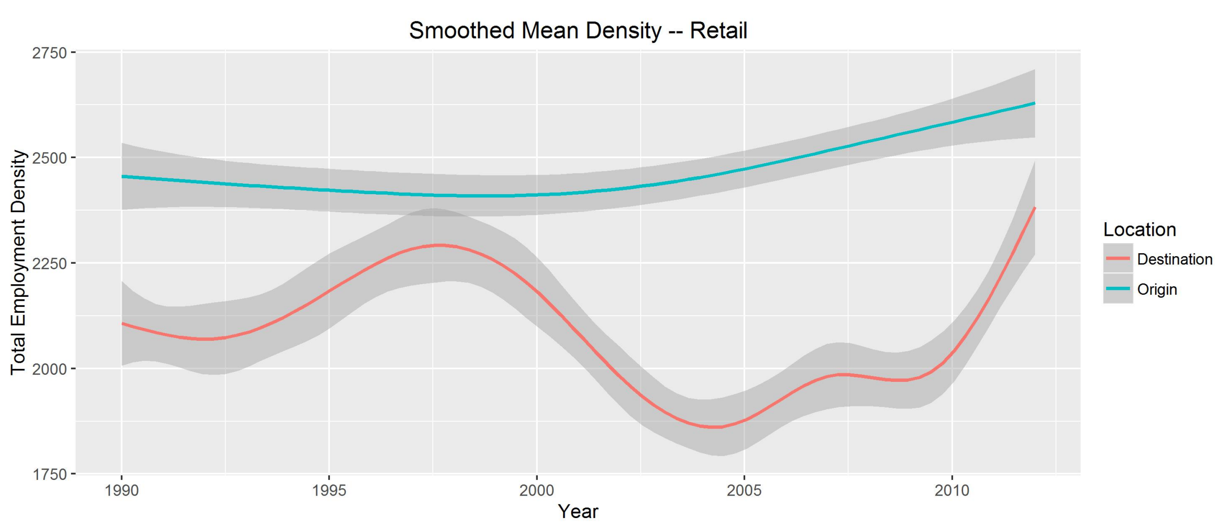

Figure 10 Retail Establishment Relocation Densities Origins vs Destinations

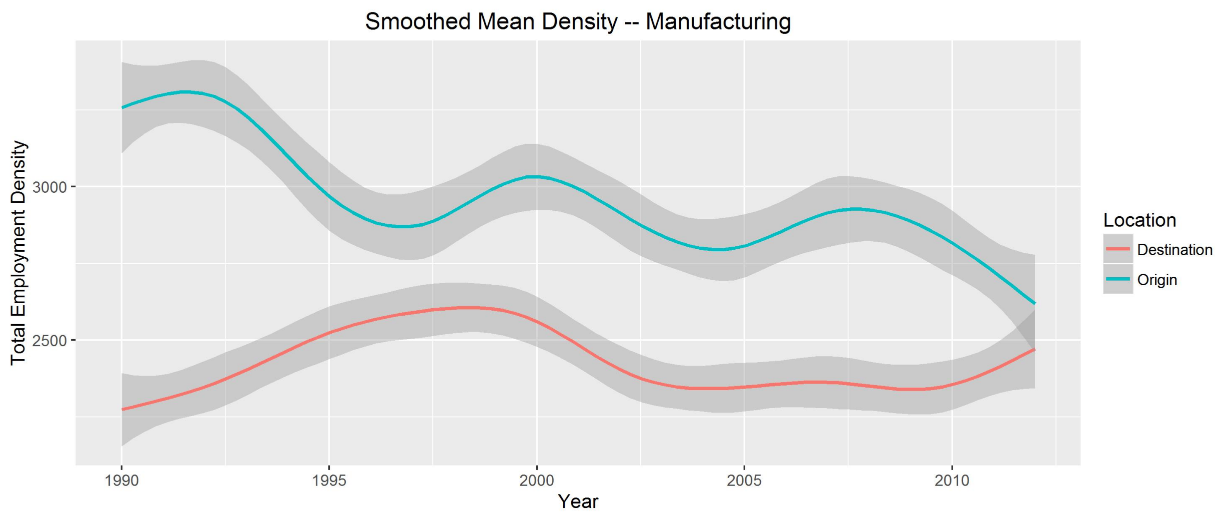

Figure 11 Manufacturing Establishment Relocation Densities Origins vs Destinations

Relocation decision-making patterns also change over course of time, as demonstrated by Figures 10 and 11. These figures display the changes in origin and destination characteristics over time, smoothed using a spline technique both for easier visualization and to reduce the noise caused by year-year variation. Darker grey parts of the graph indicate the confidence interval around the moving mean value. It is important to note that these mean values are estimated quite precisely because they are calculated off of a sample of several thousand moves each year, but the origin and destination values from individual relocation events are quite varied.

Density around move destinations never surpasses the density around move origins, which indicates that businesses are more likely to relocate to lower-density areas throughout the study period, but the relationship between the means does change over time. Retail firms became much more likely to move to higher-density areas after the last recession (2010 to present), perhaps reflecting the relative strength of urban areas, which were less badly damaged by the collapse of the housing market than were many suburbs. Both figures indicate that economic expansions (mid-to-late 1990s and 2011-present) favor moves to denser areas. While the characteristics of manufacturing moves origins is highly variable over time, the origin densities of retail firms remained remarkably stable over the study period, with a gradual increase that generally parallels the overall increase in density of California over time.

Relocation is a decision to a much greater extent than dissolution is. The formation and dissolution models can say that some areas make businesses more likely to fail or experience high degrees of business turnover, but the relocation results suggest that these results cannot be uniformly applied to the relocation process; relocation decisions change over time and are highly variable between businesses. Overall, businesses generally relocate to less-dense environments, which may indicate that relocation represents a type of failure – if businesses cannot succeed in high-competition dense environments, they may relocate to less-dense ones in order to stay open, whereas businesses that are successful in low-density environments may be more likely to expand by opening new branches rather than by relocating and abandoning a successful location. However, this relocation-as-failure interpretation is challenged by the overall high efficiency values of firms that chose to relocate.

Southern California Case Study

In order to assess the affects of transportation level of service and accessibility more directly, we take advantage of fine-grained network accessibility data from a major previous project with the Southern California Association of Governments; this dataset that contains detailed multimodal accessibility data for the six counties in SCAG in 2008 (Chen et al., 2011; Lei et al., 2012). Because the data source covers only a portion of the state at a single point in time, we limit our analysis to 2008 and a subset of the region covering much of Southern California. The results of models using this data are expected to diverge somewhat from the statewide 1995-2012 panel models because they represent a confined area and single point in time, however because this year is relatively recent within the scope of the project, we anticipate these results will be closer to what can be seen now. The smaller datasets also make it possible to include more business categories in our analysis; we add Professional Services because businesses of this type are have very dispersed locations, ranging from home offices in residential neighborhoods to suburban strip malls to downtown centers, and health care because these businesses cover a wide variety of size classes and may have very different locational strategies than other business types (small clinics and self-employed doctors may cluster around major hospitals or may choose other locations). We hypothesize that these business types will be less responsive to transportation-related factors than retail and manufacturing.

For this analysis we use the bounding box shown in Figure 12. This area was chosen because it covers the core of the region for which we have detailed data but still includes a range of land use densities, from rural areas around Thousand Oaks, the dense heart of Los Angeles, and suburban development in San Fernando Valley and Orange County. This analysis zone covers a wide range of business environments, land use, and transportation networks but is small enough that all businesses in each category can be analyzed simultaneously, without the need to extract a random sample of observations and therefore virtually eliminating sampling error. This area contains 154,727 retail businesses, 68,700 manufacturing, 345,533 professional services, and 90,320 health care.

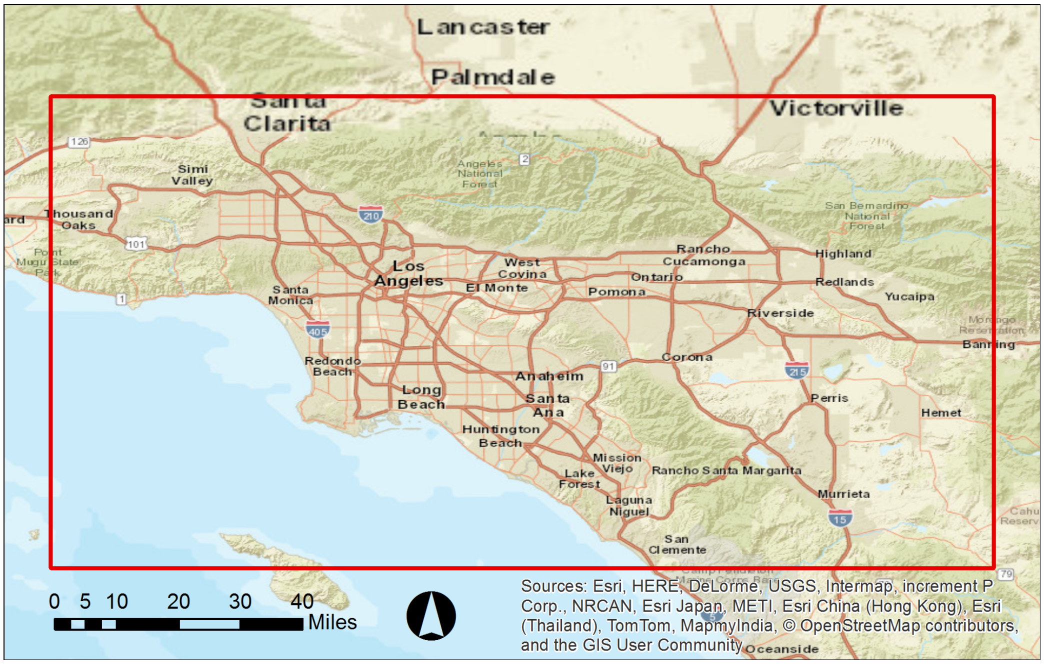

Figure 12 Bounding Box used for Southern California Analysis

To create the datasets used for regional analysis, we overlay NETS business locations on Southern California bounding box shown in Figure 12 and then generate a subset of our dataset that includes only observations of businesses at these locations in 2008. The SCAG model output provides data at the level of census blocks, which we extract to 2008 the table of business establishment observations. Variables considered in this analysis include the overall maximum number of employees reachable within 20 minutes on the automotive network and the transit network (for most blocks, this number is reached in the late morning); the total length of network links reachable from each block group within 20 minutes by car; and the number of stops reachable within 20 minutes by transit. Traffic congestion information and actual transit schedules were used to calculate these accessibility values, so they are much closer to the actual accessibility experienced by users of the infrastructure.

The small number of relocation events in any given year in this region makes it impossible to estimate a meaningful model for relocation for 2008, so we focus our analysis on the factors that predict business establishment formation and dissolution across four categories (retail, manufacturing, health care, and professional services). For these models, we started by including a suite of internal attributes, market characteristics, local land use densities, and transportation accessibilities; these variables are used to estimate a simple binary probit model for business formation or dissolution, and the list is whittled down iteratively to remove insignificant variables. Some variables are kept in either the formation model for a given business type because they are highly significant in the corresponding dissolution model and vice versa. In general, the goal is to produce models that provide a significant explanation of the observed firm life cycle events while using a small enough number of variables that the results are straightforward to comprehend. In each of these models, interaction terms were tested to determine whether accessibility has different effects on businesses of different sizes, however these tests returned no significant results; this indicates that businesses of all sizes within a given business category require similar levels of access to the transportation network, though they may be further differentiated by business sub-category.

Retail

As was the case in the statewide models, retail establishments and events are broadly distributed throughout the Los Angeles area, with more located in high-density areas and fewer in areas with less development in general. The following maps show the regional distribution of retail establishment life events in 2008; in each case, a random sample of 4,000 events is used; the lack of clear spatial patterns visible from coarse scale maps is evidence of the usefulness of regression analysis to identify significant differences between the processes leading to business formation and dissolution. These models include all 154,727 retail establishments active in this region in 2008. 18,593 new firms opened this year and 6,508 closed. Retail firm establishment formation and dissolution locations are shown in Figures 13 and 14, respectively.

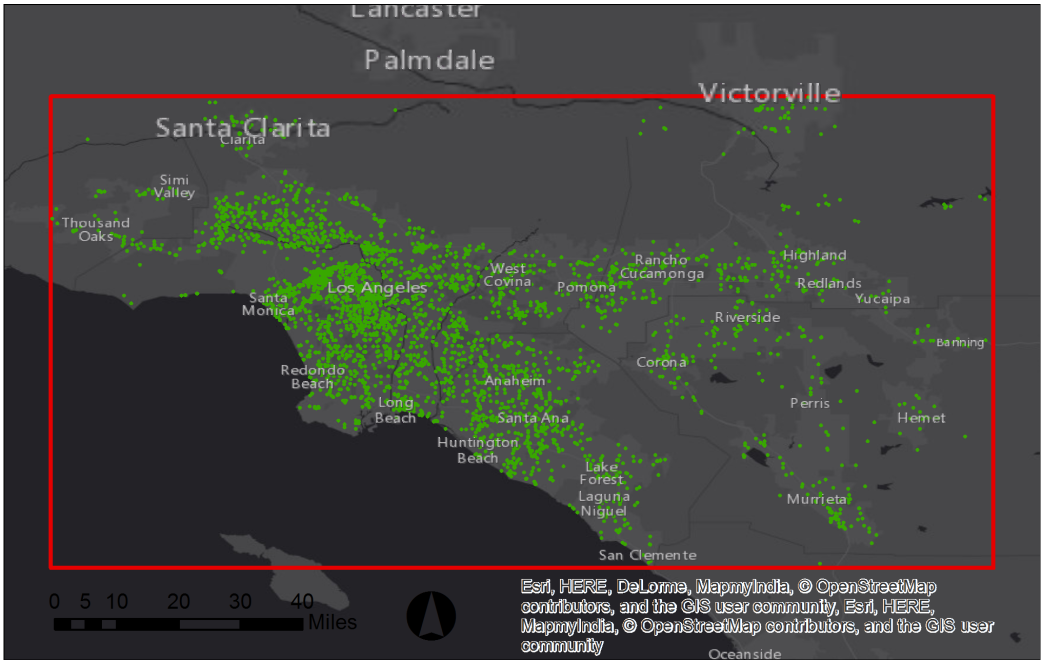

Figure 13 Retail Establishment Formation Locations 2008

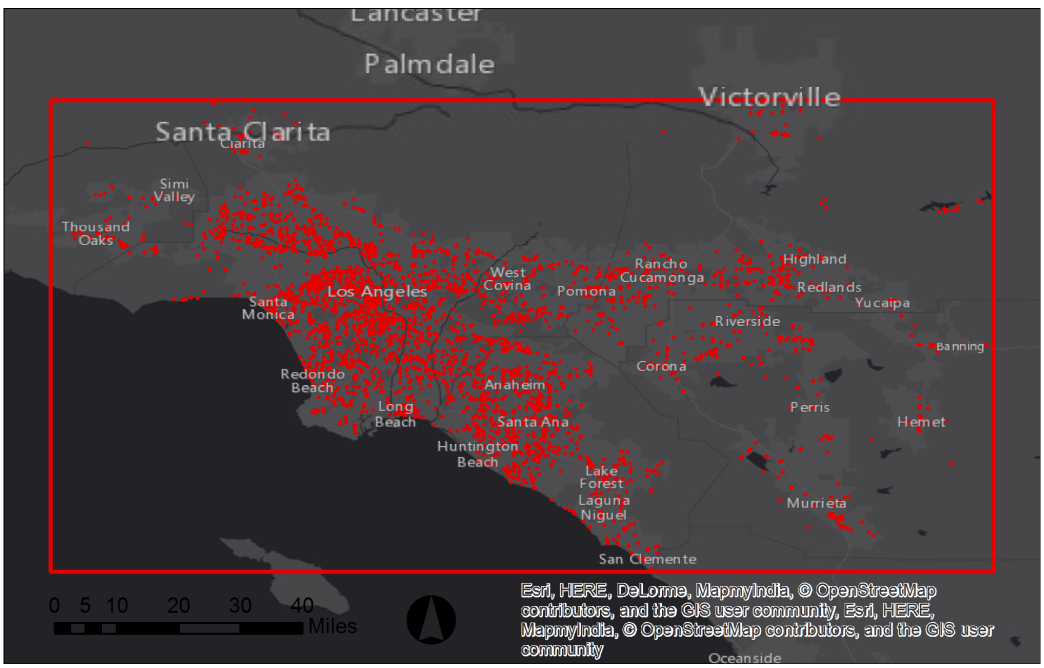

Figure 14 Retail Establishment Dissolution Locations 2008

Firm internal attributes generally have similar effects in these models as they do in the statewide full-period panel models. Smaller businesses are generally less stable – formation and dissolution rates are highest among businesses with one or two employees and higher among businesses with 3-8 employees than those with 9 or more (the difference between the coefficients on medium-sized establishments and the large-establishment reference class are insignificant in both models). Older firms were less likely to fail. Somewhat surprisingly, sales and efficiency had no significant effect on the probability of retail establishment failure in this year. One finding that does not match the panel results is that in 2008, standalone businesses were much less likely to close and disproportionately likely to open. This contradicts the panel model results, which generally found standalone businesses to be more stable. Because the specific year analyzed came at the start of a global economic crisis, this result may indicate that large firms responded to a slowing economy by closing their worse-performing locations, a freedom small firms do not have.

Market area and land use findings are somewhat similar to those from the panel models. Retail establishments in denser parts of the market are less likely to fail, and more new establishments are in areas with high population density. This may be a function of the specific year studied, if the financial collapse began affecting suburban areas sooner than it did downtowns. The 50km market area variable generally predicts lower reduction in business turnover in this year (negative on both birth and death), matching other results from this pair of models, which find more stability in dense urban environments than suburbs.

Land use effects were mixed. Retail businesses have lower turnover in areas with more professional service density, and higher turnover in areas with high retail density, indicating that there is considerable competition among retail firms. An interaction model between local retail density and establishment size indicated that larger businesses performed slightly better in areas with high competition, but all the other results remained consistent with the model presented here. Turnover is also higher in less traditional retail environments – high density of manufacturing employees predicts higher rates of both formation and dissolution. Perhaps because they correspond to the parts of urban centers with the highest rents, the presence of finance and public administration employees predicts higher risk of failure for retail firms. These models diverge from the panel results by finding no significant effect of entertainment density on establishment failure, and a negative impact on new retail formation.

The strongest transportation result from these models is the importance of public transportation accessibility. Retail firms with high transit accessibility were considerably less likely to fail, and there is a positive coefficient on transit accessibility for firm births, which indicates that retail firms are showing a genuine preference for transit-accessible parts of the Los Angeles region. Access to freeways is similarly important, as firms farther from freeways are more likely to close and new firms seek out areas closer to freeways. These findings are generally unsurprising, as retail establishments require robust transportation systems to provide access to their customers and employees.

The transit accessibility indicators used here are a function of the availability of opportunities for travelers within a day (e.g., due to opening and closing times of businesses) and a direct function of the transit routes and schedules and related spatio-temporal level of service variation during a day (Lei et al., 2012). The findings here show that increasing the frequency of public transportation service has a direct and significant positive impact on business survival and therefore the local economy.

In general, these models find that retail businesses were likelier to fail in marginal sites (low access to transportation or presence of manufacturing) or high-rent/high-competition areas (with high density of existing retail and finance firms), but this was counteracted by the presence of strong transportation infrastructure. These findings can be taken to indicate that access to transportation (and particularly public transit) is very important for retail establishment success, at least in relatively developed areas like Los Angeles.

The following figures show the spatially variable aspects of the model rendered as maps, with more intense colors identifying areas where firm births and deaths are more likely, and lighter colors indicating areas where local businesses are more stable. At this scale, relatively fine-scale local differences may not be visible (this is particularly true of the transit level-of-service variables, which are extremely localized), so these maps can only loosely be interpreted as overall suitability surfaces and the model results generally tell a more complete story.

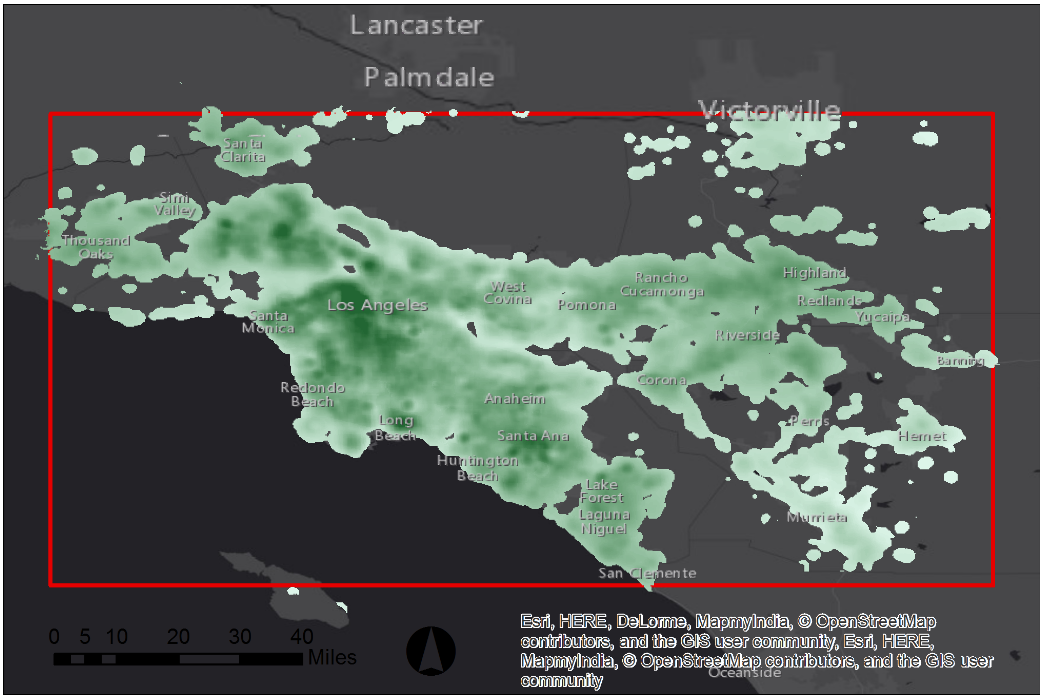

Figure 15 Southern California Retail Formation Suitability Map 2008

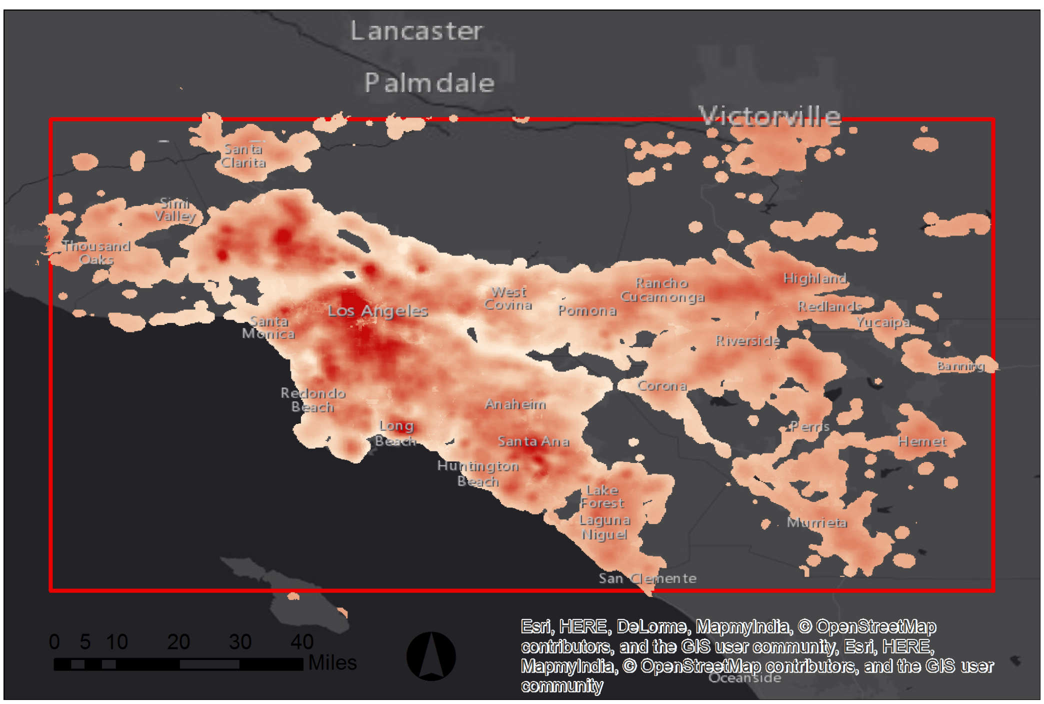

Figure 16 Southern California Retail Dissolution Likelihood Map 2008

Table 8 Southern California 2008 Retail Model Results

	Formation		Dissolution	
Variables	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**
(Intercept)	-1.9230		-1.5952	
One Employee*	0.9252	<0.0001	0.1603	0.0044
Two Employees*	0.9432	<0.0001	0.1185	0.0166
3-8 Employees*	0.5197	<0.0001	0.0905	0.0355
9-20 Employees*	-0.0357	0.5743	0.0084	0.8394
ln(Sales)	-0.0436	<0.0001	0.0060	0.5465
Standalone Business*	0.4726	<0.0001	-0.3169	<0.0001
Headquarters Business*	-0.4614	<0.0001	-0.1759	0.0004
Establishment Age (yrs)			-0.0253	<0.0001
Population within 2km (per 10,000)	0.0048	0.0173	-0.0048	0.0952
Population between 2 and 50km (per million)	-0.0137	<0.0001	-0.0033	0.2697
Utilities Sector Density	-0.0061	0.0004		
Manufacturing Density	0.0502	<0.0001	0.0146	0.0107
Retail Density	0.0297	0.0014	0.0185	0.0777
Prof. Service Density	-0.0109	0.1543	-0.0367	0.0043
Entertainment and Food Service Density	-0.0159	0.0454		
Other Service Density	-0.0315	0.0014		
Finance Density			0.0347	0.0047
Public Admin. Density			0.0045	0.0958
Segment Length within 20 minutes (per 100 km)			0.0006	0.0090
Max. 20 minute Transit Accessibility (per 10,000)	0.0139	<0.0001	-0.0067	0.0374
ln(distance to freeway)	-0.0054	0.1408	0.0125	0.0161

One additional result not presented in this is that different types of retail businesses prefer different types of environments, particularly with respect to transit level of service. A comparison of the mean values across the different business types is provided in Table 9. This table shows that some kinds of retail establishments (particularly the kind likely to be found in malls, such as clothing stores) are much more likely to be located in higher-accessibility areas for transit. In contrast, grocery stores are located in areas with higher population density, as people tend to visit them more frequently.

Table 9 Southern California Comparison of Event Occurrence and Transit Accessibility

Retail Sub-Category	**Total Ests**	**Diss. Rate**	**Form. Rate**	**Transit Accessibility**	**Pop. within 2km.**
Building Material and Garden Equipment and Supplies Dealers	6,764	3.9%	17.1%	4,466	35,745
Clothing and Clothing Accessories Stores	30,357	4.7%	14.1%	**19,756**	47,503
Electronics and Appliance Stores	10,253	4.7%	13.9%	7,025	39,833
Food and Beverage Stores	18,168	3.6%	9.5%	8,017	47,324
Furniture and Home Furnishings Stores	10,124	4.9%	13.0%	6,423	39,647
Gasoline Stations	3,489	3.3%	5.0%	5,240	38,901
General Merchandise Stores	3,208	3.2%	10.3%	9,145	**48,027**
Health and Personal Care Stores	10,896	3.9%	11.7%	8,230	43,464
Miscellaneous Store Retailers	28,955	4.0%	12.2%	7,866	40,476
Motor Vehicle and Parts Dealers	13,570	4.7%	10.8%	4,498	39,578
Nonstore Retailers	5,347	3.3%	8.5%	5,088	35,922
Sporting Goods, Hobby, Musical Instrument, and Book Stores	13,596	4.2%	10.5%	8,721	40,484

Manufacturing

Manufacturing results in this cross-section are similar to those found in the panel models, with some notable contrasts (Table 10). Firm size, somewhat surprisingly, had very little impact on the likelihood of a business's failure, though medium-sized firms (9-20 employees) were significantly less likely to fail than larger or smaller firms. Efficiency did not have a remotely significant effect in models for this cross-section, so sales were used instead, but their importance was not clear. As in the panel model, manufacturing firms have lower turnover in areas near freight processing locations. Few of the landuse densities tested were significant in this model, but denser areas in general seem to lead to higher failure rates, possibly because of competition for real estate and rents. Information sector mattered for the panel models, but does not clearly matter here (nearly significant predictor of firm birth, nowhere near significant for firm dissolution). Access to public transit increased odds of failure and decreased new starts. Manufacturing firms are often located in less densely developed areas and have no need to provide access to customers, so the higher rents of transit-served neighborhoods may not be worth it to them. These models include all 68,700 manufacturing establishments active in this region in 2008. 5,074 new firms opened this year and 2,375 closed.

Table 10 Southern California 2008 Manufacturing Model Results

	Formation		Dissolution	
Variables	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**
(Intercept)	-1.1927		-1.5526	
One Employee*	0.5894	<0.0001	0.0162	0.8016
Two Employees*	0.8654	<0.0001	0.0120	0.8338
3-8 Employees*	0.2388	<0.0001	-0.0543	0.2553
9-20 Employees*	-0.0150	0.7736	-0.1038	0.0159
ln(Sales)	-0.0736	<0.0001	0.0181	0.1336
Standalone Business*	0.1100	0.0121	-0.3481	<0.0001
Headquarters Business*	-0.3664	<0.0001	-0.1819	0.0006
Establishment Age (yrs)			-0.0273	<0.0001
Population within 50km (per million)	-0.0101	0.0252	-0.0144	0.0082
Retail Density	0.0557	<0.0001	0.0372	0.0004
Information Sector Density	0.0117	0.1238		
Other Service Density	-0.0473	0.0015		
Distance to Center (km)	0.0053	0.0002	0.0049	0.0047
Distance to Freight Processing (km)	-0.0031	0.0003	-0.0029	0.0055
Segment Length within 20 minutes (per 100 km)	0.0006	0.0517	0.0006	0.1198
Max. 20 minute Transit Accessibility (per 10,000 employees)	-0.0050	0.1920	0.0058	0.1839

Professional Services

Professional Services establishments had a fairly mixed bag of results (Table 11). Small professional services firms (2-8 employees) were the most stable in this year. Standalone firms were strongly favored, as they were in all of the cross-section models. Turnover was highest in regions with generally high values for the land use densities across the board. Professional services firms do not display clear preferences about the transportation network, though turnover was generally somewhat higher in areas with denser road networks. These models include all 345,533 professional services establishments active in this region in 2008. 53,979 new firms opened this year and 10,044 closed.

Table 11 Southern California 2008 Professional Services Model Results

	Formation		Dissolution	
Variables	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**
(Intercept)	-1.4861		-1.6719	
One Employee*	0.7827	<0.0001	0.0276	0.5220
Two Employees*	0.9016	<0.0001	-0.2109	<0.0001
3-8 Employees*	0.2412	<0.0001	-0.1197	0.0008
9-20 Employees*	0.0506	0.1659	-0.0170	0.6368
ln(Sales)	-0.0374	<0.0001	-0.0143	0.0742
Standalone Business*	0.2669	<0.0001	-0.3273	<0.0001
Headquarters Business*	-0.2635	<0.0001	-0.3106	<0.0001
Establishment Age (yrs)			-0.0173	<0.0001
Population within 2km (per 10,000)	0.0121	<0.0001	-0.0010	0.6500
Population between 2 and 50km (per million)	-0.0111	<0.0001	-0.0155	<0.0001
Manufacturing Density	0.0251	<0.0001	0.0422	<0.0001
Retail Density	0.0284	<0.0001	0.0454	<0.0001
Information Sector Density	0.0033	0.3254	-0.0197	0.0007
Prof. Services Density	-0.0512	<0.0001	0.0087	0.3386
Entertainment Density	-0.0168	0.0003	0.0262	0.0018
Educational Services Density	-0.0047	0.0719	-0.0037	0.4395
Segment Length within 20 minutes (per 100 km)	0.0008	<0.0001	0.0006	0.0014
ln(distance to freeway)	-0.0052	0.0352	0.0013	0.7456

Healthcare

Healthcare establishments show the largest imbalance in significance between internal and external characteristics, and these firms show limited sensitivity to land use and transportation network accessibility (Table 12). Small healthcare businesses show much higher turnover than large ones, and standalone businesses are less likely to fail in this year. As we find in all models, older firms are generally more stable. Health care firms experience less turnover in areas with denser population. New healthcare businesses are likely to locate in areas with low existing healthcare density and higher retail density; this suggests a shift in locational preferences, because healthcare firm locations were generally seen to cluster around major hospitals. Dissolution was not generally found to be a function of local land use. New locations have no consistent relationship to transportation infrastructure, but failures were more common in areas with denser road networks (and maybe less in areas well-served by transit). In general, the findings in the healthcare cross-section model indicate that formation and dissolution of healthcare firms is primarily driven by (unmeasured) internal characteristics (e.g., relationships with health care providers and insurance companies). These models include all 90,320 healthcare business establishments in the region, including 11,504 formations and 2,881 dissolutions.

Table 12 Southern California 2008 Healthcare Model Results

	Formation		Dissolution	
Variables	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**
(Intercept)	-0.5966		-1.4677	
One Employee*	0.3602	<0.0001	0.1446	0.0694
Two Employees*	0.4528	<0.0001	0.2581	0.0002
3-8 Employees*	0.6173	<0.0001	0.0498	0.4051
9-20 Employees*	-0.2766	<0.0001	0.0069	0.9091
ln(Sales)	-0.1330	<0.0001	-0.0050	0.7403
Standalone Business*	0.6346	<0.0001	-0.1985	<0.0001
Headquarters Business*	-0.3827	0.0005	-0.1419	0.0949
Establishment Age (yrs)			-0.0343	<0.0001
Population within 50km (per million)	-0.0120	<0.0001	-0.0077	0.0552
Healthcare Density	-0.0475	<0.0001	0.0005	0.6877
Retail Density	0.0527	<0.0001		
Prof.Service Density	0.0100	0.2242		
Segment Length within 20 minutes (per 100 km)	-0.0002	0.3650	0.0008	0.0071
Max. 20 minute Transit Accessibility (per 10,000 employees)	0.0033	0.4219	-0.0083	0.1594

Cross Section General Findings

The models estimated for our Southern California 2008 cross-section have some similarities to panel model results, but they reveal some differences possibly reflecting local conditions and their very different data structure. We find that certain types of mixed use are more beneficial than others. The failure models generally show loss of business establishments located in locations not perfectly suited for that type of business, which suggests that certain types of mixed-use development are unlikely to succeed: 1) retail establishments in manufacturing areas more likely to fail; 2) manufacturing establishments near retail more likely to fail; and 3) professional services more likely to fail anywhere there is competition for real estate. However, mixed use is not all bad: retail businesses benefit from the presence of workers in professional services and manufacturing benefits from presence of information industries, perhaps illustrating the necessity of direct collaboration between different segments of high tech industries.

A lack of symmetry between creation and dissolution results means that differences between sites affect overall success differently. These results indicate even more clearly than the panel models did that business survival is not just a matter of high-turnover vs low-turnover areas. The franchise/non-standalone penalty found in these results does not match previous research or the panel model results, and may be specific to 2008, since large companies may have begun consolidating locations earlier in the recession than smaller companies started to close.

Conclusions

In this section we provide answers in a summary format to the main research questions of this project followed by limitations and next steps.

1. **In what ways does access to transportation infrastructure affect the success, failure, and relocation of businesses?**

Business establishments of all types rely on access to transportation infrastructure, but businesses with good access to infrastructure are not uniformly more likely to succeed. Transportation impacts the initial location decisions of most firms as well as influencing success later on, and this increased demand drives up real estate costs and sometimes increases failure rates. Businesses are less likely to relocate to less-accessible areas than to less-dense ones, but local land use density appears to play a more significant part in relocation decisions than does access to transportation. The Los Angeles case study demonstrated the substantial importance of public transit accessibility to many types of industries, since it provides access for customers and workers across a range of socioeconomic status and permits even denser concentrations of activity in already dense areas that would not be possible if access were possible exclusively by car. Data limitations and the large scope of this project mean it may not be possible to extract the importance of specific transportation policies and projects on business establishment success at a statewide scale.

2. **In what ways does local land use affect the success, failure, and relocation of businesses? Which types of mixed-use environments are beneficial to business establishment success?**

Density provides lots of opportunities for collaboration but also for competition. Denser environments generally experience higher levels of turnover due to general competition and high real estate and labor costs. Mixed use is clearly beneficial in some cases, but not in others. Certain specific pairings are particularly beneficial – retail establishments are less likely to fail when they are located in environments that also provide customers with entertainment and dining opportunities; manufacturing businesses may perform better when in close proximity to information technology firms. Others are not – many business types perform worse in areas with high same-class density, since direct competition drives down profits and small professional services firms perform best in areas with very low land use / employment density across all categories. Interestingly, manufacturing firms seem to do best in limited-use developments. Manufacturing turnover is lowest in areas with high density of manufacturing employees, and other land uses (including residential) likely prefer it that way as well, since it keeps the noise and air pollution generated by manufacturing companies somewhat more contained in specific areas.

3. **How do the effects of land use and transportation accessibility vary across different business types?**

Manufacturing and retail showed similar preferences across all the models, but transit accessibility was notably important for retail and not at all important for manufacturing. Somewhat surprisingly, manufacturing showed no especially strong preference for access to freight processing centers, possibly reflecting the isolation of many of these centers in massive ports and inland train depots. Access to freeways is generally beneficial for all types of business establishments. As hypothesized, health care and professional services do not show any particularly strong preferences about transportation access, which indicates that their location decisions play a smaller role in their success/failure than do those made by retail and manufacturing firms.

4. **How are the factors that predict the formation of new business establishments relate to the factors that predict the dissolution of businesses?**

Many of the variables we examined were had a strong positive bearing on firm birth but also on firm death (or were negative in both models). Rather than representing a net cost or benefit to a type of business, these variables serve as predictors of business turnover rates. It is beyond the scope of this study to investigate whether turnover has any substantial effect on a region's overall economic health, but the consistently high rates of turnover in dense areas indicates that it is not the worst thing in the world. Turnover likely has more negative consequences for individual employees. This may be especially true in low-wage sectors like retail, where business closures mean changes in schedule and the need to learn new products, as opposed to higher-paying business types. Still, these areas also have higher rates of firm establishment, and the constant creation of new opportunities may be beneficial to workers.

Limitations and Next Steps

Although we met the objectives of this project a main limitation emerges from the focus on Southern California in our attempt to correlate level of service with business establishment events. This was done to make the analysis tractable and take advantage of detailed transit accessibility for that specific region. If other regions have similar indicators we could expand the analysis accordingly. A second limitation of the overall analysis is the lack of network-based centrality indicators statewide and for each year of the NETS business establishment inventory. This would have allowed a more direct correlation between business establishment events and central location measured using the infrastructure instead of the business density. Although it is feasible to expand the analysis here using network-based centrality indicators, this type of work may be more appropriate for city-level analysis. A third limitation is due to our focus on four types of businesses and it may be worth testing the same models for all types of business establishments. Finally, it should be noted that we analyze business establishment events and not firm evolution and correlation of events with business practices.

One key takeaway from this research is the benefit of local studies, particularly when detailed data can be made available for a particular region. The Los Angeles case study allowed us to investigate a wider range of variables in more detail than was possible for the statewide models. We recommend developing detailed local studies that examine business establishment histories and land parcel histories to build microsimulation of land use that is sensitive to the level of service and the relationship of co-locating businesses. Another takeaway from the relocation section is that businesses' preferences change over time. This is a limitation for any analysis of this sort, as predictive models based on past results will miss this change. In future work (and for planning now) it is important to consider what exactly "mixed use" means and what are the effects of this sort of development on the success of different types of businesses. This may again be more appropriate to be done with focus on a region instead of the State.

References

California Department of Transportation. Caltrans GIS Data Library. <http://www.dot.ca.gov/hq/tsip/gis/datalibrary/> (accessed February 14, 2016).

Chen, Y., Ravulaparthy, S., Deutsch, K., Dalal, P., Yoon, S., Lei, T., ... & Hu, H. H. (2011). Development of indicators of opportunity-based accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, (2255), 58-68.

de Bok, M. and van Oort, F.G. (2011). Agglomeration economies, accessibility and spatial choice behavior of relocating firms. The Journal of Transport and Land Use, vol.4, no.1, pp.5-24.

Elegar, I. and Miller, E. (2006). Conceptual model of location of small office firms. Transportation Research Record: The Journal of Transportation Research Board, vol.1977, pp.190-196.

ESRI (2015). Kernel Density. *ArcMap 10.3 Online Help.* <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/kernel-density.htm>

ESRI (2015). Focal Statistics. *ArcMap 10.3 Online Help.* <http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-analyst-toolbox/focal-statistics.htm>

Greene, W. H. (2003). Econometric analysis, 5th. *Ed.. Upper Saddle River, NJ*.

Hunt, J., Khan, J. and Abraham, J. (2003). Microsimulating firm spatial behavior. In Proceedings of 8th International conference on computers in urban planning and urban management, Sendai.

Hunt, J.D. and Abraham, J.E. (2005). Design and Implementation of PECAS: A generalized system for the allocation of economic production exchange and consumption quantities. Lee Gosselin and Doherty (eds.) Foundations of Integrated Land-Use and Transportation Models: Assumptions and New Conceptual Frameworks. Elsevier.

Kumar, S., & Kockelman, K. M. (2008). Tracking size, location, and interactions of businesses: Microsimulation of firm behavior in Austin, Texas. Transportation Research Record: Journal of the Transportation Research Board, 2077, 113-121.

Lei, T., Chen, Y., & Goulias, K. (2012). Opportunity-Based Dynamic Transit Accessibility in Southern California: Measurement, Findings, and Comparison with Automobile Accessibility. *Transportation Research Record: Journal of the Transportation Research Board*, (2276), 26-37.

Manzato, G.G., Arentze, T.A. and Timmermans, H.J.P. and Ettema, D. (2010a). Exploration of location influences on firm survival rates with parametric duration models. Transportation Research Record: The Journal of Transportation Research Board, no.2245, pp.124-130.

Manzato, G.G., Arentze, T.A. and Timmermans, H.J.P. and Ettema, D. (2010b). Location and accessibility mediated influences of office firm closure rates: A proportional hazard model. International Journal of Urban Sciences, vol.14, no.1, pp.1-15.

Maoh, H. and Kanaroglou, P. (2009). Intrametropolitan location of business establishments: microanalytical model for Hamilton, Ontario, Canada. Transportation Research Record: The Journal of Transportation Board, no.2133, pp.33-45.

Maoh, H. and Kanaroglou, P. (2013). Modeling firm failure: Towards building a firmographic microsimulation model. In Pagliara, F. et al. (eds.) Employment Location in Cities and Regions. Advances in Spatial Sciences. Springer-Verlag, Berlin.

Miller, E.J., Farooq, B., Chicgcuanco, F. and Wang, D. (2010). Microsimulating urban spatial dynamics: Historical validation tests using ILUTE model system. Presented at the Workshop in Urban Dynamics, Marbella, Chile March 22-25, 2010.

Minnesota Population Center. (2011). *National Historical Geographic Information System: Version 2.0*. Minneapolis, MN: University of Minnesota. http://www.nhgis.org

Moeckel, R. (2009). Simulation of firms as a planning support system to limit urban sprawl of jobs. Environment and Planning B, vol.36, pp.883-905.

Ravulaparthy S. K. and K.G. Goulias (2011) Forecasting with Dynamic Microsimulation: Design, Implementation, and Demonstration. Final Report on Review, Model Guidelines, and a Pilot Test for a Santa Barbara County Application. University of California Transportation Center (UCTC) Research Project. Geotrans Research Report 0511-01, May, Santa Barbara, CA.

Ravulaparthy, S.K. (2013) Spatial Perspectives in Business Establishment Behavioral Modeling: A Case-Study Analysis in Santa Barbara County. PhD Dissertation, University of California Santa Barbara.

Ravulaparthy, S.K., and K.G. Goulias (2014) Characterizing the Composition of Economic Activities in Central Locations: A Graph-Theoretic Approach to Urban Network Analysis. *Transportation Research Record: Journal of the Transportation Research Board (in press).*

Strauch, D., Moeckel, R., Wegener, M., Gräfe, J., Mühlhans, H., Rindsfüser, G. and Beckman, K.J. (2005). Linking transportation and land-use planning: The microscopic dynamic simulation model ILUMASS. Peter M. Atkinson and Giles M. Foody et al. (eds.) Geodynamics, CRC Press Boca Raton, Florida.

U.S. Bureau of Economic Analysis, 2014 Use Table (Before Redistribution, Producer Prices). http://www.bea.gov/industry/xls/io-annual/IOUse_Before_Redefinitions_PRO_1997-2014_Sector.xlsx (accessed January 14, 2016).

van Wissen, L.J.G. (2000). A microsimulation model of firms: Applications of concepts of demography of the firm. Papers in Regional Science, vol.79, pp.111-134.

Waddell, P., Ulfarsson, G.F., Franklin, J.P. and Lobb, J. (2007). Incorporating land-use in metropolitan transportation planning. Transportation Research Part A: Policy and Practice, vol.41, pp.382-410.

Walls, D. (2007). National Establishment Time Series Database: Data Overview. Presented at 2007 Kauffman Symposium on Entrepreneurship and Innovation Data.