Title

Exploratory Study of Spatial Variation in Car Insurance Premiums, Traffic Volume and Vehice Accidents

Authors

Ong, Paul
Sung, Hyun-Gun

Publication Date

2003-07-18

Exploratory Study of Spatial Variation in Car Insurance Premiums, Traffic Volume and Vehicle Accidents[1]

Paul Ong
with Hyun-Gun Sung
July 18, 2003

Introduction:

Traffic accidents are inevitable, albeit undesirable, outcomes of vehicular travel. They impose a heavy burden to individuals and society. The most obvious individual costs are related to bodily harm, damaged property, and lost time. Even those not involved bear a cost in terms of insurance premiums to offset the risks of driving. We rely on the private market to establish the price for coverage, with state oversight to address potential problems of market concentration, imperfect information and moral hazard. The fairness of variations in premiums hinges in part on the use of unbiased actuary rates, but this is a necessary but not sufficient condition. Differences are questionable when based on discrimination or when externalities are present.

One of the most controversial issues is the accusation of "redlining," a practice of charging higher premiums for those residing in low-income, minority neighborhoods.[2]

---

1. I wish to thank the members of the Lewis Center for Regional Policy Studies who aided our research in this study. In particular, we would like to thank: Douglas Houston, Douglas Miller and Margaret Johnson. The research included in this paper was supported by a grant from the University of California Transportation Center. The authors alone are responsible for any errors. Neither the University of California nor the School of Public Policy and Social Research either support or disavow the findings in any project, report, paper, or research listed herein. University affiliations are for identification only; the University is not involved in or responsible for the project.

2. Minority neighborhoods are also the same areas with relatively few insurance offices. For example, State Farm Mutual Insurance Company has no agents in most of the zip codes in central and south-central Los Angeles, areas that have high concentrations of welfare recipients (The Foundation for Taxpayer and Consumer Rights, 1999; Glionna, 1999).

According to one study of premiums in Los Angeles, the rates for individuals with identical coverage and driving histories can vary by nearly a factor of two, with the highest costs in the inner city and the lowest in affluent suburbs. (Ong, 2002) Over half of the geographic variation in the rates can be explained by two factors, minorities as a percent of the population and median income.[3]

"Redlining," however, may be due to underlying environmental risk factors rather than simple racial discrimination. (Harrington and Niehaus, 1998) Rates are higher in some neighborhoods because the risks are higher. There is, however, a more fundamental question about how the population is distributed across the urban landscape. A lack of affordable units outside the inner city and housing discrimination constrain many low-income minorities to areas with a high traffic volume, much of it generated by drivers from outside the neighborhoods. In other words, living in a low-income and minority neighborhood exposes an individual to greater odds of being in an accident, holding personal characteristics and risk factors constant. Although a higher premium may reflect a higher cost of offering insurance in these areas, the burden falls on everyone, even an individual with a decent driving record. A driver must bear an externality cost due purely to place of residence.

---

3. This finding is based on an OLS (ordinary least squares) regression with the average insurance rate as the dependent variable. The independent variables are percent minority, median household income (divided by 10,000) and the square of median income. Demographic and income data come from the 1990 census. The estimated coefficients and their t-value are

$$\begin{array}{lllll} \text{Rate} = 1488 & + 95*(\%\ \text{Minority}) & - 284*(\text{Income}) & + 26.8*(\text{Income Squared}) \\ \quad (16.72) & \quad (1.73) & \quad (7.80) & \quad (6.29) \end{array}$$

Adj R-square = .52

For some, this cost is not just related to higher premiums. When insurance is prohibitively high, some residents drive without insurance, which places them and others at great personal financial risk. (Bernstein, 1999) Others are forced to become transit dependent, which limits their employment opportunities and access to services and activities. (Ong and Houston, 2002; Ong and Miller, 2003) The inequality, then, is a consequence of the complex way a city is spatially structured.

The complexity of the problem requires careful analysis of the process generating inequality. This paper presents findings from an exploratory study. Part 1 examines the relationship between insurance costs and claims by zip-code areas in the urbanized area of Los Angeles County, and the analysis indicates that premiums are closely tied with claim rates. The highest premiums and claim rates tend to be in the downtown area, and the lowest tend to be in the outlying areas. Part 2 examines traffic volume and accident rates in the City of Los Angeles. The analysis reveals that traffic levels are highest in the downtown area and lowest in the outlying areas, and this high volume generates higher accident rates. The paper concludes with a discussion about the potential implications for low-income minority neighborhoods and recommendations for future research.

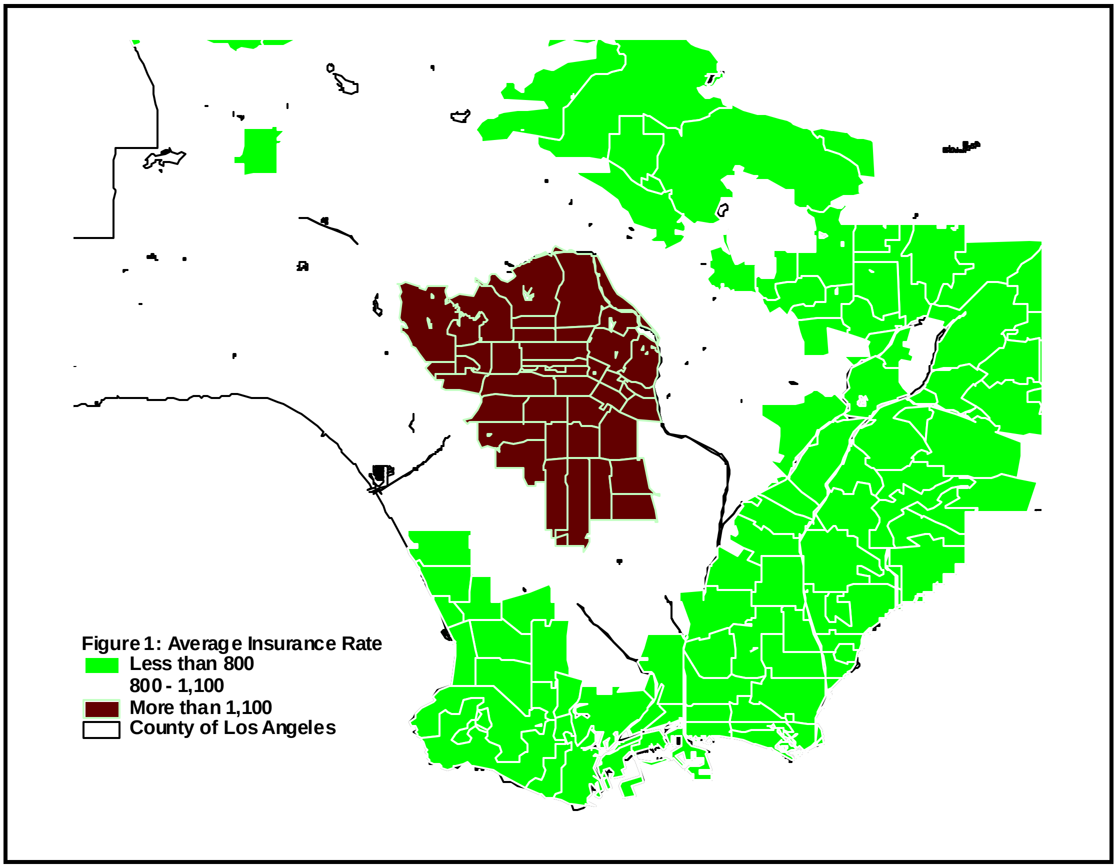
Part 1: Spatial Variation in Insurance Premiums

This section examines how insurance premiums vary across neighborhoods, and the influence of insurance claims on the variation. Insurance data come from two sources. Premiums were collected for the year 2000 from multiple quotes for each zip code.[4] To

---

4. The estimates are based on information provided by the following website: http://www.realquote.com. Multiple quotes from different insurers were requested for each zip code. To capture the "pure" geographic variation of insurance rates, we held the characteristic of the "applicant" constant by using the same demographic profile for every zip code: a 25-year old

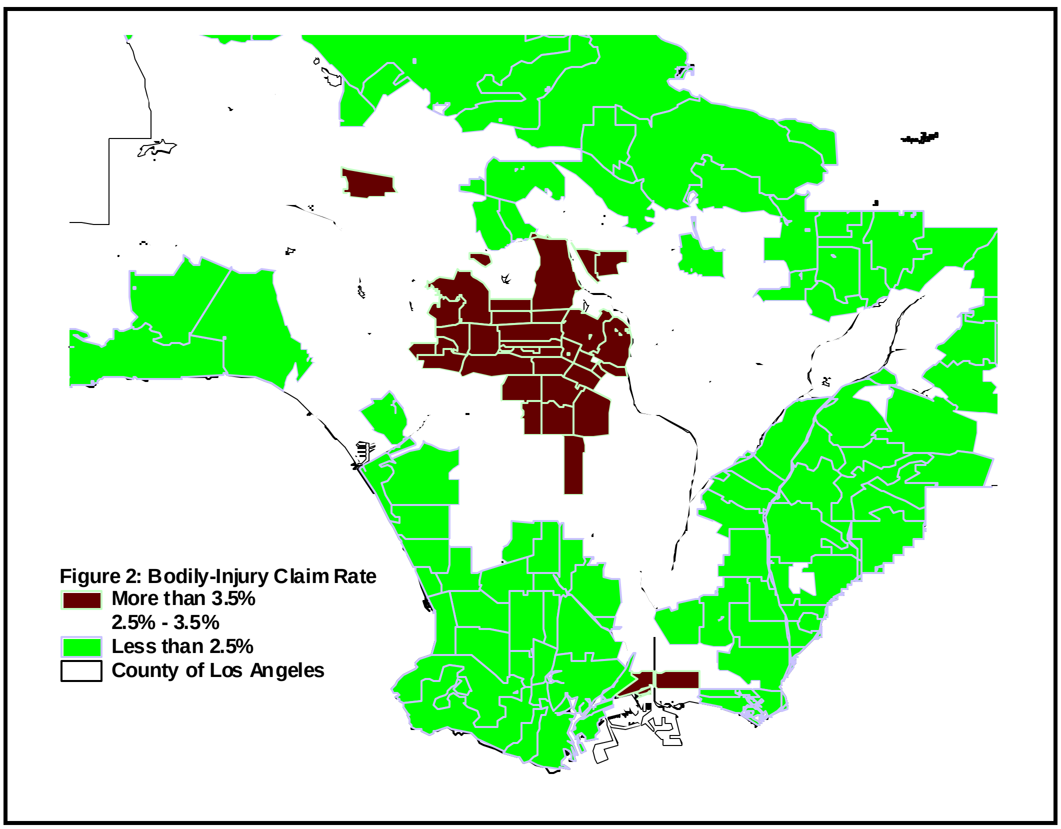
eliminate variations in premiums due to individual differences (coverage, type of car, driving history), the quotes are for an identical hypothetical person. The spatial variation, then, captures the "pure" differences associated with location. Premiums for basic coverage for a single mother with an older car and fair driving record range from $405 to $1,275 per year, with a mean of $875 and standard deviation of $183. The geographic variation in premiums is depicted in Figure 1. The highest rates are in the inner city, and the lowest rates are in the outlying suburbs.



---

employed single mother, who has been driving for seven years, had taken a driver training course, and has one moving violation, but no accidents and is a non-smoker. She owns a 1990 Ford Escort LX, 2-door hatchback with no anti-theft devices, no anti-lock brakes and no airbags, which is parked on the street. She carries only the minimum insurance required ($15/30,000 bodily liability, $5,000 property liability) with no deductibles. The insurance premium for each zip code is the average of quotes from at least a half dozen companies. Premiums for basic coverage for a single mother with an older car and fair driving record range from $679 to $1,275 per year, with a mean of $975 and standard deviation of $177.

The insurance industry states that premiums are determine in part by territorial-based risk as documented by historical claims. Data on insurance claims come from the California Department of Insurance, which provides statistics on the frequency and severity of accidents by zip code for the period from 1988 to 1993. (Hunstad, 1996) The data are reported for the location where a vehicle is garaged and not by the location of the accident. Frequency rates are calculated as the number of claims divided by the number of policies. Because the rates for different types of claims (property damage, medical, collision, etc.) are highly collinear, this study uses one index, the bodily injury rate. The geographic variation is depicted in Figure 2. The highest rates are in the inner city, and the lowest rates are in the outlying suburbs, suggesting a high degree of correlation with spatial patterns of insurance premiums in Figure 1.



We use ordinary least squares (OLS) regressions to estimate the influence of claim rates on premiums. Claims are reported per 100 policies, and the average value across zip code areas is 1.51, with a standard deviation of 0.66. A second variable is included, which is the standard deviation of the quotes within each zip code. The average value of that variable is 266, with a standard deviation of 57. This measures the relative efficiency of the insurance market in each zip code. A highly functional market is one where full information and vigorous competition drive out opportunities for excess profits, thus producing more uniform prices. The results are reported in Table 1. The estimated coefficients are highly significant and consistent with *a priori* expectations. The model explains over 90% of the variation in average premiums across zip codes. Dropping the second independent variables increases the size of the estimated coefficient for the bodily-injury claim rate but does not significantly lower the adjusted r-square.

Table 1: Regression Results

Dependent Variable: Average Premium

	Model 1		Model 2	
	Coefficient	P-value	Coefficient	P-Value
Constant	107	<.0001	250	<.0001
Bodily-Injury Claim Rate	144	<.0001	239	<.0001
Standard Deviation of quotes	1.47	<.0001	N/A	
Number of zip codes	280			280
Adjusted R-Square	.904			.823

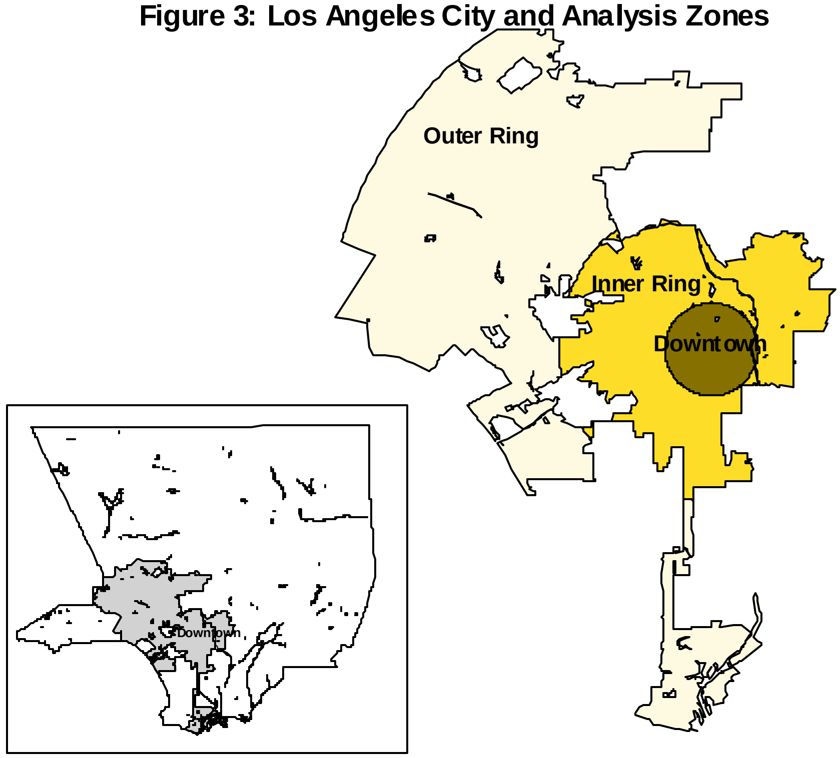
The results are consistent with the contention that the regulated private market sets premiums that are consistent with historical claims rates. It is also consistent with the argument that even good drivers in the inner city face a higher cost due to place of residence. Recall, the dependent variable is for an identical hypothetical driver, thus there is no difference in coverage or driving history.

From a societal perspective, it is important to determine what leads to spatial differences in risk. One possible explanation is that the bodily-injury rates are systematically tied to vehicular accident rates within neighborhoods. While this is a plausible assertion given that a large percentage of accidents occur close to home, the claim data are not reported by locations of accidents. Consequently, the findings in this section only infer this relationship.

Part 2: Vehicular Accidents and Traffic Volume

The available data allow us to examine spatial variations in traffic volume and accident rates in the City of Los Angeles, and an analysis of these two sets of data provides insights into how risk varies across neighborhood types. Los Angeles City is the single largest city in the Los Angeles metropolitan area and home to 3.7 million of the region's 9.5 million people counted in the 2000 census. Because of limited resources and methodological issues, the analysis uses only three zones: the downtown area, the adjacent inner-ring, and an outer ring that includes the more suburban and less densely populated areas. Figure 3 shows the City's boundaries and analysis zones.

Figure 3: Los Angeles City and Analysis Zones



Accident data come from Los Angeles City Department of Transportation. The data set contains information on type of accident by location, usually by intersections. The data set covers the years from 1993 to 1999. We are able to geocode nearly 96% of the accidents. Traffic volume data come from California's Department of Transportation, Traffic Operations Division. The data are reported as average daily traffic counts for 2000.[5] Because the LA City accident data covers only non-highway streets, traffic

5. A description of the data is posted at the department's web site, http://www.dot.ca.gov/hq/traffops/saferesr/trafdata/2000all.htm. "Annual average daily traffic is the total volume for the year divided by 365 days. The traffic count year is from October 1st through September 30th. Very few locations in California are actually counted continuously. Traffic Counting is generally performed by electronic counting instruments moved from location throughout the State in a program of continuous traffic count sampling. The resulting counts are adjusted to an estimate of annual average daily traffic by compensating for seasonal influence, weekly variation and other variables which may be present."

volume is tabulated only for local street segments.[6] Counts of private vehicles in households come from the 2000 Decennial Census (Summary File 3).

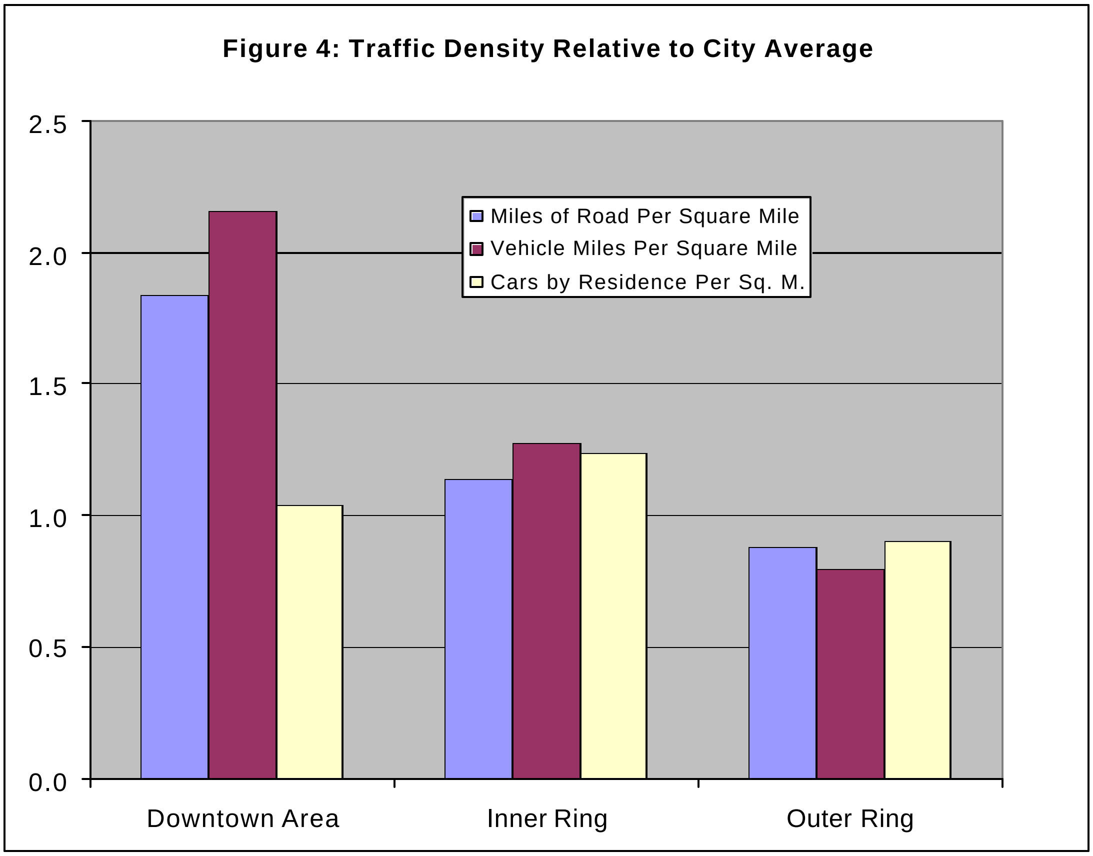


Figure 4 summarizes statistics on traffic volume per square mile in the three zones relative to the citywide averages. A value of 1 indicates a level identical to the citywide average, a value less than 1 indicates a level below the citywide average, and a value greater than 1 indicates a level above the citywide average. The street network is nearly twice as dense in the downtown area, while the network density in the outer ring is below the city average. The volume of traffic is even more unequally distributed, with the

6. Including highways and freeways inflates the values of our measurements of traffic density, but does not qualitatively alter the findings regarding the spatial variations in traffic volume and accident rates.

downtown area having over twice the number of vehicle miles per square mile. Higher traffic level is not related to the number of vehicles owned by local residents. Despite a higher housing density in the downtown area, the number of vehicles per square mile is lower than in the inner ring, due in part to a lower income in the former area. The lower vehicular density in the outer ring is due to the lower density of housing. When all of the statistics are considered together, the implication is that much of the traffic in the downtown area comes from outside the area.

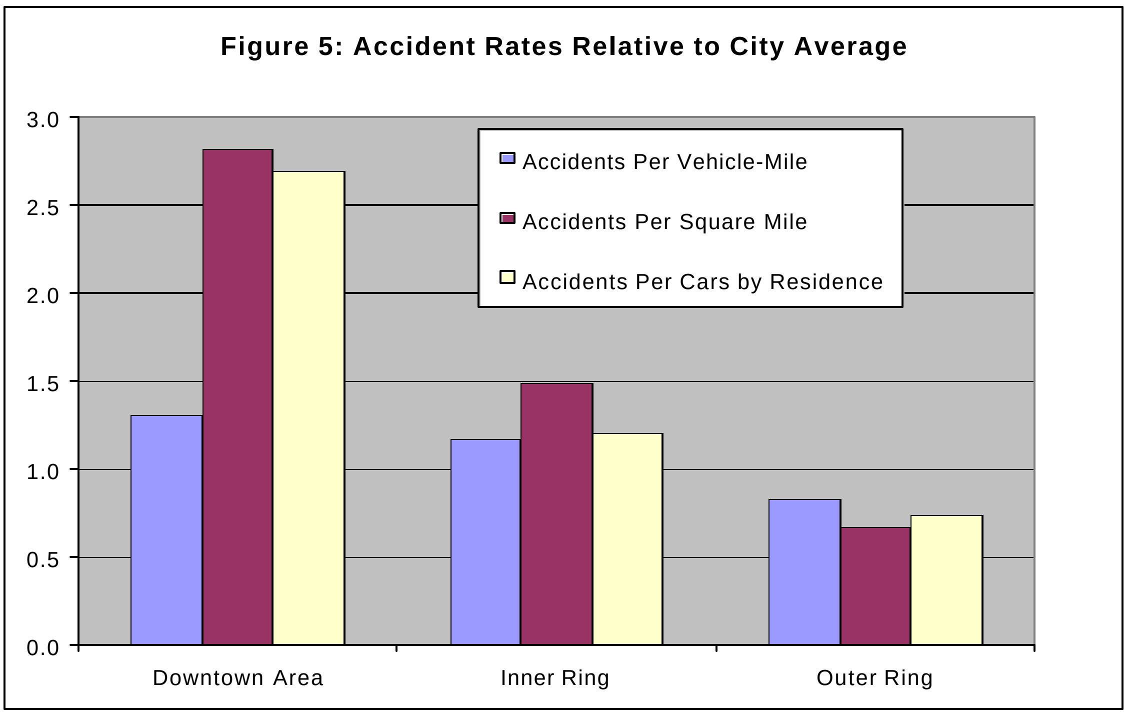
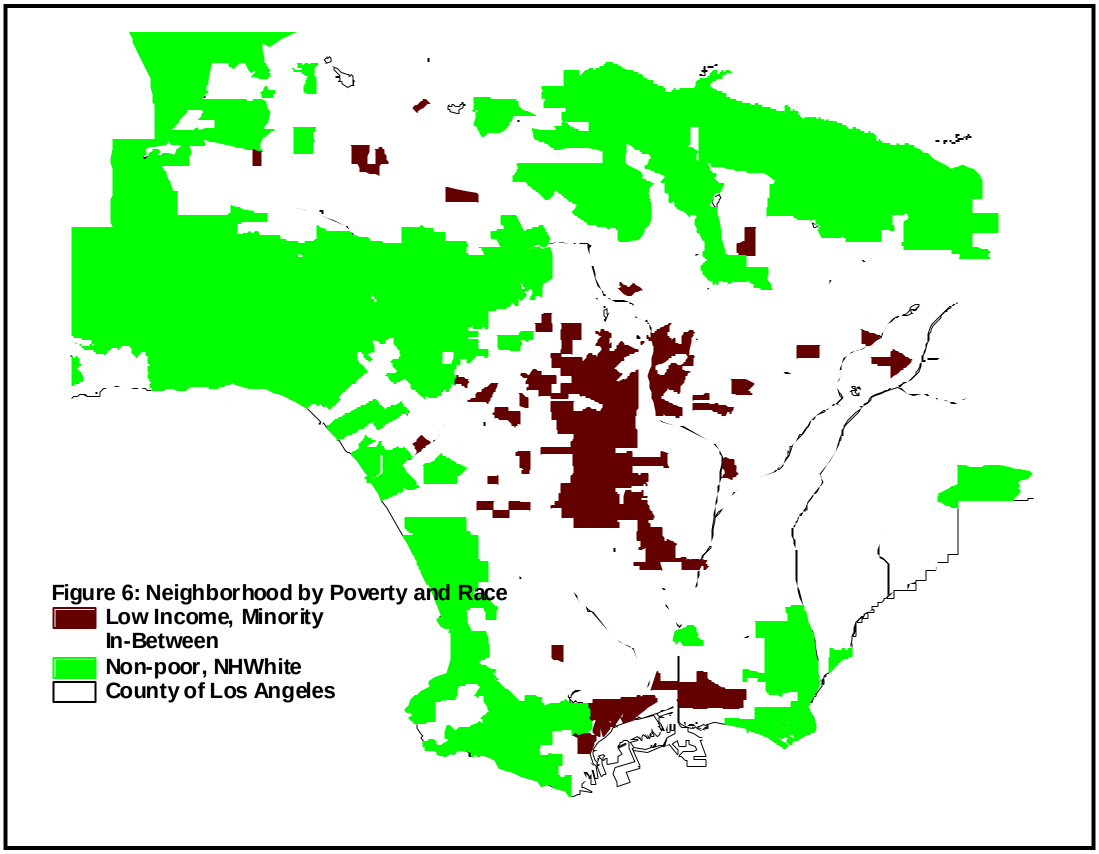


Figure 5 summarizes the statistics on accident rates. The first bar in each group of bars shows that the accident rate per vehicle mile increases with the density of traffic. In other words, the odds of an accident occurring per vehicle mile in the downtown area is over one and a half times higher than in the outer ring. Because traffic density is considerably higher in the downtown area, the number of accidents per square mile is over four times higher than in the outer ring. The geographic disparity in the number of accidents per vehicle by place of residence is nearly as large. Of course, the statistics do not mean that the number of accidents per neighborhood car is several times higher in the inner city than in the suburbs. Accidents involve both cars from a neighborhood and from outside. Despite this limitation, the statistics strongly suggest that the risk of having an accident is higher in the downtown area. Local residents spend a disproportionate amount of their driving time on these streets, so the odds are stacked against them. Moreover, a portion of the additional risk is imposed by traffic imported from elsewhere in the region.

Concluding Remarks:

The above findings are consistent with the fundamental hypothesis that some of the observed inequality in insurance premiums among neighborhoods is indirectly generated as a consequence of the way a city is spatially structured. The inner city bears a disproportionate share of traffic volume, which in turn increases exposure to the risk of a vehicular accident within this part of the urban landscape. The inner city is also the place where claim rates are above average, which accounts for higher insurance premiums, holding constant coverage and driving record. As mentioned earlier, we do not have data to show that a disproportionate number of those claims by inner-city residents are related to accidents in the inner city; nonetheless, this is a reasonable inference given the findings from the ecological analysis of traffic volume and accident rates. The outer suburban areas have the opposite characteristics: lower than average traffic volume, accident rates, claim rates and insurance premiums.



The above spatial patterns coincide with the socioeconomic geography of Los Angeles, which is depicted in Figure 6. Racial/ethnic and poverty data come from the Summary File 3 for the 2000 census. Low-income, minority neighborhoods are defined as census tracts where at least 30% of the population live in households with income below the federal poverty line and at least 80% of the population are minority. Non-poor, non-minority neighborhoods are defined as census tracts where less than 10% of the population live in households with income below the federal poverty line and at least 65% of the population is non-Hispanic white (NHW). The spatial patterns are very distinct, with low-income, minority neighborhoods heavily concentrated in the inner city around the downtown area, and most of the non-poor, NHW neighborhoods located in the outlying suburbs. This concentric pattern coincides with the pattern depicted in previous figures.

The overlapping insurance, traffic, accident and socioeconomic spatial patterns have policy implications. The findings support that assertion that insurance premiums are tied to actuary rates, but the analysis is not sufficiently refined to test whether there is residual discrimination in rate setting after controlling individual and contextual factors. Discrimination, however, is only one of the two factors that would bring variations in automobile insurance premiums into question. An equally important policy issue is how our society should address the externalities created by imported traffic on the residents of low-income, minority neighborhoods. The findings point to a form of inequality deeply embedded in the way metropolitan areas are structured. This generates a burden of living in low-income, minority communities, a burden that goes beyond the narrow findings of the above analysis. A high cost of car ownership due to insurance cost forces many to become transit dependent, which tends to isolate the residents from the rest of the region.[7]

Formulating specific policies requires additional research to better define the complex spatial interactions involving insurance, claims, traffic, accidents and the socioeconomic characteristics of neighborhoods. While the findings from the exploratory study are useful, the analysis is incomplete. The data are not consistent in terms of time period and geographic coverage, and these limitations constrain the analysis to simple statistical and non-statistical assessments. The next logical step is developing and implementing a full multivariate econometric model that tests, separates and quantifies the influence of various causal factors on outcomes at the neighborhood level. This

7. There are also associated environmental risks because greater traffic density imposes a

requires assembling and updating the data sets, geocoding records using statistically-based quality controls to eliminate spatial biases, and putting the data into common geographic units. One of the key challenges is assembling linked information on residential location and accident location, which would enable us to directly test how traffic density affects accidents of local residents within their neighborhoods. Communications with state agencies indicate that the data exist and may be available.

The subsequent step in the research is to develop and implement a multivariate model using micro-level (individual level) longitudinal data with information on residential location and accident location. This will enable us to determine more precisely the impact of local environmental factors (traffic volume and overall accident rates) on *ex ante* insurance premiums, the probability of experiencing an accident within and outside the neighborhood, and *ex post* insurance premiums. Conceptually, this could be done if the vehicles in accidents can be identified and linked to DMV data and insurance data, which include information on where the vehicles are registered or garaged, claims, and insurance rates. Because of confidentiality concerns, this research would require considerable cooperation from state agencies and administrative procedures to ensure the confidentiality of the data. Past research experience with similar administrative data indicates that conducting a micro-level project is doable but would be very costly and difficult to implement.

---

health risk. (Gunier, et al., 2003)

References:

Bernstein, Robert (1999). "California Uninsured Vehicles as of June 1, 1997," California Department of Insurance, Sacramento, February 1999.

Glionna, John M. (1999). "State Farm Suit Seeks Return of Data Critics Say Indicate Redlining," *Los Angeles Times*, December 5, 1999.

Gunier, Robert B., Andrew Hertz, Julie Von Behren, and Peggy Reynolds (2003). "Traffic Density in California: Socioeconomic and Ethnic Differences Among Potentially Exposed Children," *Journal of Exposure Analysis and Environmental Epidemiology*, 13:240-246.

Harrington, Scott E. and Greg Niehaus (1998). "Race, Redlining, and Automobile Insurance Prices," *Journal of Business* 71(3):439-480.

Hunstad, Lyn (1996). "Methodology and Data Used to Develop the California Private Passenger Auto Frequency and Severity Bands Manual," California Department of Insurance, April.

Ong, Paul (2002). "Car Ownership and Welfare-to-Work," *Journal of Policy Analysis and Management* 21(2):255-268.

Ong, Paul and Douglas Houston (2002), "Transit, Employment, and Women on Welfare," *Urban Geography*, 23(4):344-364, 2002.

Ong, Paul and Douglas Miller (2003). "Spatial and Transportation Mismatch in Los Angeles," working paper, June, 2003.

The Foundation for Taxpayer and Consumer Rights (1999). Website Map Illustrates State Farm Redlining in Los Angeles, CA, News Release, December 7, 1999. http://www.consumerwatchdog.org/insurance/pr/pr000017.php3