



National Poverty Center Working Paper Series

#06-24

July, 2006

REDLINING OR RISK?

A Spatial Analysis of Auto Insurance Rates in Los Angeles

Paul M. Ong, School of Public Affairs, UCLA

Michael A. Stoll, School of Public Affairs, UCLA

This paper is available online at the National Poverty Center Working Paper Series index at:
http://www.npc.umich.edu/publications/working_papers/

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of the National Poverty Center or any sponsoring agency.

REDLINING OR RISK?

A Spatial Analysis of Auto Insurance Rates in Los Angeles¹

by

Paul M. Ong and Michael A. Stoll

School of Public Affairs
UCLA
3250 Public Policy Bldg., Box 951656
Los Angeles, CA 90095-1656
pmong@ucla.edu and mstoll@ucla.edu

June 2006

¹ We are indebted to UCTC (University of California Transportation Center) and the UCLA Lewis Center for partial funding for this project. We are also grateful to Mathew Graham, Leah Brooks, and Cheol-Ho Lee for assistance in collecting, assembling and cleaning the data.

REDLINING OR RISK?

A Spatial Analysis of Auto Insurance Rates in Los Angeles

Abstract

Auto insurance rates can vary dramatically, with much higher premiums in poor and minority areas than elsewhere, even after accounting for individual characteristics, driving history and coverage. This paper uses a unique data set to examine the relative influence of place-based socioeconomic characteristics (or “redlining”) and place-based risk factors on the place-based component of automobile insurance premiums. We use a novel approach of combining tract-level census data and car insurance rate quotes from multiple companies for sub-areas within the city of Los Angeles. The quotes are for an individual with identical demographic and auto characteristics, driving records and insurance coverage. This method allows the individual demographic and driving record to be fixed. Multivariate models are then used to estimate the independent contributions of these risk and “redlining” factors to the place-based component of the car insurance premium. We find that both risk and “redlining” factors are associated with variations in insurance costs in the place-based component, with black and poor neighborhoods being adversely affected, although risk factors are stronger predictors. However, even after risk factors are taken into account in the model specification, SES factors remain statistically significant. Moreover, simulations show that “redlining” factors explain more of the gap in auto insurance premiums between black (and Latino) and white neighborhoods and between poor and nonpoor neighborhoods.

Introduction

Automobile ownership is essential to accessing economic opportunities in modern metropolitan labor markets, and any financial barriers to such access can have adverse impacts on employment and earnings. A key cost is insuring a vehicle, which can be a significant burden for low-income and disadvantaged households.² But car insurance rates can vary dramatically, with much higher premiums in poor and minority areas than elsewhere, even after accounting for individual characteristics, driving history and coverage (Ong, 2002). An important policy issue is whether the higher car insurance premiums faced by economically disadvantaged households are based on *de facto* discrimination or on fair and legitimate risk factors. The former explanation is widely known as “redlining,” a discriminatory practice of setting rates for individuals who live in places with a disproportionately large number of poor and minority people. On the other hand, the insurance industry has long asserted that the unequal spatial pattern of car insurance premiums is based on place-based risk factors and that the demographic characteristics of neighborhood are not part of the formula used to set prices.

The debate is not just academic. California, where the Department of Insurance is reconsidering how much weight should be given to place characteristics in setting car insurance premiums, is an example of how this debate can affect public policy. Some, such as civil rights and drivers’ interest groups, advocate entirely eliminating the use of place in setting of car insurance premiums because such a practice lends itself to “redlining,”

² Ong (2002) shows that differences across neighborhoods within a metropolitan area in average insurance costs have large and negative impacts on car ownership rates, and Raphael and Rice (2002) demonstrate that differences across states in average insurance costs have large and negative impacts on car ownership rates. The adverse impacts disproportionately affect blacks and Latinos.

especially when effective oversight is absent. On the other hand, others, such as the insurance companies themselves, are fighting for the right to continue its use in order to protect themselves from place-based risks that are not fully revealed in individuals' driving records but that may affect insurance costs (Friedman, 2006).

Developing a sound policy depends in part on determining the relative role of risk versus race and class in the existing place-based premium structure. Clearly, in this domain, race should have no role in setting such premiums because it would violate civil rights statutes prohibiting racial discrimination in civic and business life. Unfortunately, it is difficult to determine the relative contribution, if any, of these factors in the setting of insurance premiums because the research that the insurance companies use to support their position is treated as proprietary information, and thus not easily evaluated by third parties.

To fill the gap, this paper uses a unique data set assembled from numerous information sources to examine the relative influence of place-based socioeconomic characteristics and place-based risk factors on the spatial pattern of automobile insurance premiums in the city of Los Angeles. To examine this, we use a novel approach of combining tract-level census data and car insurance rate quotes from multiple companies for sub-areas within the city of Los Angeles. The quotes are for an individual with identical demographic and auto characteristics, driving records and insurance coverage. This method allows the individual demographic and driving record to be fixed. Multivariate models are then used to estimate the independent contributions of these risk and race and class factors to the place-based component of the car insurance premium. In sum, we find that both risk and "redlining" factors are associated with variations in insurance costs in the place-based component, with black and poor neighborhoods being adversely affected, although risk

factors are stronger predictors. However, even after risk factors are taken into account in the model specification, SES factors remain statistically significant. Moreover, simulations show that “redlining” factors explain more of the gap in auto insurance premiums between black (and Latino) and white neighborhoods and between poor and nonpoor neighborhoods.

The remainder of this paper is organized into four parts. Part I reviews the literature and includes our proposed approach to testing the two competing explanation for the spatial variation in premiums. Part II discusses the data on insurance premiums, risk and socioeconomic characteristics used in this study. Part III presents our findings, which show that both place-based risk and “redlining” affect the rate. The paper concludes with a discussion about the implications of the study.

Literature Review

Understanding why there is an extensive spatial variation in insurance premiums across urban neighborhoods is important because of the impact it has on disadvantaged people and neighborhoods. In Los Angeles, for example, insurance premiums can vary by a factor of two across neighborhoods for persons with identical driving records and coverage, with those living in the inner-city paying the highest rates (Ong, 2002). These variations have particular consequences. Higher car insurance rates affect one’s ability to own a car, especially for racial minorities and the poor such as welfare recipients. More importantly, for a variety of reasons, car ownership influences considerably ones’ ability to search for and find employment and stay employed, among other factors (Ong, 1996; Raphael and Stoll, 2001; Ong, 2002; Raphael and Rice, 2002).

Setting automobile insurance premiums is a complex (and proprietary) process that usually involves both individual and neighborhood factors. In general, insurers use correlations with categories of cars defined by data such as owner's residence zip code, credit score, prior or new insurance, home owning or not, occupation, single payment or installation plan, driver record, marital status, and very selectively, driver sex. Insurers theorize that the correlations distinguish cars with high-risk and low-risk drivers (Butler, 2004). These and other models of premiums as a function of risk are commonly understood and result from microeconomic theory in which the insurance premium is a product of the actuarially expected payment plus a load factor that accounts for additional risk, all of which are set in a competitive market.

While the use of individual characteristics, driving experience and driving records is well understood and generally accepted in the setting of car insurance premiums, the use of zip-code areas in setting such rates has been highly controversial and political. For example, California voters approved the 1998 Proposition 103, which states that auto insurance rates should be based primarily on three factors – the driver's driving safety record, the driver's annual mileage driven and the number of years of driving experience for the driver. This implies that the role of place in setting premiums should be reduced or eliminated. However, efforts to do so have been hindered by a number of factors including a state government administration sympathetic to the insurance company and its use of zip-code areas, and by court rulings that has allowed continued use of zip codes to complement the primary elements (Kristof, 2005). As a consequence, there are cases in which the geographic factor carries more weight than the individual factors in setting auto insurance premiums, leading to the wide spatial disparity in premiums cited earlier.

There are potentially two major explanations for the spatial variation in the premiums with respect to the place-based component. One major explanation of the higher car insurance premiums is discriminatory “redlining,” a practice of charging higher premiums for those residing in low-income, minority neighborhoods. Much of the literature on “redlining” has focused on the housing market and on home mortgages in particular. Much of this literature has found that “redlining” plays a role in the higher mortgage denial rates of minorities, especially before greater enforcement of the Home Mortgage Disclosure Act and the Community Reinvestment Act (Newman and Whyly, 2004; Nothaft and Perry, 2002; Reibel, 2000; Ross and Tootell, 2004; Zenou and Boccoard, 2000). Moreover, a review of industry underwriting and marketing materials indicates that race affects the policies and practices of property insurance industry (Squires, 2003).

Unfortunately, there are few systematic empirical tests of “redlining” in the automobile insurance market. Although suggestive and not definitive, one study of auto insurance rates in Los Angeles documents that holding constant individual drivers records, demographics, experience, and auto maker and type, inner city residents pay much higher insurance premiums than those in outer suburban areas (Ong, 2002). This study, however, does not account for potential differences in place-based risk. There is indirect evidence of “redlining” in the automobile industry more generally. All else equal, blacks are disadvantaged in the new car purchase bargaining process and on average pay more than other car buyers even after controlling for credit worthiness (Ayers, 1995). Similarly, other evidence demonstrates that blacks and other minorities are charged higher interest rates for new car loans once these car purchases are made, and that these higher interest rates are not justified by the higher credit risks of applicants (Cohen, 1993).

The other major explanation for the higher automobile insurance premiums faced by residence of disadvantaged communities is differential risk experiences across neighborhoods. While not necessarily denying that premiums are higher in disadvantaged neighborhoods, proponents of place-based rating argue that costs of insuring a vehicle are higher in some neighborhoods because of higher risks there, including higher claim and loss rates.

The number of studies that empirically examine risk-based claims is thin however. One study examined this question indirectly by investigating whether auto insurance market loss ratios are related to the percent of the population that is minority (Harrington and Niehaus, 1998). The results indicate that there are on average 36 percent more liability claims and 48 percent more collision claims per 100 covered persons in urban areas with a disproportionately high black population. This indicates that loss ratios are positively related to percent of the population that is minority. To the extent that auto insurance premiums are positively correlated with loss ratios as the auto insurance industries argues, this suggests that some of the higher auto insurance premiums in minority neighborhoods are partly accounted for by legitimate driving risks.

On first inspection, the aforementioned results seem contrary to the “redlining” hypothesis that racial discrimination increases premiums relative to expected claim costs for minorities. However, although auto related risks such as loss ratios may be higher in minority and poor areas and thus should drive higher car insurance rates there, the question is whether they are set above and beyond what is appropriate to compensate for these risks and whether insurance companies profit from these practices.

We propose to add to the literature by disentangling the two competing explanations for the higher premiums paid by those in poor and minority communities: the higher rates are due to higher legitimate costs of insuring residents in poor and minority communities as a results of the greater risks there, and the higher rates are the product of “red lining”.

Specifically, we test a simplified form of the following equation:

$$P_i = \alpha + \beta_1 X_i + \beta_2 Y_i + \beta_3 Z_i + \varepsilon_i, \text{ for observation } i \text{ where}$$

P_i is the insurance premium,

α is a constant,

β_1 is a vector of coefficients, and X_i is a vector of individual characteristics,

β_2 is a vector of coefficients, and Y_i is a vector of place-based risk,

β_3 is a vector of coefficients, and Z_i is a vector of place-based SES,

and ε_i is a stochastic term.

Unfortunately, we do not have access to individual insurance records, so we use a single hypothetical individual and gathered information on premium quotes for neighborhoods throughout a single large urban city. In other words, we hold X_i constant and allow P_i to vary across place and place-based factors:

$$P_i = \alpha + \beta_2 Y_i + \beta_3 Z_i + \varepsilon_i .$$

This approach has some limitations. Because the study does not include data for suburban and rural areas, it does not capture the potentially substantial premium variation between these geographic types.³ Moreover, because we rely on average insurance quotes for each neighborhood, we are not modeling how an individual insurance company sets rates. Instead, we are capturing the variations of the market as a whole, which is a

³ There are also very large differences across metropolitan areas and between rural and urban areas in auto insurance premiums. (Savage, 2006).

composite of potential disparate formulas used by many companies. Finally, because we have only one hypothetical person, we are not sure how premiums would vary for drivers with other characteristics. Nonetheless, given the paucity of prior research, this study is a major step forward in testing the relatively role of redlining and risk in influencing the spatial variation in automobile insurance rates.

Data and Bivariate Relationships

The data for this study come from several sources. Socioeconomic status (SES) factors related to the “redlining” hypothesis are based on tract-level data from the 2000 Census SF-3 (Summary file 3) data set. We include the percentage of the tract population that is black, Hispanic, and the percentage of the population that is below the poverty line as the main socioeconomic status variables.

Car insurance premiums, the outcome variable of primary interest, were collected over the internet for the year 2000 from multiple quotes from multiple insurance companies for each zip code in the city of Los Angeles. Insurance premium estimates were for the liability component only and were provided by the following website:

<http://www.realquote.com>. 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, employed single mother, who has been driving for 7 years, has taken a driver training course, has one moving violation, but no accidents, and does not smoke. She owns a 1990 Ford Escort LX, two-door hatchback with no anti-theft devices, no anti-lock brakes, and no airbags, and parks 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.

The risk variables included in the analysis are those that the insurance companies argue are related to higher risks. These include claim and loss rates as the primary risk factors and accident and crime rates as secondary risk factors. The data on insurance claims come from the California Department of Insurance (CDI). It contains data on private passenger auto liability from 2000 as well as the physical damage experienced broken down by coverage and program type and by zip code. The claim rate is defined as the number of claims per 10,000 insured persons in a zip code per insurance year.

The data on the amount of dollar loss for all insurance companies also come from the California Department of Insurance. The loss rate measures the average loss payments per insured vehicle year normalized by total number of exposures years across zip codes. Exposure years are defined as the number of car-years for which the insurance companies are liable.

The risk factors that insurance companies claim as secondary risk factors include the accident and crime rate. The accident data come from Los Angeles City Department of Transportation (LADOT). These files document all reported traffic collisions in the city of Los Angeles from 1994 through 2002, with a small number of records for 1993 and 2003. The data are thus restricted to 1994-2002, as the dataset is most complete for these years. The accident locations were geocoded and summed within census tracts, and we use a buffer to capture the accidents that occur on the streets that are on the tract boundaries. From this we defined the accident rate, which is measured as the number of accidents normalized by the area size (in square kilometers) of each tract.

The car crime data was provided by the city of Los Angeles Department of Police (LAPD) for reporting districts (RD) for the year 2000. RDs are the geographical units used by the Police Department to aggregate crime data, and many of these units are equivalent to census tracts. The crime variable used in this study is the aggregated number vehicle thefts. The variable is also normalized by the area size (in square kilometers) of each RD.

Since the insurance data are based on zip codes and the other variables based on census tracts, the former variables are transformed into census tracts by using GIS proportional split method. We used the proportional split method to redistribute spatial values throughout a specific area. The assumption of the method is that population is distributed evenly over a geographic surface. This is not always the case, but it serves as an inexpensive method for generating estimates quickly and impartially without resorting to more accurate, costly techniques. Because the method assumes uniform dispersion of attributes within each zip code, it might cause a problem when there is an extreme unbalance of distribution of attributes (Schlossberg, 2003). However, an analysis of comparing total numbers and patterns of maps indicates that these concerns are limited.

Table 1 reports the averages and standard deviations for the dependent (premium) and independent variables (risk and SES) in the analysis. The average car insurance premium in the city of Los Angeles over the study period for the typical driver described above is \$1,041 dollars, with the median estimate slightly lower because of high top end premiums that are pulling up the average estimates. The main risk factors indicate an average claim rate of 220 and an average loss rate of \$212 in the sample. The secondary risk factors show average accident and crime rates of 1,822 and 279, respectively. Finally, consistent with the demographic composition of the city of Los Angeles, the mean poverty

rates is 22.4 percent while the percentage of the population that is black or Latino is 11 and 43 percent, respectively.

More importantly, Table 2 examines the extent to which insurance premiums vary with these place-based characteristics. To do this, for each independent variable on the left hand column, we divide the data into terciles ranked by the values of these respective independent variables. Then, for each tercile category (for each independent variable) we calculate the average car insurance premium. The table clearly shows that insurance premiums are much higher in areas characterized by greater risk. In particular, in areas with greater claim and loss rates, insurance premiums are higher. The same is also true for the accident and crime rate measures of risk. The difference in the insurance premiums for most of the risk variables is about a \$100 or so from the bottom to the middle or from the middle to the top tercile category.

These observed relationships between the risk factors and insurance premiums are verified in the correlation estimates provided in the right side column of the table. The correlations between the primary and secondary risk factors and the insurance premiums are all fairly large in magnitude and statistically significant. More importantly, the strength of the association between the risk factors and premiums are slightly stronger for the primary than secondary risk factors as we should expect.

At the same time, the “redlining” factors are also correlated with premiums. The correlation values in the last column in Table 2 clearly show that insurance premiums are higher in areas characterized by higher poverty rates and where the percent of the population that is black is higher, as indicated from the bottom to the top of the respective tercile categories. The association between %Latino and insurance premiums is weakly

related however, perhaps because of the wider spatial dispersion of this population (than for blacks) and the existence of small Latino enclaves close to non-Hispanic white neighborhoods. Still, the rate of the increase in premiums moving from the bottom to the top terciles for the SES variables (especially for %black) is not as great as that for the risk factors. This is confirmed in the bivariate correlation estimates between the SES factors and insurance premiums. While the correlation estimates are positive and statistically significant between all SES factors and the insurance premiums, they are much more weakly associated than between the primary (and even secondary risk factors) and these premiums.

Alternatively, Table 3 divides the insurance premium data into terciles and then for each insurance premium tercile category, the averages for the risk factor and SES variables are calculated. The table shows that the primary risk factors, such as the claim and loss rates, are higher in the areas where the insurance premium is higher, as are the secondary risk factors, i.e., the crime and accident rates. For example, the average accident rate for the tracts in the most expensive insurance areas is nearly three times higher than the average accident rate for the tracts in the least expensive insurance areas. There are also significant, albeit smaller, differences in the loss rate and crime rate across these insurance premium terciles. Combined with the data above, these bivariate relationships are consistent with the hypothesis that insurance rates are set by risk factors. Still, poverty rates are higher and the percentage of the population that is black is higher in areas with more expensive premiums, providing evidence again that race and class factors may also play a role in setting auto insurance policies.

However, estimating the contribution of the risk factors and the SES factors to the place-based premium structure cannot be accomplished simply by estimating the bivariate correlations. This is true partly because for most instances, the risk and SES factors are correlated themselves thus confounding these relationships. Table 4 shows data on the simple bivariate correlations among and between the risk factors and SES variables. The table shows that, for example, the poverty rate is correlated with all the risk factors, implying that the correlation between poverty status and insurance premiums may be statistically spurious due to the collinearity between poverty and risk.

On the other hand, the percent of the population that is black variable is either not correlated or only weakly correlated with the risk variables. This may indicate that the observed bivariate relationship between %black and insurance premiums is much more robust than that shown above. Finally, the percent of the population that is Latino variable is correlated with most of the risk variables, but not in a consistent direction. This pattern may contribute to the weak relationship between the percent of the population that is Latino and insurance premiums that we observe. Given potential confounding effects of collinearity, in the next section we pursue multivariate techniques to test and estimate the independent effects of risk and SES factors on the place-based component of the insurance premium.

In sum, the previous section demonstrates somewhat sharp insurance premium differences across areas characterized by different levels of the risk related factors and by different levels of SES characteristics of areas as well. Because the SES and risk related factors are largely correlated themselves, it is difficult to estimate the independent contribution of the risk or SES factors to the place based component of auto insurance

premiums. We address this problem by employing regression analysis where the dependent variable is the auto insurance premium.

The empirical strategy in the regression analysis is to highlight the sensitivity of SES variables to the inclusion of risk related variables into the model specification. Following this strategy, we first present a naïve model including only the SES factor. Then we systematically include and exclude the primary risk factors and secondary risk factors into the model specification to estimate these independent associations.

Multivariate Findings

We use ordinary least squares (OLS) regressions to estimate the influence of the “redlining” factors and risk factors on insurance premiums. The first model examines the “naïve” “redlining” model in which only the SES factors are included. Then we examine the “naïve” primary risk factor models where we include only the claim rate and loss rate variables. Next, we examine a model in which both the “redlining” and primary risk factors are included to examine the independent effect of these factors on auto insurance rate premiums. Model 4 includes only the secondary risk factors such as the accident rate and crime rate. Model 5 includes the primary and secondary risk factor variables to examine their independent contribution on auto insurance premiums since these factors are highly correlated. Finally, Model 6 includes all “redlining” and risk factor variables in the analysis. Table 5 reports the estimated coefficients for these models.

Model 1 is the “naïve” redlining model, where only the SES factors are included. All coefficients are highly statistically significant and the model explains about a quarter of the variation in the place-based component of the auto insurance premium. The percent of

the population that is comprised of people below the poverty line and the percent comprised of blacks are positively related to the insurance premium. In particular, living in a poorer neighborhood (25% poverty rate) would increase insurance premiums by about \$116 relative to living in a neighborhood that has few in poverty (5% poverty rate). The estimated difference between an all black neighborhood and one without any blacks is \$112.

Interestingly, the estimated coefficient for the percent of Latino is negative. Again, this may be due to the fact that Latinos are more widely spread in the city of Los Angeles as compared to African American residents and thus offer little opportunity to “redline.” However, though not shown here, in models where the poverty rate is interacted with the percent of the area that is Latino, that coefficient reveals a positive, statistically significant relationship with the insurance premium, though small in magnitude. Moreover, this statistically significant coefficient is not explained away by either the primary or secondary risk factors when they are included in the model specification. This evidence is consistent with “redlining” in high poverty predominantly Latino areas. No interaction effects were found between the poverty rate and the percentage of the area that is black.

Model 2 includes only the primary risk factors deemed important by insurance companies. The results indicate that all coefficients are positive, highly statistically significant and consistent with expectations based on the risk hypothesis; that is, premiums are higher in neighborhoods with higher driving risks. The model explains about 51% of the variation in premiums, fairly large for a cross-sectional model. The coefficient for the linear loss-rate variable indicates that each additional dollar loss leads to a \$0.15 increase in premium, but the coefficient is not statistically significant. Model 2b includes the square of the loss rate since we assume that there is a curvilinear trend between premiums and the

loss rate. The results indicate that the premiums increase with the loss rate but at a slower rate at the high end of the distribution. The inflection point is at \$226 dollars, which is just above the mean value of this independent variable. In other words, premiums start to decline in the upper range of the losses. (Interestingly, the average loss value is lower in predominantly black neighborhoods and very poor neighborhoods than in non-Hispanic white neighborhoods that have very low poverty rates. These types of neighborhoods are discussed later.)

Model 3 includes both the “redlining” and primary risk factors and indicates that 62 percent of the variation in insurance premiums is explained by these combined factors. More importantly, the results show that both risk and SES are statistically significant, providing evidence of both the “redlining” and risk hypothesis of setting auto insurance premium levels. But with the inclusion of the primary risk variables into the model specification, the coefficient estimate on the percent poverty decreases by 47 percent, indicating that much of the positive relationship between the poverty rate of an area and the insurance rate premium is driven by the higher primary risk factors in such areas. Interestingly, the coefficient estimate on percent black becomes slightly, but not statistically significantly larger, indicating that primary risk factors are not driving the higher insurance premiums in areas where the percentage of the population that is black is larger. Finally, the negative coefficient on the percent Latino variable becomes less negative with the inclusion of the primary risk factor variables suggesting that much of the negative relationship between the poverty rate of an area and the insurance rate premium is influenced by lower primary risk factors in such areas.

Model 4 includes only the secondary risk factors. The coefficient estimates are highly statistically significant and demonstrate that both the accident rate and crime rate of an area are positively associated with the place-based component of the auto insurance premium. The point estimate for the accident rate indicates that an increase in accidents over a square kilometer by 1,000 is associated with an increase the car insurance premium of about \$40.00.

Model 5 includes only the primary and secondary risk factors into the model specification. The inclusion of these variables in most cases leads to slightly lower point estimates of the association of the primary and secondary risk factors and the insurance premiums. Still, when both are included, all risk factors remain positive and statistically significantly related to insurance premiums.

Model 6 includes the SES variables and both the primary and secondary risk factors into the model specification. About 64 percent of the variation in insurance premiums is explained with all variables in the analysis included. In addition, the results show that both the SES and most risk factor variables are statistically significant, providing strong evidence of both the “redlining” and risk hypothesis of setting auto insurance premium levels. With the inclusion of the primary and secondary risk variables into the model specification, the coefficient estimate on the percent poverty decreases by 32 percent from the coefficients estimate in Model 3, indicating that much of the positive relationship between the poverty rate of an area and the insurance rate premium is driven by the higher primary and secondary risk factors in such areas. However, like the results in Model 3, the coefficient estimate on percent black becomes slightly larger with the inclusion of the

secondary risk factors into the specification suggesting that like the primary risk factors, the secondary risk factors are not driving the higher insurance premiums in black areas.

To estimate the relative role of “redlining” and risk, we use results from Model 6 to simulate insurance premiums for four types of neighborhoods: (1) non-Hispanic white (NHW) and non-poor tracts where blacks, Latinos and those below poverty individually comprise no more than 5% of the population, (2) black tracts where blacks comprise at least 75% of the population, (3) Latino tracts where Latinos comprise at least 75% of the populations, and (4) poor tracts where those below the poverty line comprise at least 40% of the population. The first set of simulations is based on multiplying the parameters from Model 6 by the average demographic and risk values for each of the four types of neighborhoods. To estimate the impact of differences in risk, the neighborhood-specific average risk values are replaced by the average risk values for NHW and non-poor tracts. Finally, to estimate the impact of differences in demographic characteristics, the neighborhood-specific average demographic values are replaced by the average demographic values for NHW and non-poor tracts. Figure 1 reports the results.

Figure 1 indicates that the relative importance of “redlining” and risk varies by neighborhood type. The estimates for black neighborhoods reveal that differences in risk account for a very small part of the higher insurance premiums relative to that for NHW and non-poor neighborhoods. On the other hand, the differences in demographic composition explain an overwhelming majority of the gap. That is, for an overall gap of \$154 in insurance premiums between these areas, about 11 percent of the gap is explained by risk factors while 89 percent of the gap is explained by differences in demographic characteristics. For Latino neighborhoods, the premium gap with NHW and non-poor

neighborhoods is relatively small (\$48), but that differences in demographic composition explain more of the gap (73 percent as opposed to 27 percent of the gap being explained by risk). The premium gap for poor neighborhoods is similar to that encountered by black neighborhoods (\$156), but risk factors explain slightly more of the gap (45 percent) in comparison to black neighborhoods. Still, like in black neighborhoods demographic differences explain more of the gap between poor and the nonpoor neighborhoods (55 percent) than risk factors. Taken as a whole, the simulations indicate that “redlining” contributes more to higher premiums in black, poor and Latino neighborhoods than risk.

In sum, the results indicate that both risk and SES factors are associated with auto insurance premiums in the expected direction. Based on the analysis of the adjusted R square of our models, the risk factors are statistically more important determinants of the premiums, but the simulations indicate that “redlining” explains more of the gap in the premiums across black (and Latino) and white neighborhoods and between poor and nonpoor neighborhoods. These results provide evidence of both the “redlining” and risk hypotheses in the setting of such premium levels.

Nonetheless, there is concern of whether the coefficient estimates of the SES variables reflect potential “redlining” of poor and/or black areas or something else. For example, it could be the case that those with worse driving records (or those who might pose the highest driving risk) are concentrated in poor and/or black areas thus leading to higher auto insurance premiums there. However, this concern is tempered because the insurance premium estimate is derived from an individual with identical demographic and auto characteristics, driving records and insurance coverage across areas in Los Angeles.

Thus, this method allows the individual demographic characteristics and driving record to be fixed and therefore rules out unobserved heterogeneity that might drive these results.

Moreover, the empirical evidence on racial differences in car ownership rates also works in the opposite direction of the hypothesis that sorting by driving ability across neighborhoods is driving the results shown here (particularly for the coefficient on % black). Car ownership rates are much lower for blacks than whites and thus black car drivers are much more positively selected across education and income groups than are other drivers (Raphael and Stoll, 2002). To the extent that education and income are correlated with driving ability (as even the insurance companies claim and are reflected in the individual portion of the insurance premium), black car drivers are likely to be more positively selected than others.

Another concern related to the overall results of the model deals with whether the regressions should be weighted. The results shown here are for unweighted regressions, which is appropriate because we essentially have a single hypothetical person per area. However, there are variations in the number of policy holders across neighborhoods. To address this concern, we also estimated models where the observations are weighted by the square root of the estimated number of automobile insurance premiums in each tract, an approach normally used to correct of any potential problem with heteroscedasticity. Though not shown, the results of these regressions are qualitatively similar to those shown here. The coefficients for two of the place-based risk factors (loss rate, claim rate and accident rate loss) and two of the SES factors (percent in poverty and percent black) are significant and consistent with a priori expectations. However, both the crime rate and the

percent Latino are not statistically significant. Despite these differences, the overall qualitative findings remain the same.

Conclusion

In this paper, we used the best data available to academic researchers (not working directly for the insurance industry) and a novel approach to examine the relative importance of risk and “redlining” factors in the setting of the place-based component of auto insurance premiums. The basic results showed that both risk and “redlining” factors are associated with auto insurance premiums in the expected direction, although the risk factors were found to be statistically more important determinants of the premiums. However, even after risk factors are taken into account in the model specification, SES factors remain statistically significant. Moreover, simulations of the regression equations demonstrate that “redlining” factors explain more of the gap in the premiums across black (and Latino) and white neighborhoods and between poor and nonpoor neighborhoods.

There is probably less controversy surrounding the findings related to risk. Most reasonable people would agree that premiums should vary with risk and that insurance companies should and do take this into consideration. Certainly, there is much more controversy surrounding the findings related to “redlining.” The “redlining” hypothesis does not dispute the role of risk, but does argue that there are discriminatory practices that influence premiums. This could include a hypothetical practice of statistical discrimination, in which insurance companies use controversial methods such as examining the racial or poverty composition of neighborhoods to identify risky neighborhoods; that is, the perception that low-income minority neighborhoods are on average more risky, thus

leading to the companies setting higher rates in these areas. Because statistical discrimination is based on some factual basis, there would still be a correlation between the rate and risk factors. However, the problem occurs when the premium rate is set well above and beyond the level to truly compensate for these risks and as a consequence, insurance companies profit from these methods. At the same time, such an approach would also taint low-income minority neighborhoods that are in fact not risky.

Based on this analysis, we cannot determine the exact practices insurance companies use in setting rates. Moreover, companies refuse to give detailed information about their practices because they claim that the information is proprietary. Our analysis tests whether the observed spatial patterns are consistent with either or both hypothesized rate setting practices. Again, the finding that “redlining” factors play a role beyond risk is noteworthy and is likely to be questioned. Of course, there may still be unobserved risk variables that matter in setting premiums and that may be collinear with the SES variables, but that are not included in our models. However, the models used in this analysis include the set of risk factors that auto insurance companies claim are important in setting premiums. Moreover, without greater access to information, particularly individual insurance records, it is difficult to do additional analyses. Given this constraint, the above analysis should be treated as the current and standing findings until additional evidence that can be verified becomes available.

Still, the findings have implications. They show that automobile insurance rates are higher in low-income and minority neighborhoods because of both direct and indirect effects; that is, these neighborhoods tend to have higher risk rates, and even after accounting for this, there is a residual adverse impact in the setting of higher premiums

there. This story is truer for residents in predominantly black neighborhoods. Residents in African American neighborhoods pay more, ceteris paribus. This can be interpreted as the discriminatory cost of simply living in such areas. Residents in poor neighborhoods pay higher premiums as well but almost entirely because neighborhood driving risks are greater there. The story for Latinos is more complicated. The Latino variable is either negatively related or unrelated to premiums (unweighted regression and weighted regression results, respectively), indicating no discriminatory practice. Moreover, the percent Latino is not consistently related to the risk factors, so there are offsetting indirect effects through these variables. However, predominantly Latino neighborhoods tend to have very high poverty rates (McConville and Ong, 2003), thus they tend to experience higher insurance costs because of the role of poverty in setting premiums.

Regardless of how the insurance rates are set, residents in disadvantaged neighborhoods suffer direct and indirect impacts of a higher insurance premium. The direct cost is the higher out-of-pocket expense for auto premiums, which can equal a thousand dollar per year for basic coverage, an extremely high amount for those with low-incomes. When insurance is prohibitively high, some residents drive without insurance, which places them and others at greater personal financial risk (Bernstein, 1999). There are also indirect effects because higher premiums can be a barrier to automobile ownership, which in turn limits access to social activities, services and economic opportunities that are dispersed throughout the metropolitan region. One well documented impact on inner city residents' lower ability to own cars is the adverse impact on the ability to gain employment (Ong, 2002; Blumenberg and Ong, 1998; Raphael and Rice, 2002; Raphael and Stoll, 2001; Ong and Miller, 2005). In turn, those forced to rely on public transit, have very limited mobility

for a variety of reasons (Taylor and Ong, 1995; Ong and Houston, 2002; Stoll, Holzer and Ihlanfeldt, 2000). The spatial variation in insurance premiums, then, adversely affects disadvantaged neighborhoods in a complex way that is not immediately apparent. These impacts make addressing the spatial inequality of auto insurance premiums a matter of public policy.

References

- Ayers, Ian (1995) "Further Evidence of Discrimination in New Car Negotiations and Estimates of Its Cause," *Michigan Law Review* v94(1), pp: 109-147.
- Bernstein, Robert (1999). "California Uninsured Vehicles as of June 1, 1997," California Department of Insurance, Sacramento, February 1999.
- Blumenberg, Evelyn, and Paul M. Ong (1998) "Job Accessibility and Welfare Usage: Evidence from Los Angeles," *Journal of Policy Analysis and Management*, 17(4): 639-657.
- Butler, Patrick (2004). "An alternative to the high-risk-driver theory: adverse selection induced by per-car premiums," Paper presented August 9 at the 2004 Annual Meeting of the American Risk & Insurance Association in Chicago.
- Cohen, Alma (2005). "Asymmetric information and learning: Evidence from the automobile insurance market," *Review of Economics and Statistics* 87 (2): 197-207.
- Cohen, Mark A., 2003, "Report on the Racial Impact of GMACs Finance Charge Markup Policy," Expert Testimony/Report on the Matter of Addie T. Coleman, et. al., Vs. General Motors Acceptance Corporation (GMAC), found at: <http://www.consumerlaw.org/initiatives/cocounseling/content/GMAC/CohenReport.pdf>, on February 5, 2006.
- Ferguson, Michael F. and Stephen R. Peters (2000). "Is Lending Discrimination Always Costly?" *The Journal of Real Estate Finance and Economics* 21(1): 23 – 44.
- Friedman, Josh (2006). "Steadfast Believer in Numbers: The State's Plan to Revamp the Setting of Car Insurance Rates Doesn't Add Up for Mercury's George Joseph," *Los Angeles Times*, Business Section, February 5, 2006.
- 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.

- Houston, Douglas, Jun Wu, Paul Ong and Arthur Winer (2004). Structural Disparities of Urban Traffic in Southern California: Implications for Vehicle Related Air Pollution Exposure in Minority and High Poverty Neighborhoods, *Journal of Urban Affairs* 26(5): 565-592.
- Hunstad, Lyn (1996). "Methodology and Data Used to Develop the California Private Passenger Auto Frequency and Severity Bands Manual," California Department of Insurance, April.
- Klein, Robert W. and Martin F. Grace (2001). "Urban Homeowners Insurance Markets in Texas: A Search for Redlining," *The Journal of Risk and Insurance* 68 (4): 581-613.
- Kristof, Kathy M. (2005). "Insurance Plan May Shift," *Los Angeles Times* December 23.
- Newman, Kathe and Elvin K. Wyly (2004). "Geographies of Mortgage Market Segmentation: The Case of Essex County," *New Jersey Housing Studies* 19 (1): 53-84.
- Nothhaft, Frank E. and Vanessa G. Perry (2002). "Do mortgage rates vary by neighborhood? Implications for loan pricing and redlining," *Journal of Housing Economics* 1 (3): 244-265.
- McConville, Shannon and Paul Ong (2003) *The Trajectory of Poor Neighborhoods in Southern California, 1970-2000*, Washington DC: Metropolitan Policy Program, The Brookings Institution.
- Ong, Paul (1996). "Work and Automobile Ownership Among Welfare Recipients," *Social Work Research*, 30(4):255-262.
- 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.
- Ong, Paul and Douglas Miller (2005). "Spatial and transportation mismatch in Los Angeles," *Journal of Planning Education and Research* 25 (1): 43-56.
- Ong, Paul and Hyun-Gun Sung (2003). "Exploratory Study of Spatial Variation in Car Insurance Premium, Traffic Volume and Vehicle Accidents," UCTC working paper no. 654.
- Pastor, Manuel, Jr., James L. Sadd, and Rachel Morello-Frosch (2004). "Waiting to inhale: the demographics of toxic air release facilities in 21st-century California," *Social Science Quarterly* 85 (2): 420-441.

- Pulido, Laura (2000). "Rethinking Environmental Racism: White Privilege and Urban Development in Southern California," *Annals of the Association of American Geographers* 90 (1): 12-40.
- Raphael, Steven, and Lorien Rice (2002). "Car Ownership, Employment, and Earnings." *Journal of Urban Economics* 52(1): 109-30.
- Raphael, Steven and Michael A. Stoll (2001). "Can Boosting Minority Car Ownership Rates Narrow Inter-Racial Employment Gaps?" *Brookings-Wharton Papers on Urban Affairs*, 2: 99-137.
- Reibel, Michael (2000). "Geographic variation in mortgage discrimination: evidence from Los Angeles," *Urban Geography* 21 (1): 45-60.
- Ross, Stephen and Geoffrey M.B. Tootell (2004). Redlining, the Community Reinvestment Act, and Private Mortgage Insurance, *Journal of Urban Economics* 55(2): 278-297.
- Savage, Mark (2006). "Auto Insurance Premium Vary Widely by Zip Code for Good Drivers Across California," Consumers Union of United States, San Francisco, CA.
- Squires, Gregory D.(2003). "Racial Profiling, Insurance Style: Insurance Redlining and the Uneven Development of Metropolitan Areas," *Journal of Urban Affairs* 25 (4): 391-411.
- Squires, Gregory D., Sally O'Connor, and Josh Silver (2001). "The unavailability of information on insurance unavailability: Insurance redlining and the absence of geocoded disclosure data," *Housing Policy Debate* 12 (2): 347-372.
- Stoll, Michael (2005). "Geographical skills mismatch, job search and race," *Urban Studies* 42 (4): 695-717.
- Stoll, Michael A., Harry J. Holzer and Keith R. Ihlanfeldt (2000). "Within Cities and Suburbs: Racial Residential Concentration and the Distribution of Employment Opportunities Across Sub-Metropolitan Areas" *Journal of Policy Analysis and Management*, 19(2): 207-231.
- Taylor, Brian D., and Paul M. Ong (1996). "Spatial Mismatch or Auto Mobile Mismatch: An Examination of Race, Residence and Commuting in U.S. Metropolitan Areas," *Urban Studies*, 32(9): 1453-1473.
- 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>.

Zenou, Yves and Nicolas Boccoard (2000). "Racial Discrimination and Redlining in Cities,"
Journal of Urban Economics 48 (2): 260-285.

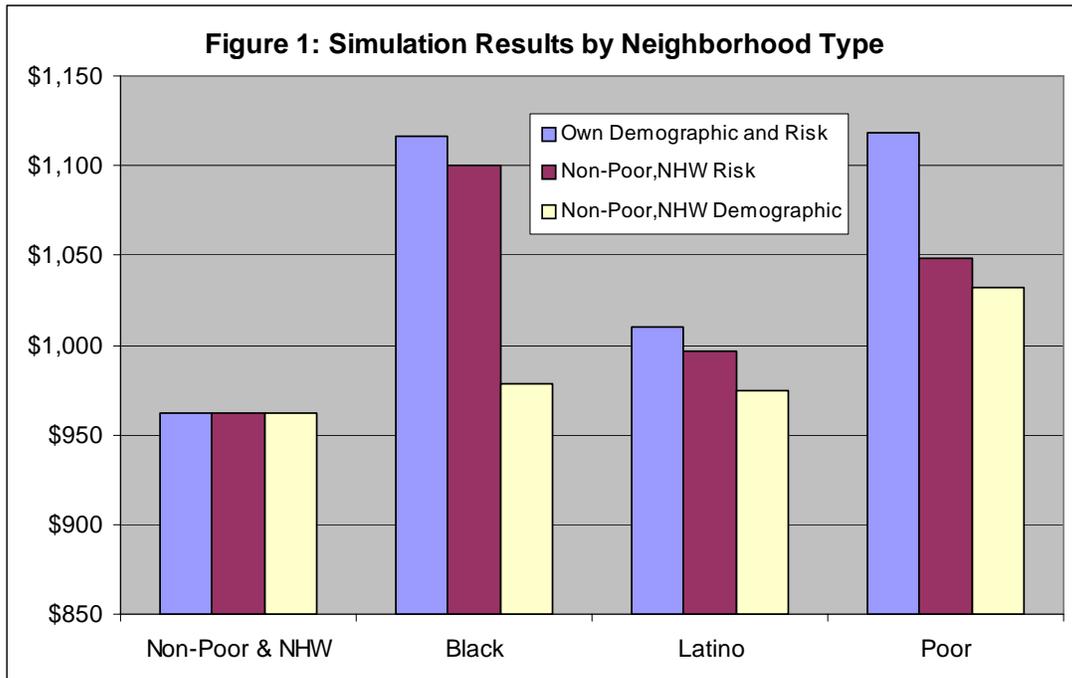


Table 1. Descriptive Statistics

	Variables	Mean	Median	Std Deviation
Dependent Variable	Premium	1,041.4	1,028.1	151.5
Neighborhood Socioeconomic Status	% Poverty	22.4	20.4	14.2
	% Black	10.8	4.1	16.7
	% Latino	43.2	44.3	27.7
Insurance-based Risk	Claim rate	220.1	212.9	48.4
	Loss rate	212.7	204.8	53.3
Other Risk Factors	Accident	1,821.7	1,359.0	1,659.0
	Crime	279.2	221.3	235.1

Table 2. Average Insurance Premiums by Risk and SES Characteristics of Neighborhoods (Trisected by each variable)

Mean	Insurance Premium			Correlation to Premium
	Bottom 3rd	Middle 3rd	Top 3rd	
% Poverty	977.4	1017.9	1128.2	0.4089***
% Black	998.6	1034.3	1090.9	0.2677***
% Latino	1015.8	1059.5	1048.8	0.0761*
Claim Rate	903.8	1060.4	1159.2	0.7139***
Loss Rate	939.4	1019.9	1164.1	0.6230***
Accident Rate	933.4	1033.0	1156.9	0.6143***
Crime Rate	945.4	1033.9	1144.2	0.5312***

P-value: * p<0.05; ** p<0.01; *** p<0.001

Table 3. Average Risk and SES Characteristics by Insurance Premium Levels of Neighborhoods (Trisected by Insurance Premium)

Mean	Bottom 3rd	Middle 3rd	Top 3rd	Correlation to Premium
% Poverty	16.8	18.5	31.6	0.4089***
% Black	4.5	9.9	17.8	0.2677***
% Latino	47.5	32.4	49.4	0.0761*
Claim Rate	180.7	218.8	259.7	0.7139***
Loss Rate	174.4	211.4	251.1	0.6230***
Accident Rate	802.4	1499.8	3125.0	0.6143***
Crime Rate	155.5	231.8	445.5	0.5312***

P-value: * p<0.05; ** p<0.01; *** p<0.001

Table 4. Correlations among Variables

	% Poverty	% Black	% Latino	Losses	Accidents	Crime
% Poverty	1	0.2273***	0.6328***	0.1544***	0.5170***	0.5840***
% Black	0.2273***	1	-0.0773*	0.0054	0.0261	0.0848*
% Latino	0.6328***	-0.0773*	1	-0.1059**	0.2576***	0.2523***
Claims Rate	0.2724***	0.0772*	0.0538	0.8530***	0.6361***	0.5625***
Losses Rate	0.1544***	0.0054	-0.1059**	1	0.5463***	0.4553***
Accidents Rate	0.5170*	0.0261**	-0.2576***	0.5463***	1	0.7888***
Crime Rate	0.5840***	0.0848*	0.2523***	0.4553***	0.7888***	1

P-value: * p<0.05; ** p<0.01; *** p<0.001

Table 5. Regression Result – Dependent Variable: Insurance Premium

	Model 1	Model 2a	Model 2b	Model 3	Model 4	Model 5	Model 6
Constant	960.2	548.4	295.7	267.1	933.2	278.1	298.7
% Poverty	5.822***	N/A	N/A	3.106***	N/A	N/A	2.127***
% Black	1.122***	N/A	N/A	1.207***	N/A	N/A	1.410***
% Latino	1.416*** ⁻	N/A	N/A	-0.409*	N/A	N/A	-0.358*
Claim rate	N/A	2.097***	2.225***	1.778***	N/A	1.588***	1.467***
Loss rate	N/A	0.147	2.264***	2.644***	N/A	3.173***	2.866***
Loss rate squared	N/A	N/A	0.005*** ⁻	0.005*** ⁻	N/A	0.007*** ⁻	0.006*** ⁻
Accident rate	N/A	N/A	N/A	N/A	0.047***	0.023***	0.023***
Crime rate	N/A	N/A	N/A	N/A	0.079**	0.061*	-0.014
Number of Observation	836	836	836	836	836	836	836
Adjusted R-Square	0.233	0.509	0.530	0.619	0.382	0.592	0.644

P-value: * p<0.05; ** p<0.01; *** p<0.001