

**A LOOK AT “RACE MATTERS: THE CONCENTRATION OF PAYDAY LENDERS IN AFRICAN-AMERICAN NEIGHBORHOODS IN NORTH CAROLINA”\***

By Diego Saltes  
Research Director, Community Financial Services Association  
dsaltescfsa@multistate.com  
September, 2005

**CRITIQUE HIGHLIGHTS**

- **The data collection method used in this study appears suspect because the authors do not fully explain the key steps taken to build their store database.** The few details that are mentioned provide the impression that the database used for the analysis may contain significant inaccuracies which would invalidate any conclusions based on assumptions about the true population of payday loan stores in North Carolina.
- **The study does not base its hypothesis on standard economic theory.** Standard theory would be unlikely to suggest that race and ethnicity are part of the rational decision-making process followed by payday loan companies to decide store locations.
- **The regression analysis used by the authors does not take into account the date in which the stores first opened.** This is a relevant factor since the racial composition of the neighborhoods may have changed after the stores began operating at their current locations.
- **The regression model in this study appears to suffer from multicollinearity, which would severely affect the output statistics, likely leading to incorrect conclusions.** Because several explanatory variables in the model are related, it's no surprise that one of the results was counterintuitive; this is a typical consequence of multicollinearity. Moreover, variables known to form the profile of payday loan customers did not make significant contributions to the regression model. These are unreasonable results and therefore clear red flags regarding the accuracy of the statistics put forth in the “Race Matters” study.
- **The authors admit having omitted at least two relevant explanatory variables from the model.** In a regression, finding a relationship that isn't actually there – as the authors of this study found – can be a consequence of omitted variable bias.
- **As a result of these troublesome calculations, the conclusion that race and ethnicity explain stores’ locations is likely the product of biased and inaccurate estimates, and therefore should not be upheld.**

---

\* Published by the Center for Responsible Lending, March 22, 2005

## **PURPOSE OF THE STUDY**

This paper intends to show that payday lending stores target communities in North Carolina where the share of African-American residents is relatively high. Specifically, the paper states that the concentration of payday stores in North Carolina communities with the highest share of African Americans is as much as 5 times higher than in communities where the share of African Americans is lowest (as ranked by US Census tracts).

## **BACKGROUND**

### ***Neighborhoods' Racial and Ethnic Composition***

As it will be explored later, it would be very difficult to prove that the main reason payday advance companies choose store locations is the racial composition of neighborhoods. Other independent factors, which may be correlated with race and ethnicity, are likely to be the true explanatory variables as it relates to the choice of location for payday loan stores. Moreover, to fairly judge the variables affecting the choice of store locations, the date in which the store first opened should be considered. The racial composition of neighborhoods often changes overtime, so many stores may have opened when the racial mix in those areas was significantly different.

Finally, the racial composition of people who live nearby a payday loan store does not necessarily imply that the store's customers have the same characteristics. For instance, a store that is located in an area with many commercial establishments is likely to provide many of its loans to people who work, but not necessarily live, close to that store.

### ***Market Drivers for Store Location***

An important consideration is that the economic reasoning behind the decision to locate a payday advance store is in all likelihood driven by actual market conditions (i.e., unfulfilled demand for the payday advance product). Assuming that it could be shown that race and ethnicity are to some degree proxies for the relevant variables determining payday loan demand as backed by economic theory, why would companies use these proxies instead of the relevant, readily available socioeconomic indicators?

Any market analysis of the payday loan industry that uses these proxies instead of the appropriate variables would likely fail to achieve optimal results. Household income, marital status, number of children in the household, and the unemployment rate are easily accessible data for virtually any community in the U.S. It would only make sense to use proxies if these variables were not available or suffered from accuracy problems, severe lags, etc. Race and ethnicity data often do suffer from respondent bias, and are only available as estimates that cannot be updated as often as the relevant indicators mentioned above.

### ***Predictors and Causality***

The authors conclude the background section by citing a study by Buckey and Simkins, which states that "after controlling for a number of variables, race is a powerful predictor of the location of payday lenders." An important point to be made about predictors is causality. Do Christmas cards "predict" Christmas? Obviously, Christmas card sales must precede the holiday, but they certainly do not *cause* Christmas.

In the same light, do large shares of minority residents cause the payday loans to be located where they are? The factors that economic theory asserts are relevant for the payday store location decision process surely appear to have more causal attributes than race and ethnicity.

## **DISCUSSION OF FINDINGS**

### ***Data Collection***

The authors explain that since the latest available official database of licensed payday lenders from the North Carolina Commissioner of Banks was for 2000, they resorted to their own data collection method to assemble a list of stores that operate in the state. The method used to identify the stores consisted of reviewing "company websites, newspaper articles, company advertisements, and advocates' reports."

It is further stated that the authors submitted the names on their list "to a phone database to obtain 2,982 telephone numbers and shop addresses in 15 states." It is not clear why the authors chose 15 states (as supposed to, say, 5, 10, 20 or all states), or which states those were, besides North Carolina. It is also not clear how many total establishments were submitted to obtain the 2,982 store list (presumably a higher number since it's likely that contact information could not be obtained for all the submitted establishments).

The authors indicate that they randomly selected 200 stores for follow-up calls to verify contact information and payday loan availability. However, the study does not mention why 200 stores was a good number to verify (i.e., based on statistical significance). Also omitted is any mention of the proportion of stores that could not be reached, declined to provide a response, or simply stated that they did not provide the payday loan product. It appears that the authors make the tacit statement that all the stores they contacted verified that their addresses were correct and that they indeed offered payday loans, which would be commonly unlikely.

The paper states that the above method "ultimately yielded the addresses of 385 payday loan storefronts in North Carolina." The authors admit, "while the list of 385 store locations may not be comprehensive, we have no reason to believe that our methods induce distortions along racial or ethnic lines. Moreover, it is more than double the 170 stores reported by Stephens, Inc. to be operating in North Carolina... Still, to the extent that our data set is a sample as supposed to a complete census, our statistical methods allow us to extrapolate findings."

There are actually several reasons to believe this methodology is not optimal. First, there are no statements regarding the statistical confidence level or margin of error corresponding to the sample used for the analysis.

The fact that the authors' list yielded a number of stores that is more than double a previous estimate is actually discouraging given that information was verified only for a small number of stores in North Carolina (assuming it was some share of the 200 verifications). The fact that the total number of stores was so high actually suggests the possibility that a number of outlets on the list may not actually be offering payday loans. That is, the establishments may just be providing check cashing or some other type of service.

The authors' statistical methods to "extrapolate findings" are not disclosed in this paper. The only statistical results that are presented pertain to the analysis of the racial composition of the U.S. Census Bureau census tracts based on the pre-established location of the stores. If these stores are not actually offering payday loan services or simply don't exist (since not all were verified), any statistical analysis that follows should not be considered relevant.

However, even in the unlikely event that the authors' data were accurate, the analysis that ensues brings up a number of questions of concern.

### ***Model Characteristics***

First, the authors state: "Unfortunately, [because there is] a highly significant correlation between African-American and Hispanic populations in North Carolina census tracts, we find it difficult to clearly interpret the meaning of the results of our analysis for Hispanic populations... The Pearson correlation coefficient between African-American and Hispanic concentration is 0.33, and it is highly significant at a 99% confidence level, which suggests that Hispanics tend to live in the same areas as African Americans."

If both Hispanics and African Americans tend to live in the same areas, how can one discern which group is purportedly being targeted by payday lenders? The authors do not show how the data could be controlled to

isolate the effect on African Americans away from Hispanics; i.e., excluding the potential effect that it is in fact the presence of Hispanic groups and not African Americans that is causing payday loan stores to choose their locations.

Second, the authors make the assertion that factors known to form the profile of payday loan customers are *unproven* characteristics of payday loan customers. (Such factors include being employed and relatively young, having a checking account, having an annual income between \$25,000 and \$50,000, having a high-school diploma or some college, as well as being a woman and/or a renter and have children living in the home.)

It is not clear what a "proven" set of parameters would look like or how it would be measured under the authors' criteria. Is race *the* proven factor determining payday loan store location choice? It would probably be difficult to prove that this hypothesis is correct since it just does not adhere to standard economic theory. Market research conducted by payday loan companies is likely guided by hypotheses based on standard economic principles rather than on elusive and imprecise factors such as race and ethnicity.

Census Bureau measurements of race and ethnicity suffer from drawbacks because of the self-reporting nature of the data, as respondent biases are difficult to correct. The fact that Hispanic populations include persons of any race complicates any analysis that seeks to identify segregated groups. In fact, the authors state: "Theoretically, one can be classified as both African-American *and* Hispanic...For the purpose of this study, any individual identified in the Hispanic ethnicity, regardless of race, was included in the Hispanic count. Therefore, African-Americans in the Hispanic ethnicity were counted as Hispanic, not African American." Undoubtedly, income levels, unemployment rates, or marital statuses are much more straightforward and precise measurements than race and ethnicity.

### **The Multicollinearity Problem**

The authors admit: "[Since] many of these descriptions [i.e., the above-mentioned characteristics] might be correlated with race...we designed multivariate regression models to evaluate whether race would continue to be a significant factor after controlling for these alternative explanatory variables." This is an important point to consider since regression analysis is severely constrained when analyzing the effect that inter-correlated explanatory variables have on a specified dependent variable.

If the variables that are believed to be explanatory are correlated (i.e., race, income, and education), they cannot be considered independent. The common result from this is called multicollinearity. Generally, collinear variables contribute redundant information and can cause other variables to seem less relevant than they really are.<sup>1</sup>

Moreover, when two explanatory variables are highly correlated, they essentially provide nearly the same information, and their relative effects cannot be accurately measured. If this is the case, neither variable may contribute significantly to the model after the other one is added; however, together they may appear to make a significant contribution. If two variables that are correlated were removed from the regression, the fit would be significantly worsened. Thus, while the overall model may appear to fit the data well, none of the explanatory variables by themselves makes a significant contribution after they are added to the regression model. The authors found a result that is therefore worrisome: "unemployment rate, age, education, and gender were generally insignificant in our models." These variables were added to a model that included number of children per household, income, and rural vs. urban areas.

The authors used a method called negative binomial regression modeling to attempt to isolate (control) the effects of the explanatory variables other than race. Nevertheless, the study doesn't show how the multicollinearity problem was addressed since variables such as income, education, and percent of people under the poverty line were included in the models along with race. Both theory and empirical observation would suggest that all these variables are correlated. Considering this likely problem with the model, one must

---

<sup>1</sup> An illustrative discussion on multicollinearity can be found in Kennedy, *A Guide to Econometrics*, Fifth Edition, pp. 205-217.

question the following result found by the authors: "The inclusion of nine control variables that purportedly describe the payday lending customer base made surprisingly little difference in our model."

Finally, the authors state that "[g]iven the skewed distribution of the dependent variable and high correlation between independent variables, the small sample size ... for the lowest and highest 10% buckets is not large enough for a revealing multivariate negative binomial regression." Aside from sample size, it is likely that the other two problems, particularly the correlation between independent variables, do not disappear for the rest of the distribution, as the authors claim: "On the other hand, the sample size ... for the 20% lowest and highest buckets is reasonably large enough for a revealing multivariate negative binomial regression in this context."

### **Omitted Variables**

The authors state that data "on credit worthiness is not available in the context of this analysis." Since the model is trying to identify the variables that determine the location of payday loan stores, this variable should have been included given its preponderance within the theoretical framework. The authors specifically state: "[The] inclusion of that variable [i.e., credit worthiness] would allow a more direct control measuring the availability of alternatives to payday loans for a particular neighborhood." However, the authors also claim that "our controlling for income and homeownership...serves a similar purpose."

Replacing income and home ownership for credit worthiness would make sense only if the latter were highly correlated with both of the former. However, it is likely that credit worthiness is not necessarily correlated with income since low credit scores can be found at virtually any level of the socioeconomic ladder. Moreover, homeownership does not guarantee good credit scores since foreclosures can occur for heavily indebted homeowners at most levels of wealth.

The authors claim that data on commercial zoning was not available when they conducted their study. However, location choice for payday lending stores, which are commercial outlets, is likely heavily influenced by this variable. The authors point out that "[it is] unlikely that low-income African-Americans are significantly more likely to live in census tracts with a disproportionate share of commercial zoning than low-income whites." Why ethnic groups locate in specific areas is a complex question since there might be a number of cultural, historical, social and economic reasons for this to happen. In fact, the existence of particular commercial establishments may indeed be a factor affecting migration patterns of certain ethnic groups. Thus, income, homeownership rates, and the other factors that the authors believe can replace the effects of the commercial zoning variable are unlikely to be adequate surrogates, unless a theoretical justification can be found to support this assumption.

A major assumption made in this study is that the racial and ethnic composition of customers of a payday loan store reflects the composition of the population living nearby the store. However, and as it was explained earlier, for areas with large numbers of commercial establishments, many of the customers are likely to be people who work, and not necessarily live, in those areas. Since the racial and ethnic composition of the workers (and potentially the customers) may not be the same as that of the residents, omitting the commercial zoning variable is clearly a likely source of problems for the model.

In sum, omitting theoretically relevant independent variables such as credit worthiness and the degree of commercial zoning from this model is likely to increase the risk of model misspecification. As a result, the parameter estimates from the regression may suffer from biases in the coefficients in terms of their sign as well as the magnitude of their effect on the dependent variable. Finding an apparent relationship in a regression that actually doesn't exist can be a consequence of omitted variable bias.

### **General Model Misspecification**

Given possible model misspecification and multicollinearity, it is not surprising to find this statement in the authors' study: "Finally, the proportion of households with children was significant in three of our four models; however, the result was the opposite of what might be expected. Our models suggest that tracts with higher proportions of households with children should be expected to have lower concentrations of payday lending stores." This result is unreasonable based on results from the Georgetown study and other observable evidence, which would suggest a high likelihood that the authors' models generated the "wrong sign" for this parameter coefficient, making the results likely unreliable and inaccurate.

## **CONCLUSION**

The research presented by the authors appears to suffer from a number of problems. First, the data collection process appears suspect because the authors do not clearly outline the steps taken to identify the sample elements (i.e., payday advance establishments). Moreover, statistics that would indicate the degree of statistical significance of the sample as well as its composition and distribution are not available from the study.

Second, the hypothesis presented by the authors (i.e., that payday loan companies target African American consumers) does not adhere to standard economic theory. Socioeconomic indicators such as income, educational level and marital status, among others, are believed to be consistent elements of the payday loan customer profile. Race and ethnicity might be at best proxies for some of these variables, and only under certain circumstances. When analyzing the choice of store location, rational business decision-makers would obtain less than optimal results by using proxies such as race and ethnicity rather than the socioeconomic variables that more accurately describe the profile of payday loan customers.

Third, even if it could be shown that payday loan stores are disproportionately located in areas with relatively large shares of African American residents, it would be extremely difficult to prove that there was intent on the part of the companies to choose those locations in order to target a particular racial group. Since many stores opened in their current locations a number of years ago, the racial composition of the areas may have changed over time, in which case the stores just happened to already be there as the demographic transition took place.

Moreover, as the authors admit, race and ethnicity are correlated with at least some of the variables that best encompass the profile of payday loan customers; and some of those variables are correlated with each other. Thus, it would be rather difficult to isolate their effect using regression analysis if it is determined that multicollinearity exists among the variables. If this is the case, the parameter coefficients from a regression analysis would be biased and likely inaccurate.

When multicollinearity exists, the likely reason is that the model has been misspecified. Respecifying the model to adhere to a reasonable hypothesis backed by economic theory would likely help solve the problem.

Finally, the effect of omitted variables such as credit worthiness and commercial zoning will also negatively affect the parameter coefficients, leading to potentially incorrect conclusions regarding the hypothesis test.

## APPENDIX

### **SUPPORTING BACKGROUND**

#### ***Duration and Use of Payday Loans***

The authors argue that, according to their previous research, only 1% of payday loans go to one-time emergency users, while 91% go to customers who receive five or more loans a year. The authors make the uncorroborated statement that the latter are "caught in a cycle of debt" just by virtue of the frequency of the loans that they presumably took in a year. However, since the duration of payday loans is usually about two weeks, it's hard to imagine that someone who had payday loans outstanding for a total of two and a half months out of the year is trapped in a cycle of debt.

Unfortunately, this paper doesn't mention the proportion of borrowers that had more than 5 loans per year. It is also not clear what the sample size was, whether it was statistically representative of the population, or what kind of methodology was used to obtain such sample. A review of *Quantifying the Economic Cost of Predatory Payday Lending* (December 18, 2003), the previous research cited, is necessary in order to attempt to answer these questions.

*Payday Advance Credit in America: An Analysis of Customer Demand*, a study by Georgetown University's Credit Research Center, provides different results. From a total of 427 completed interviews conducted in 2001 with payday loan customers across the country, this survey found that 65.7% of the respondents used their latest payday advance due to an emergency. Similarly, 65.5% (not 91%) of payday loan consumers received 5 or more loans in the previous twelve months, according to the Georgetown study.

The authors cite a publication by the Washington State Department of Financial Institutions (*Payday Lending Report: Statistics and Trends for 2003*) which found that 2% of borrowers received one payday loan advance per year. This is a weak corroborating finding for the authors' suppositions since there is no mention of the total distribution of borrowers, which could show that only a tiny percentage actually borrowed more than 5 or 6 times in a year. Still, a review of this publication is warranted in order to verify the nature of the distribution. Furthermore, this is data for just one state, hardly enough to make generalizations about the industry as a whole or in any other one state.

The authors assert that the vast majority of payday borrowers cannot solve their emergencies in two weeks. This is an unsupported statement since the authors do not present any survey findings indicating how long it takes for borrowers to solve the emergencies that prompted the payday advance. It is also stated in the paper that "rollovers can continue for months or years, with the borrower paying fees without the payday lender advancing the borrower any additional cash". This scenario is not likely for customers of CFSA member companies, which, as mandated by the association's Best Practices, cannot allow customers to rollover advances more than four times. State laws also can and often do limit the number of rollovers, which in many cases is even fewer than four.

#### ***Payday Loans' Cost to Consumers***

The authors assert that "this debt trap of repeated transactions costs five million U.S. borrowers over \$3.4 billion each year." This estimate comes from *Quantifying the Economic Cost of Predatory Payday Lending*. While a review of this study is necessary in order to assess the soundness of the estimate, it would be expected that a benefit/cost analysis be undertaken to measure the net cost or benefit of payday loans to consumers and the overall economy. A brief review of this study revealed that its authors did not conduct such analysis.

A benefit/cost study would take into account the cost of bounced checks as well as of late and non-payment fees. The cost of alternative sources of credit, such as credit card interest and other finance charges would also need to be included in the analysis. Many consumers who use the payday advance product are currently avoiding these costs. If such avoided costs are higher than the cost of using payday loans, then consumers on aggregate are better off using payday loans. The authors' earlier study only shows that payday advance customers pay \$3.4 billion annually in fees, with no accounting of any direct or indirect benefits.

***Payday Loans vs. Check Cashing***

The authors tend to lump together check-cashing and payday loan businesses to support their hypothesis that payday loan companies specifically target areas with large shares of minority residents.

Precisely, the authors point out that according to Kolb, "there were at least four times as many check cashers (sic) in zip codes that were 70 percent or greater minority as in zip codes that were less than 10 percent minority."

These findings cannot be readily verified since, according to the authors, Kolb's research has not been published. If they could be verified, the findings would have to show, among other things, that the proportion of payday loan establishments was overwhelmingly higher than that of stores offering only check cashing for this reference to be pertinent to the authors' hypothesis. It's intriguing that the authors clearly lump together these two services when in fact they are quite different. Moreover, the legal constraints regulating payday loan activity are very different from those regulating check-cashing transactions.