A Critique of

"Race Matters: The Concentration of Payday Lenders in African-American Neighborhoods in North Carolina"

By Thomas E. Lehman, Ph.D.
Associate Professor of Economics
Indiana Wesleyan University

Preface

About a year ago, I was privileged to collaborate with Dr. Thomas E. Lehman and three fellow economists to develop "Payday Lending, A Practical Overview of a Growing Component of America's Economy." We sought and succeeded in producing a clear, plainly readable summary of the emerging payday lending industry. We intended this initial vehicle to serve as common ground from which further discussions and writings would spring. My prior premise and that of my co-authors was and continues to be that the reason payday lending is such a fast growing industry is because it provides an important short-term economic service to many consumers. Even so, we all recognized that more scientific study of the industry was required to validate or refute our suspicions, and we are continuing to pursue the inquiry with a spirit of scholarship.

In 2005, the Center for Responsible Lending in North Carolina released "Race Matters: The Concentration of Payday Lenders in African-American Neighborhoods in North Carolina." An accompanying press release stated:

> "[Race Matters] confirms that the abusive loans made by payday lenders are not just an issue of fair and responsible lending, but are a civil rights issue as well."

Dr. Lehman's dispassionate examination of "Race Matters" suggests rather than a balanced assessment of the possible determinants of storefront location in the payday lending industry, "Race Matters" may have been designed to impugn the motives of the payday lending industry by suggesting it consciously preys upon minorities. Dr. Lehman's examination of the "Race Matters" study spotlights severe weaknesses within it, and I am persuaded that the results in "Race Matters" should be approached with extreme caution and a healthy skepticism.

Dr. Lehman's critique notes the data collection methods employed by "Race Matters" are unclear, causation is asserted upon the basis of weak correlations and that there are significant omitted variables which bias and erode confidence in the study's stated findings. He also discerns an underlying tone within the "Race Matters" study which suggests a clouded objectivity—perhaps motivated by an ideological bias against payday lending—not often seen in scholarly works. Thus, I welcome Dr. Lehman's

critique. It offers an important palliative against those who would let their passions affect their analysis.

Finally, The Center for Responsible Lending declined my 2005 request for their supporting data. In this context, Dr. Lehman's thoughtful critique is especially important.

Michael T. Maloney, Ph.D.
Professor of Economics
Clemson University

This critique was supported by the non-profit ***Consumer Credit Research Foundation****, which is working to improve research and data along with public discourse on the availability of short term consumer credit, including payday loans.*

Summary of the "Race Matters" study conclusions

The "Race Matters" study (King, Li, Davis and Ernst, 2005) presented data on the census tract locations of payday lending stores in North Carolina in an attempt to ascertain some of the location determinants of payday loan outlets. In particular, the authors of this study sought to explain the location of payday lending retail stores as a function of the racial composition of census tracts in North Carolina. Having found a positive correlation between these variables, the authors assert a causal connection. Specifically, the authors of the study claim to show that payday lending firms purposely target locations with a high concentration of African-American residents in North Carolina communities. Their findings, based upon multiple regression analysis, reveal that payday loan stores are more concentrated in census tracts with a higher proportion of African-American residents. In some instances, census tracts with relatively higher proportions of African Americans exhibit as much as three to five times as many payday loan stores per 100,000 residents. From this finding, the authors infer a causal link between the proportion of African-American residents in a neighborhood and the decisions of payday lending firms to locate stores in those neighborhoods.

The following analysis of the "Race Matters" study will provide a critique of the methods, findings, and causal inferences that should draw into question the conclusions reached by the authors of this study. It will be argued that the authors' conclusions are overstated at best, and misleading at worst, owing to the faulty data and research methods employed in the "Race Matters" study.

Critique of the "Race Matters" study

Data Collection Methods and Sampling Bias

The data collection method employed by the authors is unclear and thus open to criticism. The authors do not reveal specifically how the payday loan stores chosen for the study were actually selected, and their explanation of the data collection method seems obfuscated and purposely vague.

The authors state that their "approach ultimately yielded 385 payday loan storefronts openly operating in North Carolina" without explaining in clear detail how or why these particular 385 stores were selected. Unless the stores were selected using a random sampling technique (simple, stratified or cluster), the study data and findings are vulnerable to sampling bias. That is, a non-random sample of stores may not be representative of broad payday lending location decisions, and the findings would not be representative of industry practices. This would open the study to the criticism that it lacks reliability. Indeed, this appears to be the case.

The authors state that they assembled a list of payday lending companies "based upon company websites, newspaper articles, company advertisements, and advocates' reports." This appears to be a "convenience sample" as opposed to a true random sample, and is thus subject to several criticisms. Most importantly, if there are specific attributes of payday loan firms with websites or that engage in newspaper or other types of advertising considered by the authors, and these attributes differentiate them from other payday lending firms that do not engage in these activities, then the findings drawn from an analysis of these types of firms will yield results that are biased by the sample taken. By ignoring other types of payday lending firms, the authors' data are vulnerable to sampling bias. The authors either fail to acknowledge this weakness in their study, or simply do not recognize it.

Further, the authors concentrate their study on stores located only in North Carolina, but clearly suggest that they intend their results to be extrapolated beyond North Carolina markets. Why was North Carolina chosen? What makes North Carolina data necessarily representative? These questions go unanswered.

For these reasons alone, the sample of payday loan stores used by the authors cannot be said to be representative of the larger payday lending industry, and the findings are likely plagued by large sampling errors. At the very least, the authors might have provided the confidence levels or margin of error used in collecting the sample, but do not do so (Saltes, 2005). Again, this may be because the sample itself is a non-random sample, in which case these measures would not be relevant.

Correlation and Causation Inferences

Based upon their model of correlation and regression analysis, the authors assert that payday loan firms locate stores in geographic locations with a high density of African-Americans, arguing that firms "prey" on African-American customers. However, the authors do not possess data on the actual customer base of the specific stores used in the analysis, so there is no way to know whether, for example, these stores have a proportionately higher or lower number of African-American customers, or whether those customers even reside in the census tract in which the store is located. Theoretically, the stores may have a proportionately high or low number of African-American customers relative to the census tract in which they are located. And, theoretically, the stores may have a mix of customers who reside both inside and outside of the census tract in which the store is located. This is particularly a concern if the stores serve customers who work in close proximity, or who shop at retail outlets close to the payday loan store, but who have a residence in another census tract. The data gathered by the authors of the study leave these questions unanswered.

Although the link between a store's customer base and its geographic location may be plausible, the data used in this study fail to show any direct connection between location in a census tract and the actual customer base of the stores themselves. The authors are simply asking readers to make a "leap of faith" that the makeup of the census tract is necessarily the same as the makeup of the store's customer base, and then to infer from this loose connection that the payday lending firm made a calculated decision to locate the storefront in this census tract to achieve this result. In reality, the data used by the authors show no such connection. As Saltes (2005) accurately points out, "[T]he racial composition of people who live nearby a payday loan store does not necessarily imply that the store's customers have the same characteristics."

In sum, the authors fail to demonstrate causality in the model that they use, and this failure weakens their argument as well as their remedial policy prescriptions. They find a statistically significant correlation between store location and African-American concentration in census tracts. But, correlation does not equal causation. For example, it is widely believed that ice cream consumption and crime are directly correlated (they tend to rise and fall together). Yet, no one asserts that the consumption of ice cream "causes" people to commit crimes, nor that committing crimes "causes" people to consume more ice cream. As people generally recognize, the direct relationship between these two variables is likely explained by some third variable, such as weather patterns or hours of daylight. The same is surely true of the correlation between payday loan storefront locations and African-American residential concentrations, a finding that is likely explained by an omitted variable of some type.

Omitted Variable Bias

The authors of the "Race Matters" study fail to fully develop a broad economic theory of the many possible determinants of payday loan storefront locations. Instead, they confine their analysis to the census tract demographic data available to them. Upon finding a positive correlation between some of the census tract demographic characteristics (particularly racial composition) and the location of payday loan stores, they assert causation.

This is a major methodological weakness of the study, for it fails to consider a host of potential determinants of storefront location that go well beyond the demographic composition of the census tract in which the store is located. Thus, there are likely a number of omitted variables that may explain storefront location decisions but which were excluded from the model developed by the authors of the study. Omitted variables lead to severe weaknesses in regression analysis because they may lead researchers into drawing spurious relationships that do not exist. "Finding an apparent relationship in a regression that actually doesn't exist can be a consequence of omitted variable bias" (Saltes, 2005).

While it is plausible that demographic composition partially explains some of the reasons behind storefront location decisions, certainly there are other factors as well. Yet, the authors do not consider, and thus do not include in their model, variables which might measure these other factors.

As a result, the independent or explanatory variables that are included may be correlated with some other excluded variables that more clearly explain the variation in both the independent and dependent variables employed in the model.

For example, it is plausible to assume that stores make location decisions based upon the rental price of real estate, the lease costs for buildings, potential property tax abatement incentives in the urban area, the college education level of the local labor force, and the potential payroll costs of local labor. Additionally, firms' location decisions may be governed by commercial and residential zoning regulations, as well as storefront proximity to other retail outlets that may be complimentary to the payday lending business (i.e., the nearby location of an auto-repair garage, alternative consumer finance retailers, or a Wal-Mart or similar big-box retailer). These factors, more than the racial composition of the census tract, could be expected to explain a great deal more of the variation in payday loan storefront locations. And, to the extent that these omitted variables may also explain census tract racial composition, the authors' findings are overstated and likely spurious. At the very least, the magnitude of the effects of the independent variables in the model cannot be confidently taken at face value given the potential number of omitted variables.

Multicollinearity and Model Specification Problems

Perhaps the most severe weakness of the "Race Matters" study is the multicollinearity between the various demographic measures used as independent variables in the model. While the authors admit of multicollinearity problems, they do so only in passing in a footnote buried in the Appendix to the study. Further, they do nothing to investigate the degree of multicollinearity or to caution the reader when interpreting the findings.

Multicollinearity occurs when two or more independent variables included in a regression model are highly correlated. The effect of their simultaneous inclusion in the model is to mask the independent effect that each has on the dependent variable, and to depress the magnitude of the effects of other variables in the model. Multicollinearity problems among independent variables in the regression model create a redundancy in explaining the variation in the dependent variable; a redundancy that

cannot be filtered out so that we can observe the magnitude of the separate effects of the independent variables on the dependent variable (which is the ultimate benefit of multiple regression analysis to begin with). This is typically the result of a misspecified model that includes an overabundance of independent variables measuring the same or similar attributes. This appears to be the case in the "Race Matters" study.[1]

The authors include a number of independent variables in the regression model that measure similar demographic characteristics and that can be expected to be highly correlated with each other, thus masking the effect that each would seem to have on the storefront location decisions of payday loan firms. For example, the authors include census tract median income, poverty level, home ownership level, unemployment rate, and education level as independent variables in their models, all of which theory and empirical evidence suggest may be correlated with one another and/or with the other independent variable measuring racial composition.

For example, both the proportion of African-Americans living in a census tract and the median household income in that census tract are found to be correlated (positively and negatively, respectively) with store location in some of the authors' regression models. However, median income and racial composition variables have been repeatedly found to be highly negatively correlated to each other. When both are inserted into the same regression model, their effects upon the dependent variable, payday loan store concentration, are likely corrupted. At best, the magnitude of the effect of each variable is distorted by the simultaneous inclusion of the other into the regression model.

Nonetheless, the authors confidently conclude, based upon the significance of correlation coefficients, that payday loan firms are "targeting" African-American neighborhoods, holding other factors constant, without acknowledging that the multicollinearity problems in their models do not permit them to draw an unconditional conclusion of this sort. For instance, the authors also find that the proportion of census tracts that are rural is statistically significantly and negatively correlated with the presence

[1] For an excellent explanation of the statistical problem of multicollinearity in regression models, and one relevant to this study, see Saltes, 2005.

of payday storefront locations (i.e., payday loan storefronts are located overwhelmingly in urbanized census tracts). Does this mean that payday lending firms are "targeting" urban customers only? And, are urban customers more or less likely to be African-American than rural customers? Evidence suggests that urban populations are much more likely to contain higher concentrations of minorities, so which of these two variables has greater explanatory power? What is the separate effect that each variable has on payday storefront location? Or, is there some other omitted variable that may explain payday storefront location concentrations as well as these other two variables?

One way of isolating the effect of collinear independent variables would be to isolate these variables in separate regression models. Yet, even here, if the theory behind the model says that they should be included, the separate models will suffer from omitted variable bias. The authors do not reveal whether this was attempted, much less reveal conditional results from this type of adjustment.

Given that the authors do not reveal the level of multicollinearity between the independent variables with a correlation matrix or a listing of tolerance levels for each variable in the model, their findings that payday loan storefronts are strongly correlated with African-American population in North Carolina census tracts are overstated at best, and misleading at worst. Until we have data that can overcome these multicollinearity problems, such an assertion should be made only with the utmost caution and with the caveat that multicollinearity is an issue. However, the authors do not offer this caveat, and thus overstate their case.

Non-normally Distributed Dependent Variable

One of the basic assumptions of the multiple regression statistical technique used by the authors of the "Race Matters" study is that the dependent variable is "normally" distributed. In other words, the values in the dependent variable should not be severely skewed, either negatively or positively. If the values in the dependent variable are heavily skewed so that there are only a few cases with extreme high or low values, the results of a multiple regression statistical technique are severely weakened. The greater the degree of non-normality in the dependent variable, the less confidence we can place in the results of the regression accurately repre-

senting the relationship between the independent and dependent variables.[2]

In technical terms, for each value of a given independent variable, there must be a group of dependent values that are normally distributed so that the mean of the residuals (Y – Y') are equal to zero and lie roughly on the regression plane. This can be approximately achieved by ensuring that the dependent variable in the regression model exhibits a normal distribution with low skewness. In the "Race Matters" study, this appears not to be the case. This is not necessarily surprising, since, as mentioned previously, the data do not appear to be gathered using commonly accepted random sampling techniques. One of the ways in which researchers may obtain normally distributed data and guard against non-normally distributed variables is by taking a random sample.

The dependent variable in the "Race Matters" study is the number of payday loan stores in a given census tract, where the census tract is the unit of analysis. However, of the 1,554 census tracts included in the study, only 185 actually have at least one payday loan store. As the authors themselves explain, their data collection method yielded "a dataset of 385 total payday loan storefronts in 185 of 1,554 North Carolina census tracts. Among these tracts, 96 census tracts have one store each, 30 have two stores each, 27 have three stores each, 19 have four stores each, and 13 have more than four stores each."

In other words, only 12 percent of the census tracts examined by the authors contain a payday loan store. This leaves the data on the dependent variable with a high level of positive skewness where only a small proportion of the total census tracts contain a payday loan store, and, among these, payday loan stores appear to be even further concentrated into a few tracts.

One of the statistical methods used in adjusting non-normally distributed variables is to transform them into a log value of themselves to obtain a more normal distribution. However, the authors do not attempt this transformation, and refer only once to the "skewed distribution of the dependent variable" in a footnote in the Appendix.

[2] For a basic primer on the assumptions of multiple regression and correlation analysis, see the following: Berry, 1993; Kennedy, 1997; Tabachnick and Fidell, 2001.

Given the extreme positive skewness in the dependent variable in this study, the finding that payday loan stores are concentrated in neighborhoods with a higher concentration of African-Americans is severely weakened. At a minimum, the authors should have identified this weakness and cautioned the reader against reading too much into the results. At best, the authors could have transformed the dependent variable and run a separate model revealing the results, perhaps in the appendix. However, the authors do neither and thus, again, appear to overstate their case.

Tone of the Study

Rather than remaining objective in the explanation of their findings, the authors of the study frequently use normative and loaded language to describe the payday lending industry, leading the casual reader to believe that payday lending is a social ill that provides no benefits. For example, the authors make loaded statements about "predatory" payday lenders who "trap" borrowers in a "cycle of debt" without providing credible references to outside supporting research.

The references used to support these declarative statements are either 1) to research conducted by the CRL itself, the same organization supporting the "Race Matters" study, or 2) to the Stegman and Farris (2003) study which uses similar data on North Carolina and has several methodological weaknesses of its own. Thus, the approach taken by the authors seems open to the criticism that it is ideologically motivated, rather than a balanced study of the location decisions of payday lending stores. That is, rather than a balanced professional investigation of the possible determinants of storefront location in the payday lending industry, it seems the author's intent was to impugn the motives of the payday lending industry by suggesting that the industry consciously "preys upon" minorities.

Conclusion

In summary, the "Race Matters" study contains severe weaknesses, and the results should be interpreted with extreme caution, for the following reasons:

- The data collection methods are unclear. It does not appear that the data were collected using a true random sample, opening the research sample to possible (and likely) sampling bias.
- Causation is asserted on the basis of weak correlation. The assumption that the characteristics of census tract residents are similar or equal to the characteristics of payday loan store customers located in those census tracts is undemonstrated.
- Omitted variable bias. There are certainly omitted variables in the regression models developed by the authors, and these models could have been significantly improved if the authors had employed a more fully developed economic theory of payday loan storefront location, including real estate costs and the proximity of payday loans to complimentary retail businesses. For example, it seems likely that payday lenders would have great business interest in population densities, automobile traffic flows, and rental housing concentrations near their retail facilities. Variables like these are of keen interest to all retail service providers.
- Multicollinearity among the various independent variables employed in the regression model make it difficult to determine the independent strength that each has in explaining payday loan storefront location, including the effect, if any, of the concentration of African-American residents on storefront location decisions.
- The dependent variable used in the multiple regression analysis is non-normally distributed and highly positively skewed, severely weakening the findings.
- In the modern American marketplace, sellers of goods and services are driven by an invisible hand-in-glove dynamic of intense competition. This is a continually evolving process which benefits sellers and consumers alike. The economic targeting of consumers who are likely to need the services provided by payday lenders based upon their income, population density, proximity to transportation and other variables does not equate to the specific targeting of racial groups who may—or may not—have similar demographic characteristics.
- The tone taken by the authors of the study suggest a lack of objectivity perhaps motivated by an ideological bias against the payday lending industry. This may explain why the authors appear to overstate their case given the weaknesses of their research.

Because of the numerous weaknesses identified with the research methods employed in the "Race Matters" study, the conclusion that neighborhood racial composition is a critical or even plausible determinant of payday loan storefront location decisions must be rejected, at least until better data and a more complete theoretical model are available. As mentioned above, the tone of the study, combined with the rather severe missteps and oversights involved with the data collection and interpretations of findings, compels one to conclude that the "Race Matters" study adds very little to the understanding of the locational decisions of payday lending stores. In fact, contrary to the claims of the study, it is likely that "race matters" not at all.

References

Berry, W.D. (1993). *Understanding Regression Assumptions*. Newbury Park, CA Sage Publications, Inc.

Kennedy, P. (1997). *A Guide to Econometrics, 3rd ed.* Cambridge, MA: The MIT Press.

King, U., Li, W., Davis, D., and Ernst, K. (2005). Race matters: The concentration of payday lenders in African-American neighborhoods in North Carolina. *Center for Responsible Lending.* www.responsiblelending.org.

Saltes, D. (2005). A look at "Race matters: The concentration of payday lenders in African-American neighborhoods in North Carolina." Alexandria, VA: *Community Financial Services Association of America.*

Stegman, M. and Faris, R. (2003). Payday lending: A business model that encourages chronic borrowing. *Economic Development Quarterly, Vol. 8, No. 1.*

Tabachnick, B.G. and Fidell, L.S. (2001). *Using Multivariate Statistics, 4th ed.* New York: Harper Collins College Publishers.

Consumer Credit
Research Foundation

www.consumercreditresearchfoundation.com