

TO: Federal Trade Commission  
FROM: Delvin Davis and Joshua M. Frank, Center for Responsible Lending  
DATE: March 30, 2012  
RE: *Under the Hood* research report and methodology

Thank you for the opportunity to participate in the Federal Trade Commission (FTC) Roundtable discussions on auto financing and sales issues. We think the discussion was helpful in understanding auto finance and abusive practices in that market.

The purpose of this memo is to respond to Dr. Thomas Durkin's remarks and criticisms of our research report, *Under the Hood*, made during Panel 8 of the August 3, 2011 roundtable session in San Antonio on "Fair Lending: Interest Rates, Markups, and Payments". We felt it important to have our response on record.

We stand behind our research and its findings. We used the best publicly-available data to study this issue. All of the data we obtained reported directly from auto industry sources, including figures from Securities and Exchange Commission (SEC) prospectus filings on auto loan asset-backed securities. Other data used to measure the scope of rate markups came from a National Auto Finance survey of auto lenders, as well as data from CNW Marketing Research that tracks annual auto sales.

We believe that *Under the Hood* furthers the discussion around interest rate markups, their scope, their role in distorting incentives, and their impact on loan performance because of the data analysis included in it. Since the report's release, the industry has been quick to criticize the report yet has not provided any data to the contrary. We also asked Dr. Durkin during the panel to provide us with specific concerns about the paper, and he has not done so.

Dr. Durkin spent a considerable amount of time during the panel session discussing our data, methodology, and even the format of the report. His criticisms fell into two general categories. Some of his criticisms were specific to our paper. The others were very general issues that could be raised with any quantitative research study, not just ours.

In essence, he criticized our report for using generally-accepted and widely used statistical methods.

We believe that to advance the conversation about this important issue, an examination of existing data is imperative for any fact-based decision making. We know that the data has been collected, but those data sources are unavailable to those outside the industry. We urge the FTC to procure and examine this data given the industry's seeming refusal to make data available publicly.

### **Issues Specific to *Under the Hood***

Dr. Durkin raised two specific concerns with our paper. First he said that he did not see descriptive statistics of our variables. They are provided in Appendix 4 of the report on page 20.

Dr. Durkin said that he did not see a well-defined hypothesis or research objective. The first sentence of page 9 in the Data and Methodology section reads, “The objectives of this research are threefold: To estimate the magnitude of markups nationally, explore loan conditions that influence rate markup, and investigate rate markup’s impact on loan performance.”

### **General Statistical Issues**

Dr. Durkin raised a host of unsupported concerns including omitted variables, multicollinearity, over-fitting, a non-normal dependent variable, heteroscedasticity, and no specified functional form. All of his criticisms are those that can generally be leveled at any statistical analysis, including those Dr. Durkin has published.

These and other statistical objections thrown out by Durkin can be adequately addressed by recognizing two facts about our study. The first is that the high level of statistical significance of the results which, as already noted, makes many minor technical objections unimportant (even if there was evidence that these violations existed).

The second fact is that when analyzing the main result of the research—the relationship between higher markups and higher losses—it is important to note that we split the data into subprime finance companies and other lenders (which is also made clear in the report).

Dr. Durkin did not, however, offer any specific evidence of where each issue is present in our paper. He generally asserted that the data is incomplete, that we both omitted variables and had too many variables, and that we skewed the research to our advantage. We disagree:

- We used the best data available to us. If there is other data, then the industry should release that data in a form that allows for public analysis.
- In any analysis, some variables will be omitted. If there are available variables that Dr. Durkin or others in the industry think were crucial, then they should tell us what they are and why they should be included.
- Our methodology and analysis were based on accepted statistical practice, and we were careful to ensure that the results were valid.

It is important to note the high level of statistical significance for the most important findings. This is particularly relevant for our results on who gets a high markup where we did not split the data into two groups. Anyone with familiarity of statistical methods can levy general and random objections with any piece of empirical research. We also bifurcated the data to separate loans made by subprime finance companies from the other

loans in the pool to help ensure that the issues Dr. Durkin raised did not affect the statistical validity of the study.

However, statistical expertise and statistical intuition can help one differentiate between what is merely a technical objection and one likely to materially affect the outcome of the results. When a result is significant at the 0.1% level, as ours is, minor objections over whether certain technical statistical assumptions hold (such as moderate levels of heteroscedasticity, non-normality, and outliers) don't impact the overwhelming statistical validity. Our findings are statistically valid, and as such stand up to scrutiny.

Omitted Variables

Durkin raised the issue of omitted variables, but did not specify which variable may have been missing. Of course, this is a potential issue for all statistical analyses not involving a tightly controlled laboratory experiment, not just ours. It is a generic accusation that can be leveled at any piece of research when someone disagrees with its conclusions, and the question can arise in any research.

Durkin does not specify what additional variables should be included that would significantly change our findings. The most likely candidate is risk, which we attempt to account for in the final report (at least in part) by separating sub-prime finance companies from securitizations of loans from other issuers. We also ran numerous multiple regressions that we did not include in the final report, which continued to show that markups were significantly related to delinquencies and losses even after accounting for FICO and other risk markers. In fact, markup is more predictive than any other factor (including risk variables) of losses (see Tables 1 and 2 below).

Table 1: Multivariate Regression with 60-Day Delinquency Rate as Dependent Variable

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.736	2.046		2.804	.011
	Avg Amount Financed (in '000s)	-.032	.035	-.153	-.902	.377
	Weighted Avg FICO	-.002	.004	-.188	-.509	.616
	% Used Sales in Pool	-.010	.004	-.600	-2.616	.016
	Weighted Avg Term	-.071	.030	-.534	-2.370	.027
	Markup Cap Present	-.530	.204	-.452	-2.604	.017
	Captive	.279	.259	.267	1.079	.293
	Finance Co.	.389	.223	.361	1.746	.096
	Estimated Rate Markup	.126	.060	.926	2.112	.047

Table 2: Multivariate Regression with Cumulative Loss Rate as Dependent Variable

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-5.682	9.708		-.585	.565
	Avg Amount Financed (in '000s)	.330	.167	.261	1.979	.061
	Weighted Avg FICO	-.020	.019	-.295	-1.031	.314
	% Used Sales in Pool	-.022	.019	-.211	-1.186	.249
	Weighted Avg Term	.023	.143	.028	.158	.876
	Markup Cap Present	-.624	.966	-.087	-.646	.525
	Captive	1.561	1.227	.244	1.272	.217
	Finance Co.	-.659	1.057	-.100	-.624	.540
	Estimated Rate Markup	.808	.283	.971	2.859	.009

Multicollinearity

The reason these tables were not added to the final report even though they are quite compelling is that we take a conservative approach in reporting our results. We did this because we were concerned with multicollinearity. Multicollinearity is when variables included in the analysis overlap with each other. In this case, FICO scores and dealer interest rate markups overlap. In any multivariate regression analysis, there will be multicollinearity in some form or fashion. While this can have an impact on the findings, it does not necessarily invalidate the outcome.

We intentionally chose to remove the multivariate relationships in our model in order to address that issue. We split our data into two groups to account for risk in our model after we removed a risk variable from the model to eliminate multicollinearity. We are also clear in our conclusions to not imply causal statements between variables.

The multicollinearity issue could be adequately addressed and a much more powerful multiple regression could have been run with a larger dataset. Currently the kind of data necessary to satisfy this concern is not available to the general public, which again speaks to the need for greater data transparency from the auto finance industry. The data we do have is strongly representative of the auto finance industry and provides statistically significant results.

Over-Fitting

While on the one hand Dr. Durkin felt we did not include enough variables, he went on to argue that we included too many. Dr. Durkin cited over-fitting as an issue, which is the use of inappropriate variables to skew the results or including too many variables without a strong theoretical basis in the hope that one will be statistically significant.

Dr. Durkin did not indicate which of our variables was inappropriate to include. We gave careful consideration to the variables that were included, as each could have a foreseeable influence on a loan's financing structure and performance. Each variable is also well populated, as shown in the report's descriptive statistics (Appendix 4, page 20). If each variable has a legitimate place in the regression model, then the over-fitting argument is invalid.

### *Non-Normality and Heteroscedasticity*

Dr. Durkin also mentioned non-normality and heteroscedasticity as potential issues with the paper. Yet again, these are general issues that Dr. Durkin referenced which present themselves in any statistical analysis, and did so without citing any concrete issue with our research.

Non-normality occurs when the variable does not show the traditional bell-shaped curve. In instances where the data set is smaller, like this one, data will often not visually appear "normal" even when it is. If the data truly is not normally distributed, one simply needs to take steps to recognize this issue and determine whether that non-normality changes the outcome. We determined that this was not an issue in our analysis because of the statistical significance of the analysis, and the act of splitting the data further eliminated any indication of a non-normal distribution.

Heteroscedasticity occurs when variance changes across variables. Heteroscedasticity can affect the standard error in statistical analysis, which is used to determine whether the outcome of the analysis is statistically significant. When the outcome is highly significant, heteroscedasticity does not invalidate the outcome. In our research, heteroscedasticity exists if all loans are combined in one pool. When separated into prime and subprime pools as was done in our research, heteroscedasticity is no longer an issue.

### **Criticism of Data Used**

Andrew Koblenz from the National Auto Dealers Association, who also served on the panel, incorrectly claimed that our use of the National Automotive Finance Association's (NAF) survey data is inconsistent with figures reported in the NAF data report. The NAF data summary presents its aggregated figures in two ways: one representing data from all 25 survey respondents, and other set of tables detailing responses from only the 15 respondents that have taken the survey in previous years. Data from repeat respondents allows NAF to draw year-to-year comparisons. However, the dealer reserve averages Koblenz referred to in the panel were averages for the 15 repeat respondents, whereas the averages in our report use figures from the full 25 dealer sample.

Dr. Durkin also implied that Federal Reserve G.19 data shows that indirect auto lending may be cheaper for the consumer than direct auto lending. This is not only an inaccurate conclusion, it is entirely impossible to make this comparison based on publicly reported G.19 results. To be fair, Durkin did state in response to a question from the audience regarding direct vs. indirect lending rate, "The Fed data is not well defined. You'd have to

go there and ask them exactly what's in there, so I wouldn't want to make too much of it.” But, Dr. Durkin then implied that the Fed data in fact *did* shed light on the subject.

We have asked the staff at the Federal Reserve what is included in the G.19 data. We used this information in a prior report to analyze disparities in credit card interest rates reported in G.19 (“Credit Card Clarity” available at <http://www.responsiblelending.org/credit-cards/research-analysis/credit-card-clarity.html>).

The aggregated G.19 data includes both direct and indirect loans from banks and is not helpful in determining the difference in rates from these sources. It also may not be helpful in determining the rate difference between subprime-focused finance companies and banks, which is what we separated in our research. It appears based on the Federal Reserve’s FR3033p form that finance companies may include “captives” (i.e. financing sources affiliated with an auto manufacturers) in addition to finance companies not affiliated with an automobile manufacturer.