Disclosure of Time-Barred Debt and Revival: Findings from the CFPB's Quantitative Disclosure Testing

June 30, 2020 Peer Review

by

Members of the CFPB's Academic Research Council:

Michael R. Baye, Bert Elwert Professor of Business Economics and Public Policy, Kelley School of Business, Indiana University

Karen Dynan, Professor of the Practice, Department of Economics, Harvard University

Terri Friedline, Associate Professor, University of Michigan School of Social Work

John G. Lynch, University of Colorado Distinguished Professor and Senior Associate Dean for Faculty and Research, Leeds School of Business, University of Colorado-Boulder

Brigitte C. Madrian, Dean Brigham Young University Marriott School of Business

Thomas W. Miller, Jr., Jack R. Lee Chair of Financial Institutions and Consumer Finance, Department of Economics and Finance, Mississippi State University

Joshua D. Wright, University Professor of Law and Executive Director of the Global Antitrust Institute, George Mason University Antonin Scalia School of Law

Overview and Summary of Conclusions

The Academic Research Council (ARC) was asked to conduct a peer review to evaluate and comment on the theoretical and empirical merit of information in a report, entitled *Disclosure of Time-Barred Debt and Revival: Findings from the CFPB's Quantitative Disclosure Testing*, produced by CFPB Staff. We were asked to consider, among other things: (1) whether the methodology and assumptions employed are reasonable, appropriate, and consistent with accepted social science theory and statistical practices; (2) whether the data used are reasonable and of sufficient quality for purposes of the analysis; and (3) whether the conclusions, if any, follow from the analysis. Our review recognizes that standards for evaluation are not necessarily the same as those one might apply in evaluating studies for publication in a professional journal. For example, it is not necessary that the study present new or novel theoretical results or empirical techniques. We have not been asked to provide advice on policy, nor have we been asked to evaluate the policy implications of the study.

The report summarizes results from a survey designed to assess consumer understanding of different disclosure notices related to time-barred debt. The study tested 11 different potential disclosures which varied in how information about time-barred debt and the possibility of its revival was explained to consumers. The survey was designed to test different aspects of consumer comprehension, and the results of the analysis demonstrate that modest changes to existing and proposed disclosures related to debt collection result in a statistically significantly increase in consumer comprehension.

The CFPB staff designed and implemented this survey well; the analysis is scientifically sound. We especially appreciated the clarity and accessibility of the exposition—the authors obviously took a care to make the study accessible to a non-expert audience. Overall, readers gain a materially better understanding of how consumer knowledge might be shaped by disclosures.

The report provides an overview of the methodology and disclosure notices that were tested; the appendices and supplemental materials provide more detailed information about the survey instruments, potential disclosures, and sample demographics. A key takeaway from the report is that certain disclosures enhance respondents' abilities to comprehend and/or recall information about time-barred debt. For example, in Figure 7, there is about a 50 percent difference in "correct" responses between respondents that received the time-barred debt (TBD) notice alone and those that received the TBD notice with revival (with respondents that received the TBD notice with revival responding correctly more often). The TBD notice with revival information provided respondents with the clearest and most consistent understanding. Section 3.1 provides appropriate caveats regarding whether the results of this study may or may not generalize, and why. Additionally, the survey design included questions to gauge respondents' confidence in their answers, and the report has a number of helpful pedagogical features. For example, Figures in the report highlight "correct" answers to make it easier for readers to interpret results.

The report offers an introductory paragraph describing that the survey was undertaken to test several versions of disclosures, but does not contextualize the results by providing a literature review or describing whether or how the results of the survey will be used. Our review focuses on the scientific merit of the results as presented, and not on the degree to which the report should inform policy or rulemaking. Readers must interpret the implications of the results for themselves.

The remainder of our peer review discusses potential limitations of the study and provides a few minor comments. While many of the limitations we discuss are beyond the scope of the report, they may be important to consider before using the results of the study to address certain policy questions.

Potential Limitations

The report does not draw inferences about the economic significance of the survey results. Instead, the findings indicate that different types of disclosures result in statistically significant differences in the response to various questions about time-barred debt for the particular sample assessed. In some cases, the differences were quite large—for example, the results in Figure 6. Those results suggest that disclosure design is very important. But, the results do not indicate the level of economic harm arising from existing disclosures or whether alternative disclosures reduce the potential harm to consumers of the current laws around debt collection. Returning to the example of Figure 6, the report emphasizes that 70 percent of the sample assessed got the answer "correct" after seeing the TBD-with-Revival notice, but does not measure the economic benefits of this "improvement" or the extent of harm from the fact that 30 percent still got the answer "wrong."

More broadly, in interpreting the results of this study, it is important to recognize that the report is silent on the definition of "success." To judge which disclosures are most effective as a matter of public policy, one must have a clear definition of a successful disclosure. As the following questions reveal, this may be a tricky task:

> What behaviors would indicate a successful disclosure?
>
> Is the goal to induce people with debts past the statute of limitations to not be drawn back into an obligation to pay the debt? Put a different way, would a consumer be making a "mistake" to acknowledge a debt that is past the legal statute of limitations even if he or she wants to repay it for "ethical" reasons?
>
> What are the consequences of inaction when a consumer is dunned by a debt collector for time-barred debt? Would a self-interested consumer refuse to acknowledge that debt for any reason other than, say, a moral belief that one should pay one's own debts? For example, might acknowledging and paying one's debts (and repaying) past the statute of limitations improve one's credit score or

have some other material benefit other than a sense of living up to a moral obligation?

What should a consumer do if dunned over an old debt that is improperly recorded? If a consumer gets a debt collection communication and there is some aspect of the claim that seems wrong, is the "right" decision to keep one's mouth shut and not to acknowledge the debt? If the debt negatively impacts the consumer's credit score, it is not obvious that the right decision is to stay silent. This was the subject of Section 3.9.

Addressing questions such as these are clearly beyond the scientific goals of the report, but they may be relevant before applying the results to public policy. Additional research may be necessary to address these and other issues relevant for particular public policy applications of the study. For example, additional research would be necessary in order to determine the level of consumer harm stemming from 30 percent (referencing the example above) of those seeing an "enhanced" disclosure still getting the answer "wrong."

While the study examines many information treatments, the report is largely silent on why specific treatments were included, which treatments were hypothesized to be "better" or "worse," or why the version other than the one that seemed "best" in the end had been entertained. Depending on the policy application, it may (or may not) be appropriate to compare "model" and "status quo" notices that do (do not) inform subjects about their legal rights.

The sample in the present study may be appropriate for some policy applications but not others. One could envision a policy question for which the appropriate sampling population is representative of the United States; one can also envision a policy application in which the appropriate sample is representative of those who have engaged with a debt collector in the past. The discussion of population coverage on pages 5 and 37 of the report is relatively brief, but it appears that the actual sample in the report is a blend of these two populations. Some policies may be targeted to subpopulations with specific demographic characteristics (e.g., gender identity, income level, race, disability status), and these may be underrepresented in the sample. For instance, the *Board of Governors of the Federal Reserve System's 2015 Consumers and Mobile Financial Services Report* used GfK's KnowledgePanel and reported a response rate of 44% for an oversampling of Black and Latinx adults, compared to response rates of 63% and 79% for non-oversampling.

Another consideration that may limit the utility of the present study in some policy applications is the possibility that treatment manipulations interact with "background factors" that are policy-relevant but not focal in the investigation.[1] We highlight three examples.

---

[1] Lynch, John G., Jr. (1982), "On the External Validity of Experiments in Consumer Research," *Journal of Consumer Research*, 9 (December), 225-239; Lynch, John G., Jr. (1999), "Theory and External Validity," *Journal of the Academy of Marketing Science*, 27 (Summer), 367-376; Hutchinson, J. Wesley, Wagner A. Kamakura, and John G. Lynch, Jr., (2000) "Unobserved Heterogeneity as an Alternative Explanation for 'Reversal' Effects in Behavioral Research."

First, the results in Section 3.5.2 show that the disclosures are more effective in improving the comprehension of more highly educated consumers and those with higher incomes. This may be a relevant policy consideration if there is evidence that the "vulnerable" populations targeted by a particular policy are correlated with income or education.

Second, hypothetical survey responses may differ from real responses in the field. The supporting document (CFPB-2019-0003-0003.pdf) discusses critiques from the comment period regarding "*Use of Hypothetical Scenario in Survey Questions."* On p. 8 of that document, CFPB researchers address a critique that the use of hypothetical scenarios might affect the generalizability of the findings, and state that "the Bureau is interested in relative differences between groups in disclosure comprehension, depending on the disclosure that each group receives; the Bureau does not intend to rely on this research project to understand incidence rates in the population. The hypothetical nature of the questions should have similar effects (if any) on participants in all experimental groups, and therefore would be a common factor across groups. Comparing relative responses across groups, as opposed to measuring the incidence rate of responses for a particular group, should render any effect of the hypothetical nature of the questions irrelevant for the Bureau's purposes." If the hypothetical versus real nature of responses affects the *relative* means or interacts with treatments, this could impact the external validity of the study in certain applications.[2]

Incentives can also result in differences in responses to hypothetical survey questions and the actual behavior (or understanding) of consumers in the field.[3] While the report includes a thoughtful discussion of how responses may be distorted by monetary incentives, the there is no discussion of the effects of respondents confronting a hypothetical situation regarding a stranger. It is possible that actually engaging with a debt collector about one's own past-due debt might induce stress that could impede one's ability to synthesize the information in the disclosure. Additionally, respondents to these types of surveys are typically compensated. There is also some ambiguity regarding whether all participants were compensated for participating, and if so, how much.

Third, differences in responses may stem from differences in consumers' prior debt experiences. The sample used in the current study includes those with and without prior debt collection experience. On p. 7 of CFPB-2019-0003-0003.pdf the authors note: *"In previous*

Journal of Consumer Research, 27 (December), 323-344; Lynch, John G., Jr. (2019), "Interdisciplinary Behavioral Science and Consumer Protection," Public Commentary for US Consumer Financial Protection Bureau Behavioral Law and Economics Symposium. Washington, DC. Available at: https://www.consumerfinance.gov/about-us/events/archive-past-events/cfpb-symposium-behavioral-economics/
https://files.consumerfinance.gov/f/documents/cfpb_lynch-written-statement_symposium-behavioral-economics.pdf

[2] Lynch, John G., Jr. (1982), "On the External Validity of Experiments in Consumer Research," *Journal of Consumer Research*, 9 (December), 225-239.
[3] See, for instance, Michael R. Baye and Joshua D. Wright, "How to Economize Consumer Protection," T*he Antitrust Source*, Vol. 17, No 4 (February 2018), pp. 1-15.

versions where consumers were asked to estimate their own behavior rather than that of a hypothetical Person A, researchers found that consumers without debt collection experience dwelled on the idea that they would never be in the position of owing a debt, which interfered with their ability to complete the survey. Switching to a third person proved easier for both those with and without debt collection experience to answer questions about the information on the form." For policies in which the disclosure is intended to help people who already have time-barred debt, the results may not be directly relevant if the disclosure works only for those who don't have such debt. For policy issues like this, additional research using a different sampling strategy and running the study only on those who did in fact have such debt may be useful. We note that a strength of the current study is that the patterns of treatment means were quite similar for those with and without prior debt collection experience (see, e.g. Section 3.5.1 and Figure 2).

Other Minor Comments

We conclude with a few expositional and methodological comments.

1. Some types of incorrect answers seem more concerning than others. People erring by not acknowledging the amount of debt they have (as in Figure 8), may be less concerning than some of the other "incorrect" actions they might take. At face value, one might argue that Figure 8 represents a reason not to use a TBD with Revival Notice.

2. The results by education and income group were very interesting. It would have been interesting to also see results by age.

3. It would be useful to explore whether the results might have depended in some way on the order of the measures in the survey. One way of doing this is to consider a study design that counterbalances the order of measurement of the various individual difference moderators and the manipulations of the various treatment conditions.

   a. In most cases, it is best to measure moderators before and not after the treatment manipulations and the main DVs. In the study conducted, Q33-Q36 follow rather than precede the manipulations, making it possible that the manipulations affect the responses to questions that are supposed to reflect prior beliefs and experiences. The results would be more convincing if measures of one's moral beliefs had preceded, rather than followed, the main independent and dependent variables. It would be quite interesting to know how the effects of the treatment manipulations varied as a function of various background factors – like demographics and attitudes toward moral obligations to repay debt. But that argues for measuring all the moderator variables first, before the manipulations.

b. The counterargument to that critique is that measuring these constructs first would result in "self-generated validity"[4]; prior measurements of attitudes toward repaying debt might cause one to think about those constructs and influence responses in a survey in a way that would not be true in a real-world situation where nobody is priming those moral thoughts.

c. One way to handle these possibilities would have been to counterbalance which came first. If the results were invariant to the order, results like those in Figure 16 would be more informative.

4. It might be possible to make use of machine learning techniques for understanding factors that affect the coefficients for treatment effects in randomized experiments.[5]

5. In Figure 3, the labels that describe the response options are so shortened from what was in the actual survey questions (in the header above the figure) that to a casual reader, they may appear to be a mistake. This could readily be remedied by adding more context from the answer options so that a casual reader knows that all of the options had to do with the timing of when a person must write or call a debt collector to dispute the debt.

6. There are a number of figures where a stacked bar graph might be more effective in displaying results (e.g., Figure 4, Figure 15, Figure 16, and Figure 19).

7. Adding a header to the notices tested, which are helpfully included in Appendix B, would enable readers to quickly discern which notice corresponds to which treatment in the analysis.

8. There is an open-ended question on the survey that asks respondents, "When you answered the two questions above, what were you thinking about?" (The two questions asked how likely a respondent would be to make a payment or ignore the notice). The only factor discussed in the report from this open-ended question was the mention of credit reporting. Expanding the discussion to include these other questions may be of value.

9. Section 3.7 is titled "Electronic Delivery"; however, by a wide margin, respondents preferred to receive these notices by postal mail (Figure 18). This section might be better titled "Method of Delivery," or something similar, in order to indicate multiple possible delivery methods and/or that the preferred method is by postal mail (not electronic).

---

[4] Feldman, Jack M. and John G. Lynch, Jr. (1988), "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology*, 73 (August), 421-435; Simmons, Carolyn J., Barbara Bickart, and John G. Lynch, Jr. (1993), "Capturing and Creating Public Opinion in Survey Research." *Journal of Consumer Research*, 20 (September), 316-329.

[5] Wager, Stefan, and Susan Athey (2018), "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523),1228-1242.

10. In Appendix Table 33, it might be useful to add row and column percentages for sample demographics to ease interpretation by public audiences interested in this report. Taking gender as an example, the four cells add to 100%; however, some audiences will be interested in knowing, for example, the percentages of females and males among those having debt collection experience.