



▲ Highland Park, Michigan, July 2010
▼ Lehigh Acres, Florida, May 2009 (Elvin Wyly)

Good Data, Good Deeds

Testimony prepared for “Regulation C: The Home Mortgage Disclosure Act (HMDA)
Public Hearing, August 5, 2010.”

Federal Reserve Bank of San Francisco, San Francisco, California

Elvin Wyly, University of British Columbia



Thank you, Director Braunstein, members of the Consumer Advisory Council, and dedicated staff, for the opportunity to comment on the revisions to Regulation C. My name is Elvin Wyly. I am Associate Professor of Geography and Chair of the Urban Studies Program at the University of British Columbia, in Vancouver, Canada.¹ I have used HMDA in my research since 1995.

Common sense tells us that a source of “data” should provide an accurate measure of reality. True, but there’s more: one of the dictionary definitions of “datum” is “a known fact,” and the word fact is Latin for “a thing done.”² A good data source like HMDA is *a thing done*. While it reflects certain aspects of reality, it also plays a role in producing other conditions of possibility. These are real too. Good data are good deeds.

Today, everyone across the political spectrum agrees that in the distant past, American mortgage markets were shaped by discrimination and redlining. This broad consensus did not come easy, however: it required many years of work to marshal the kind and volume of evidence that could not be dismissed. Before HMDA, financial institutions were able to deny the collective realities created by their individual actions. Residents of particular neighborhoods suffering disinvestment could point to local evidence of redlining and disinvestment, but without broad, systematic quantitative data from cities

¹ I have lived in Canada since 2002, but I remain a citizen, taxpayer, and registered voter in the United States.

² For an extended etymology and its application to housing data, see David Listokin, Elvin Wyly, Ioan Voicu, and Brian Schmitt (2003). “Known Facts or Reasonable Assumptions? An Examination of Alternative Sources of Housing Data.” *Journal of Housing Research* 13(2), 219-251.

nationwide, their claims were easily ignored.³ HMDA changed that. It allowed residents, public officials, advocates, researchers, and financial institutions themselves to measure the evolving inequalities among people, places, and various market segments across American cities and suburbs. (See the Appendix for a few examples of the wide range of outcomes that can be mapped across different cities, neighborhoods, and institutions). HMDA presented financial institutions with a new reality: a culture and infrastructure of public accountability, responsible and rigorous policy analysis, and informed, evidence-based organizing in the finest traditions of representative democracy.⁴ Unfortunately, most of the financial industry has fought disclosure every step of the way. Right now, lobbyists are working hard to shape the implementation of the Dodd-Frank legislation.⁵ The Board should resist these pressures, and should protect and expand the achievements of HMDA.

Are there successes in how HMDA has been implemented that should not be changed? Two successful features of HMDA should not be changed. First, HMDA is *extensive* -- it is a full enumeration rather than a sample, it is microdata rather than aggregate, and it now covers most institutions across all of metropolitan America, all the way down to the neighborhood scale. It is true that HMDA cannot tell us all the unique

³ For histories of the disclosure and reinvestment movement, see Gregory Squires, ed. (1992). *From Redlining to Reinvestment*. Philadelphia: Temple University Press; and Gregory Squires, ed. (2003). *Organizing Access to Capital*. Philadelphia: Temple University Press.

⁴ For an assessment of the record after the second-generation revisions to HMDA in 1989, see Allen J. Fishbein (1992). "The Ongoing Experiment with 'Regulation from Below': Expanded Reporting Requirements for HMDA and CRA." *Housing Policy Debate* 3(2), 601-636.

⁵ In the latest private-sector raid on the public treasury, industry lobbying firms are receiving bailouts in the form of influential human and political capital. Reports indicate that more than 500 officials have recently walked through the "revolving door" from finance-related government agencies to the private sector. Almost a third of these are former regulatory officials who have registered as paid lobbyists. Eric Lichtblau (2010). "Ex-Regulators Get Set to Lobby on New Financial Rules." *New York Times*, July 27, B1.

circumstances of individual borrowers. Yet every time the unique circumstances of undeniably predatory transactions are showcased in court cases or press accounts, critics dismiss the evidence *precisely because it is too detailed* -- it is “anecdotal,” and thus cannot be generalized beyond the unique circumstances of a particular case. Whatever else one can say about HMDA, it cannot be dismissed as anecdotal. It is the only publicly available data source that allows anyone to undertake econometric analyses of millions of transactions every year, revealing systematic inequalities among and within cities, suburbs, and neighborhoods.⁶ It is true that these extensive analyses, alone without additional evidence, cannot prove discrimination by particular lending institutions -- in large part because the industry for years resisted proposals to expand HMDA to include the very applicant risk factors cited as legitimate business-necessity reasons for observed disparities in credit outcomes. The Board should consider the logic carefully: detailed, well-documented individual cases of predatory exploitation are just anecdotal, while generalizable studies of millions of HMDA records are misleading because of the missing details on applicant risk factors. The only solution to this paralyzing dilemma is for the industry to accept the additional reporting burden of disclosing every data element required to paint an accurate, complete picture of the general risk factors and unique circumstances that are cited as justifications for observed credit outcomes.⁷ This will not

⁶ The extensive coverage of HMDA allows us to analyze 8.8 million conventional mortgages originated across the nation’s metropolitan areas in the peak year of the subprime boom, for example, and to determine that African Americans were 2.3 times more likely to wind up with rate-spread loans compared to otherwise identical non-Hispanic Whites, after controlling for income, loan-to-income ratio, loan purpose, lender type, securitization, and a proxy for credit history developed by researchers to deal with the absence of true credit information in the files. See Elvin Wyly, Markus Moos, Daniel Hammel, and Emanuel Kabahizi (2009). “Cartographies of Race and Class: Mapping the Class-Monopoly Rents of American Subprime Mortgage Capital.” *International Journal of Urban and Regional Research* 33(2), 332-354.

⁷ If history is any guide, anything the Board does to address the warnings raised by the industry will simply be used a few years from now to claim that HMDA cannot prove anything. This is exactly what happened

be easy.⁸ But it has been done before. For a successful model achieved nearly twenty years ago, the Board will certainly recall the explicit “search for omitted variables” of the Boston Fed Study.⁹ For a contemporary successful model that captures all of the dynamic innovations of the fast-changing subprime market of recent years, the Board should consider the data elements collected in the National Mortgage Data Repository built with support from the Ford Foundation and the National Consumer Law Center.¹⁰

The second success is freedom of access. The data are created by *public* law, and they are a *public* resource. The Board should maintain maximum, free public availability of as much data as possible. Privacy concerns will certainly place limits on the scope and detail of public data release (and I will address these below). But privacy must not be used as a tactic of fear, to undermine the principle of public use of the data to achieve the core purposes of HMDA, CRA, and related statutes.

when the industry resisted adding borrower credit scores back in 2002. Researchers and community reinvestment advocates recommended adding borrower credit information to HMDA in the 2002 Regulation C revisions, but industry lobbyists succeeded in killing the idea. When the new high-cost lending data were released in early 2005, lobbyists explicitly attacked the use of the data to infer inequality or discrimination, because there was no information on borrower creditworthiness. See Dan Immergluck (2004). *Credit to the Community: Community Reinvestment and Fair Lending in the United States*. Armonk, NY: M.E. Sharpe, pp. 216-219. See also the account of the fall 2004 lobbyist convention in Washington, DC devoted to spin tactics to be used upon the release of the data, in Elvin Wyly, Mona Atia, Elizabeth Lee, and Pablo Mendez (2007). “Race, Gender, and Statistical Representation: Predatory Mortgage Lending and the U.S. Community Reinvestment Movement.” *Environment and Planning A* 39, 2139-2166.

⁸ HMDA certainly involves reporting burden, particularly for small institutions. But reporting burdens must be evaluated in relation to the existing, privatized reporting burdens demanded by securitizers, investors, ratings agencies, and other private-sector actors. In many ways, the new data elements specified in Dodd-Frank simply harmonize HMDA with existing private data collection efforts and information disclosed to investors through SEC filings.

⁹ Lynne Elaine Browne and Geoffrey M.B. Tootell (1995). “Mortgage Lending in Boston: A Response to the Critics.” *New England Economic Review*, September/October, 53-78.

¹⁰ Elizabeth Renuart, Patricia A. McCoy, and Stephen Ross (2009). *National Mortgage Data Repository*. Boston, MA: National Consumer Law Center.

How do recent events change the way we should think about HMDA? Using HMDA to study mortgage markets is like surfing, riding the big, fast wave of market innovation -- but standing backward on the surfboard.¹¹ It may well be impossible to anticipate the next “innovations” of securitization and mortgage credit. But it is clear that the new data elements specified in Section 1094 of Dodd-Frank are not much more than what researchers recommended a decade ago. The Board, therefore, should begin with the new Dodd-Frank elements as an absolute minimum baseline. To provide an accurate picture of contemporary market realities, though, we also need information on

loan to value

combined loan to value

home equity lines of credit

reverse mortgages

balloon payments

cash-out refinances

front- and back-end debt to income ratios

loan modifications

Given the broad diversity of experiences of people of varied racial and ethnic identities¹² -- especially in the “Asian” and “Hispanic or Latino” categories -- HMDA race/ethnicity reporting should be harmonized with current Census Bureau procedures. Adding English-language ability, and the language in which the loan documents were prepared,

¹¹ This is the metaphor used to describe inter-generational relations in the “quantitative revolution” in the discipline of geography in the 1960s. See Peter R. Gould (2000). *Becoming a Geographer*. Syracuse: Syracuse University Press.

¹² See Wyly et al. (2007), “Race, Gender, and Statistical Representation,” *op. cit.*

would also further two of the three explicit purposes of the statute. HMDA reporting thresholds should be simplified and reduced, and should cover all entities making or purchasing any significant number of mortgage loans. Subsidiary structure was used to evade regulation during the housing bubble,¹³ and corporations are already reorganizing themselves to avoid parts of Dodd-Frank.¹⁴ Level the playing field for all in the market.

How should we balance public disclosure with privacy concerns? Privacy concerns are serious.¹⁵ Given the explicit mention of only two elements -- age and credit history -- in the privacy directives in Section 1094,¹⁶ the Board should preserve the historical continuity of today's Loan Application Register (LAR) format, as far as possible.¹⁷ To

¹³ Patricia A. McCoy and Elizabeth Renuart (2008). *The Legal Infrastructure of Subprime and Nontraditional Home Loans*. Cambridge, MA: Joint Center for Housing Studies, Harvard University. See also Elvin Wyly (2010). "The Subprime State of Race." In Susan J. Smith and Beverly A. Searle, eds., *The Blackwell Companion to the Economics of Housing: The Housing Wealth of Nations*. West Sussex, UK: Wiley-Blackwell, 381-413.

¹⁴ The most vivid example is GM's purchase of a subprime lender immediately after Dodd-Frank was signed into law with an exemption for auto dealers. See Nick Bunkley (2010). "GM Spends \$3.5 Billion for Lender to Subprime." *New York Times*, July 22, B3.

¹⁵ Even so, there is a rich irony in the fact that privacy concerns are so often raised by an industry that is built on personal surveillance systems that are far more invasive than anything ever contemplated for HMDA. The choices available to every consumer in this room are governed in large part by something each of us helped to create, but that is deployed about us, without our consent, among private corporations for their strategic purposes. Consumers have very few rights in this coercive data regime. Until a few years ago, the financial services industry insisted that consumers should not even be allowed to know the details of how consumers' actions resulted in particular kinds of data used to target them. I am referring, of course, to the ubiquitous FICO scoring algorithm. For a broader perspective on the profitability and use of consumer data, see Jon Goss (1995). "We Know Who You Are and We Know Where You Live: The Instrumental Rationality of Geodemographic Systems." *Economic Geography* 71(2), 171-198. For an update on how place-based data are being interwoven with the ecosystem of internet data sources, see Emily Steel and Julia Angwin (2010). "On the Web's Cutting Edge, Anonymity in Name Only." *Wall Street Journal*, August 4, p. A1, A2.

¹⁶ The specific directive is to "modify or require modification of itemized information, for the purposes of protecting the privacy interests of the mortgage applicants, that is or will be available to the public." *Dodd-Frank Wall Street Reform and Consumer Protection Act* (2010), Section 1094, p. 723. The "rules for modification" specify "credit score data" (sub-paragraph i) and age (sub-paragraph ii), but also include "any other category of data" in the new disclosure data elements "as the Bureau determines to be necessary to satisfy the purpose" of protecting the privacy interests of the mortgage applicants or mortgagors.

¹⁷ Some have argued that the global financial crisis was an epic event, a once-in-a-lifetime collapse on par with the Great Depression, and thus unlikely to be repeated anytime soon; others suggest that the slow and timid regulatory response -- including the limits and exemptions of Dodd-Frank -- are already setting the stage for another collapse. See Simon Johnson and James Kwak (2010). *Thirteen Bankers: The Wall*

minimize the risks of the new data elements, the Board should distinguish the concepts of privacy, confidentiality, and anonymity. Privacy refers to peoples' interest in controlling the information about them, while confidentiality specifies the rules of sharing certain kinds of information. Anonymity means the absence of personal identifiers.¹⁸

Regulation C already maintains anonymity and confidentiality. Unfortunately, HMDA records are *even now* associated with privacy risks -- a fact that has nothing to do with HMDA itself. The problem is that information in the *anonymized* HMDA files can be connected with *non-anonymized*, publicly available land records information filed in courthouse deed and lien records. To be sure, the new data elements will increase the number of things that can be discovered once an individual is identified; but the new data elements should not be blamed as the portals to individual identification.¹⁹ That problem lies elsewhere.

There are several technical and procedural choices the Board can consider in order to enhance the anonymity of the HMDA records themselves.²⁰ But in light of the existing

Street Takeover and the Next Financial Meltdown. New York: Random House. Reconciling these competing interpretations -- and their wide-ranging implications for the future of American housing and public policy -- requires a time series of data with as much consistency as possible.

¹⁸ Joan E. Sieber (2009). "Planning Ethically Responsible Research." In Leonard Bickman and Debra J. Rog, eds., *The Sage Handbook of Applied Social Research Methods*. Thousand Oaks, CA: Sage Publications, 106-142; see, especially, pp. 117-128.

¹⁹ The only exception would be if true, geo-coded parcel numbers were collected and included in the public release application records. This would obviously be a terrible idea, and unnecessary (see the next footnote).

²⁰ The Board has several options. First, the most sensitive data elements -- those most susceptible to reverse engineering -- can be separated and issued in a dual file structure -- similar to the approach used for the GSE public use databases. If this approach is followed, the Board should issue public use microdata files for counties and county equivalents, or metropolitan areas. Second, the Board can reserve the full details of the most sensitive data elements for regulatory purposes, and issue more simplified -- but still useful -- codes for the public release files. This approach, of course, is already used to exclude application dates from the raw public data. A variant of this technique involves converting continuous, interval-ratio data for such items as credit score or LTV into categorical or binary indicators -- making data mining much more difficult. Another variant is to take a 'raw' data element -- such as the proposed parcel identification number -- and direct the FFIEC to use the information to create a useful but anonymized synthetic variable

possibility of identifying individuals from sources other than HMDA, the Board needs to reconsider the meaning of privacy threats. In this area of policy and regulation, the threat comes not from the availability of anonymized information. The threat comes from *how the information is used*. And the most serious risks to privacy come from the private sector itself -- individuals and institutions, large and small, who have powerful economic incentives to find and target individual consumers for various kinds of commercial transactions.²¹ Some of these transactions are legal and fair. Some are technically legal, albeit unethical and far from optimal. And others -- such as identity theft -- are blatantly illegal and unfair, but nonetheless driven by the same profit motive that defines the essence of the private sector in the ruthlessly competitive American marketplace.

What does this imply for Regulation C? The Board can and should take steps to preserve and enhance the existing anonymity of the HMDA records. But further efforts will require the Board to rethink the broad meaning of public disclosure. Raw, anonymized application-level data should be provided quickly and freely to all members of the public

for public release. For the example of parcel number, a useful but anonymized variable would indicate the calendar year of the most recent loan made on the subject parcel. A third approach is risky and undesirable: wholesale geographic aggregation of major sections of the HMDA release. Geographic specificity down to the tract level has been one of the most important and valuable features for research on the highly uneven geography of credit, investment, and risk -- and so geographic aggregation should be avoided at all cost. If any aggregation is used, it should not simply default to a higher level in the standard Census hierarchy (i.e., replacing tract codes with county or metropolitan codes). Instead, confidentiality thresholds should be established based on combinations of the most sensitive variables, and aggregation used only in those census tracts with insufficient numbers of applications or mortgages. Tracts would be aggregated until thresholds for anonymity are exceeded -- with provisions in an attempt to ensure consistency for future years. In such a scenario, a loan record would have the full range of Dodd-Frank data elements, but if it were located in a sparsely-populated tract where a small number of applications risked privacy, the single tract code would be replaced by two, three, or more: we would only be able to tell that the application property was located somewhere in, say, tracts 248.1, 248.2, or 248.3.

²¹ Community reinvestment advocates have no interest in reverse-engineering HMDA to identify individual borrowers. Researchers are tightly regulated in any proposal to reverse-engineer the data: faculty and students at nearly all educational institutions across North America must strictly adhere to human subjects ethical protocols. Extensive reviews -- including specific procedures for confidentiality and/or anonymity - - are required prior to the commencement of any research involving any personally identifiable information.

wishing to use the data for the purposes specified in HMDA. These purposes are broad indeed: a wide variety of research, organizing, and advocacy activities are covered by the stated goals of helping determine whether institutions are serving community credit and housing needs, helping public officials target public investment, and assisting in identifying possible discriminatory lending patterns and enforcing antidiscrimination statutes. But targeting individuals for commercial, profit-driven schemes is clearly not one of these purposes.

The Board should therefore explore the legal possibilities of issuing parts of HMDA under a Creative Commons Non-Commercial Use License. The Board should explore this and any other strategy that would prohibit the use of the data for the purpose of identifying individual applicants or mortgagors. Narrow exceptions should be permitted for non-commercial purposes -- where the use of personal identification is a means to an end for research, local government policy development, or public interest law or journalism; these exceptions should be governed by reasonable equivalents to the human subjects ethical review procedures that already prevail for academic research involving any personally identifiable information.

Any individual using the data for the purposes clearly specified in HMDA, however, should be able to obtain the data quickly and freely: whatever procedures are instituted to guard against misuse of the data must be quick and efficient, and must maintain the open access principles that have made HMDA an important part of informed public debate on matters of credit, fair lending, and urban investment.

HMDA has been one of the most successful policy innovations of our lifetime. It promotes transparency, accountability, and the kind of market knowledge that nourishes fairness and efficiency. It reinforces a respect for the value of measurement, evidence, and reason. To the degree that the evidence is incomplete and the measurements are imperfect, Dodd-Frank gives the Board the opportunity to fix these problems. Regulation C should be modernized to capture the rich, vast landscape of American financial innovation. Expanded disclosures will increase our knowledge of a fast-changing market, and will allow us to spot danger signs earlier than ever before. Good data are good deeds. Please, do good deeds.

Respectfully submitted,

Elvin K. Wyly, Ph.D.
Associate Professor
Chair, Urban Studies Program
Department of Geography
The University of British Columbia
1984 West Mall
Vancouver, BC V6T 1Z2 Canada
778 899 7906
<http://www.geog.ubc.ca/~ewyly>
ewyly@geog.ubc.ca

APPENDIX

A small sample of the kinds of analysis made possible by the 2004-and-later generation of public-release loan application register files. The great success of HMDA, as amended, is the free public availability of anonymized data on most of the mortgage market, at the applicant and loan level, across the full range of city and suburban neighborhoods across metropolitan America. Extensive coverage and maximum public availability, subject to privacy protections, should be preserved.

1. Debt leverage and foreclosure risk, New York City census tracts. Elvin Wyly, Kathe Newman, Alex Shafran, and Elizabeth Lee (forthcoming). "Displacing New York." *Environment and Planning A*.
2. Black-White disparities in subprime lending across counties in the continental U.S. Joe T. Darden and Elvin Wyly (2010). "Cartographic Editorial: Mapping the Racial/Ethnic Topography of Subprime Inequality in Urban America." *Urban Geography* 31(4), 425-433.
3. Latino-White disparities in subprime lending across counties in the continental U.S. Joe T. Darden and Elvin Wyly (2010). "Cartographic Editorial: Mapping the Racial/Ethnic Topography of Subprime Inequality in Urban America." *Urban Geography* 31(4), 425-433.
4. Black-White subprime inequalities by metropolitan area, after accounting for estimated credit risk and other controls. Elvin Wyly, Markus Moos, Daniel J. Hammel, and Emanuel Kabahizi (2009). "Cartographies of Race and Class: Mapping the Class-Monopoly Rents of American Subprime Mortgage Capital." *International Journal of Urban and Regional Research* 33(2), 332-354.
5. Segmentation by borrower income and securitization network, by metropolitan area, after accounting for estimated credit risk and other controls. Elvin Wyly, Markus Moos, Daniel J. Hammel, and Emanuel Kabahizi (2009). "Cartographies of Race and Class: Mapping the Class-Monopoly Rents of American Subprime Mortgage Capital." *International Journal of Urban and Regional Research* 33(2), 332-354.
6. Racial-Ethnic Segmentation and Lender Specialization. Elvin Wyly (2010). "The Subprime State of Race." In Susan J. Smith and Beverly A. Searle, eds., *The Blackwell Companion to the Economics of Housing: The Housing Wealth of Nations*. West Sussex, UK: Wiley-Blackwell, 381-413