Racial Inequality, Data Politics, and Federalism

A First Glance at a Small but Important Sample of the New HMDA Data

A First Glance at A Small but Important Sample of the new HMDA Data

Elvin K. Wyly
University of British Columbia

September 9, 2005
ewyly@geog.ubc.ca

Draft; please request permission prior to citing or quoting.

Politics and Data: Continuity and Change

The thirtieth anniversary of the Home Mortgage Disclosure Act (HMDA) of 1975 turned out to be a busy year, illustrating once again the pervasive parallax of data in American public policy debates. Empirical indicators designed to measure socially-defined (but initially invisible) policy problems are developed and implemented, creating a flood of interest, concern, and criticism. When the indicators are revised and refined in later years, the measures are selectively redefined amidst the altered context of evolving public concerns, regulatory priorities, and socio-economic change.

Today's debates represent the third major shift in the politics of HMDA. The first generation of data created and organized by the Act established a vocabulary that enabled the first systematic public discussion of racial-geographic redlining, but over the next dozen years conservative analysts and industry advocates launched increasingly strident challenges to redlining research. Critics charged that spatial disparities provided no evidence of redlining, because the data only captured the volume and location of approved loans, obscuring the role of market demand (e.g., variations in application volumes and approval rates, differences in borrower incomes). The 1989 amendments responded directly to those criticisms, inaugurating a second generation of policy and regulatory debate in the aftermath of potent and unexpected results from the Boston Fed Study. After a long period of regulatory avoidance and dismissal in the Reagan years, debates over the new loan-level data were revived when the Clinton Administration sought to use a fairly conservative, bipartisan tool (homeownership) to achieve traditional Democratic goals. The conservative backlash siezed on the absence of credit history and asset data in the HMDA files. HMDA also provided no information whatsoever on loan terms, making it impossible to measure consumer-lender negotiations designed to arrange a mutually acceptable transaction. Meanwhile, each year that passed after the Boston Fed Study weakened the policy implications of its aging findings, which documented substantial racial disparities that persisted despite the inclusion of an extraordinary array of underwriting variables. Finally, after repeated delays, the latest generation of HMDA has revealed persistent and pervasive racial disparities in the cost of credit. Once again,

the disparities are blamed on differences in credit, and the credit disclosures resisted by the industry lobby are now cited as a conveniently confidential factor that explains away all evidence of discrimination. For thirty years, limited but important improvements in public data have invariably led to increasingly sophisticated explanations for why observed disparities are no cause for concern. In the face of these changes, the underlying disparities have remained surprisingly durable.

In this paper, I take advantage of an unusual opportunity to jump across the trenches dug in this Thirty Years' War. The public release HMDA files suffer from many limitations, but they offer a crucial first view of inequalities in credit. Not long ago, several large banks had the opportunity to resolve the limitations of HMDA-based studies by sharing precisely those data elements that the institutions claim exonerate them on charges of discrimination. The banks instead went to court to challenge the request on the grounds of federalism and jurisdictional precedent.

I begin with a review of the legal struggle, followed by an econometric analysis of the lending records of the banks involved.

Clearing House v. *Spitzer*

In early 2005, the Civil Rights Bureau of the Office of the Attorney General (OAG) of the State of New York began to review the newly-available loan-pricing data available in HMDA. After finding substantial racial disparities, Eliot Spitzer's Bureau Chief, Dennis D. Parker, sent two letters on April 19 to Wells Fargo and HSBC Bank USA; letters were sent the next day to JPMorgan Chase and Citigroup. All letters to these banks (henceforth, "the Banks") were substantively identical. In the letter to Wells Fargo, Parker noted that "Published reports analyzing Wells Fargo's 2004 HMDA data suggest that African-American and Hispanic borrowers are substantially more likely than white borrowers to receive higher-cost loans .... These ... disparities are troubling on their face, and unless legally justified may violate federal and state anti-discrimintion laws...."[1] Parker went on to state that the OAG had begun "a preliminary inquiry ... to determine the causes of these disparities" before asking that Well Fargo "voluntarily provide ... certain information regarding its HMDA reportable loans and applications."[2] Parker requested two separate categories of data: 1) 2004 HMDA loan application register (LAR) records for property located in the State of New York, and 2) several types of non-public data, including applicant credit score and other variables used to determine APRs, underwriting "formulas or algorithms," detailed loan terms, any internal analyses of racial disparities in Wells Fargo's 2004 HMDA loans, and policies or procedures regarding exceptions to the lender's standard underwriting formulas.

Wells Fargo complied with request for the first category of data (as required by Regulation C) and a few weeks later supplied non-public lending information for its non-bank subsidiaries; at the same time, Wells Fargo told Spitzer's office that it would provide the non-public data for its

---

[1] Dennis D. Parker (2005). *Letter to Richard M. Kovacevich, President and CEO, and James M. Strother, Executive Vice President and General Counsel, Wells Fargo & Company*. April 19. Albany and New York City: New York Department of Law, p. 1.

[2] Parker, *Letter to Kovacevich and Strother*, p. 1.

national bank subsidiaries directly to the Office of the Comptroller of the Currency (OCC). The General Counsel of Wells Fargo Home Mortgage then informed the OCC of Spitzer's request; after a few weeks of discussions between the OCC, the OAG, and Wells Fargo, Parker wrote again on June 6 asking for voluntary cooperation; otherwise, Parker stated, Spitzer "would either subpoena the lending information or file an action in New York State Court against Wells Fargo regarding Wells Fargo Bank's lending practices."[3]

On June 16, the OCC and an industry trade group representing the Banks (the Clearing House Association) filed suit in U.S. District Court in Manhattan, charging that Spitzer's actions had no authority, and that the OCC's supervisory role pre-empted state investigation or enforcement against OCC-regulated banks. The Clearing House suit begins by invoking the National Bank Act of 1864, which provides that "No national bank shall be subject to any visitorial powers except as authorized by Federal law…." and proceeds to cite recent interpretive regulations by the OCC that equate "visitation" with the conduct of examinations or requiring various kinds of records or documents.[4] The Clearing House presents the challenge as "a purely legal issue regarding the primacy of federal law with respect to national banks,"[5] but the *Complaint*'s recitation of negotiations between Spitzer's office and the Banks hints at deeper concerns over disclosure. The Banks have provided the public HMDA LAR data to Spitzer's office, the *Complaint* states, but the additional data "sought by the Defendant is, however, nonpublic lending information that the Clearing House Members are not required to disclose to state officials,"[6] and therefore the Banks refused.[7] The *Complaint* cites "continuous regulation, supervision, examination, and monitoring by the OCC,"[8] and asserts that "By seeking to inspect the books and records of the Clearing House Members and to otherwise exercise visitorial powers over them, the Defendant threatens to increase - and if not enjoined will increase – the compliance burden faced by the Clearing House Members in a manner expressly prohibited by federal law."[9]

Despite the sharp, narrow focus on visitorial powers and the primacy of federal law, the *Complaint* nevertheless includes an attempt to justify the racial-ethnic disparities cited as *prima facie* evidence of discrimination:

> "Either directly or through their operating subsidiaries, the Clearing House Members each engage in residential mortgage lending. Their mortgage lending programs are designed to promote home ownership and, consistent with the Community Reinvestment Act, include mortgage lending to less creditworthy

[3] David L. Moskowitz (2005). "Declaration of David L. Moskowitze." *Clearing House v. Spitzer* (05-CV-5629), June 14. New York: U.S. District Court, Southern District of New York, p. 4.
[4] The Clearing House Association, L.L.C., Plaintiff, v. Eliot Spitzer, Attorney General of the State of New York, Defendant (2005), *Complaint*, June 20. New York: U.S. District Court, Southern District of New York, p. 1.
[5] Clearing House *Complaint*, p. 2.
[6] Clearing House *Complaint*, p. 6.
[7] There is a discrepancy on this point between the Complaint (stating that the Banks refused), later public statements by Spitzer (stating that only Wells Fargo refused), and the declaration by the Chief Counsel of Wells Fargo Home Mortgage (stating cooperation with regard to its non-bank subsidiaries). See Moskowitz, "Declaration," p. 3.
[8] Clearing House *Complaint*, p. 8.
[9] Clearing House *Complaint*, p. 9.

> borrowers. *As a matter of prudent and sound lending practice, loans to borrowers with different risk factors bear different interest rates*."[10]

U.S. District Court Judge Sydney Stein denied the Clearing House request for immediate relief, declining to issue an injunction against Spitzer's data requests until he had time to rule on the merits of the suit; but Stein also refused to force the Banks to turn over the data, and he stated that he would reconsider the injunction if Spitzer pursued subpoenas prior to Stein's ruling on the merits of the case.[11] Stein issued an order for Spitzer to show cause, and scheduled arguments.

Spitzer's response came on August 5, in a counterclaim with supporting briefs by the attorney generals of 34 states and a coalition of public interest groups, including: the National Community Reinvestment Coalition, the Center for Responsible Lending, the National Fair Housing Alliance, AARP, NAACP Legal Defense and Educational Fund, and a half-dozen advocacy organizations and service providers in New York State. The National Association of Realtors and the New York Association of Realtors also filed a supporting brief.[12] Spitzer's arguments responded directly and forcefully to the federalism, jurisdictional, and definitional aspects of the Clearing House and OCC suits. Spitzer argued that a) the National Bank Act does not prohibit *state* enforcement of *state* laws regulating the activities of national banks, b) Congress explicitly vested authority in state attorneys general to bring cases under the federal Fair Housing Act, and c) the OCC's pre-emption actions constitute an attempt to shield national banks from federally-recognized state authority, in both legislative and enforcement powers.[13] Spitzer also challenged the Clearing House definition of the activities at issue:

> "As history and case law reveal, 'visitorial powers' refers to the agency activity of administratively supervising the affairs of a national bank to ensure safety and soundness, plus compliance with relevant laws. To the extent a suit by the Attorney General involves the exercise of 'visitorial powers,' it is the court that performs the visitation by ordering disclosures and remedies."[14]

'We have not gotten complete data from anyone'

Public statements by Spitzer and his staff underscored two key issues in the struggle over disclosure. First, Parker described the Banks' behavior as "stonewalling" since late May or early June, when it became clear that they would get support and relief from the OCC (and, they hoped, the federal judiciary).[15] Second, Spitzer and Parker both described the early stages of the

---

[10] Clearing House *Complaint*, p. 4, emphasis added.
[11] Gemma Westacott (2005), "Banks Denied Order Against Spitzer," *The Lawyer.com*, July 25.
[12] New York Department of Law (2005), "Broad Coalition Joins Spitzer in Opposition to Effort to Shield Banks from Scrutiny." Press Release, August 5. Albany and New York City: New York Department of Law.
[13] Eliot Spitzer (2005), "Defendant's Memorandum of Law in Opposition to Plaintiffs' Request for Injunctive Relief and Declaratory Relief and in Support of Counterclaim." *Clearing House v. Spitzer* (05-CV-5629), *OCC v. Spitzer* (05-CV-5636), August 5. New York: U.S. District Court, Southern District of New York.
[14] Spitzer, "Defendant's Memorandum," pp. 1-2.
[15] Quoted in Eric Dash (2005), "Spitzer Says Banks are Resisting Inquiry on Lending Practices." *New York Times*, August 6, p. C1.

investigation as involving a considerable amount of communication and negotiation; federalism and jurisdictional issues remained in the background during a period of initial, hesitant, and partial cooperation: "With the exception of Wells Fargo, all of the banks initially promised cooperation and provided some of the requested materials," Spitzer wrote in his press release; "The cooperation ended abruptly on June 16 with the filing of lawsuits" by OCC and the Clearing House.[16] Parker drew attention to this chronology, and went on note how the abrupt shift had ended the flow of information; early on, some banks had supplied a few materials, but Parker stated that "We have not gotten complete data from anyone."[17]

Parkers' remark should encourage us to reconsider the politics of data woven into the legal arguments in the Clearing House *Complaint*. Three tensions and contradictions stand out. First, if it is true that "As a matter of prudent and sound lending practice, loans to borrowers with different risk factors bear different interest rates,"[18] then the Banks had a powerful incentive to provide the applicant and underwriting information to Spitzer on an informal, voluntary basis. If the data would exonerate them, the Banks had a clear interest in negotiation, even if it were to become adversarial. If the Banks accepted Spitzer's stated motives as legitimate and transparent, voluntary disclosure would convince him of the business-necessity defense against allegations of discrimination, such that litigation would be seen as unwarranted. If the Banks harbored suspicions of Spitzer's political motives, however, then voluntary disclosure would convince Spitzer that a discrimination case would likely fail; the Banks would have benefited by disclosing the results of the detailed internal loan modeling that all large banks were advised (by lobbyists and the OCC itself) to undertake in preparation for the barrage of publicity associated with the new loan-pricing data.[19] Either way, the Banks could easily have provided a considerable amount of non-public information carefully selected in order to mount a defense against Spitzer's *suggestion* of a *possible* anti-discrimination enforcement action.

Second, the Banks could have pressed the Clearing House to argue that all proprietary data requests are inherently burdensome, regardless of the regulatory status of the financial institution receiving such requests. But the effort to use the OCC as a federal shield required a sharp focus on regulatory jurisdiction and division of powers; thus the *Complaint* "does not challenge the authority" that Spitzer has "over other companies that are not chartered by the federal government" and also does not assert that Spitzer "has been unreasonable or unfair in his enforcement efforts against such other companies."[20] The implication is that a state attorney general's data request threatens to increase the compliance burden only for those institutions

[16] New York Department of Law, "Broad Coalition," p. 1.
[17] Dash, "Spitzer Says," p. C1.
[18] Clearing House *Complaint*, p. 4.
[19] It is also crucial to note that if Spitzer's investigation was motivated by simple publicity goals, then the *prima facie* disparities are sufficient for a full-court press. The political case for public-opinion victories against the Banks is only weakened and obscured by requesting precisely those things that the Banks claim will exonerate them, including detailed "policies and procedures concerning the circumstances under which the APR offered to a loan applicant may depart (upward or downward) from the rate determined by the application of any formulas or algorithms" used in underwriting. Parker, *Letter to Kovacevich and Strother*, p. 2.
[20] Clearing House *Complaint*, p. 2.

chartered as national banks (and who are reporting to an agency that is currently trying to pre-empt state authority).

Third, the Banks could have adopted an explicit position on the relevance, reliability, or meaning of HMDA data on questions of discrimination. If the Banks asserted that public disclosure of HMDA LARs provided all the information required to demonstrate the Banks' compliance with relevant laws, they would have greatly strengthened their compliance burden argument by offering the interpretation that even infinite disclosure would provide no meaningful evidence. Alternatively, asserting that HMDA data are completely useless, even for a *prima facie* allegation of discrimination, would allow the Banks to undercut the rationale for Spitzer's initial data request. The former strategy would, of course, risk granting credibility to discrimination allegations already circulating in the public domain, while the latter strategy would invest considerable legitimacy in the public-interest goal of scrutinizing the Banks' non-public lending information. It remains to be seen whether the Clearing House will adopt an explicit position on the use of HMDA in oral arguments in September; but in the first salvo, the only solution to the dilemma was to maintain complete silence on the question. Moreover, the Plaintiffs' response to Spitzer – while mounting forceful challenges on the interpretation of the National Bank Act, the Congressional and regulatory definition of "visitation," and the scope of pre-emption exceptions under the Fair Housing Act – remains silent on the substantive question of the *prima facie* disparities. The Reply includes a single line on the issue in the context of a challenge to Spitzer's standing to sue as an "aggreived person" under the FHAct: "defendant's investigation is based solely on general HMDA statistical data."[21] By skirting the obvious questions (What, exactly, are the flaws in "general HMDA statistical data"? What explains the large disparities? Are the disparities the result of the Banks' practices, or the result of systematic data biases of some sort?) the Clearing House suit erases all meaning from Spitzer's data request. The implication is that non-public loan data would neither confirm nor deny the existence of discrimination. The intense legal-laser focus on visitation and federal pre-emption strips the data – and the large racial disparities they indicate – of all context and meaning.

Beyond First Appearances: Measuring and Modeling Racial Disparities

The OAG's analysis of 2004 HMDA disclosures "found substantial racial disparities in interest rates charged by various banks on their New York State loans."[22] African American borrowers at Wells Fargo, for instance, were three times more likely as whites to receive high-cost, "rate-spread" loans; this ratio was nearly two at JP Morgan Chase and Citigroup, and about 1.5 at HSBC. For our analysis, we consider the combined effects of gender and racial-ethnic disparities in the nationwide business of the Banks and their subsidiaries – a total of 1.9 million originations by a total of 217 operating units (Table 1).[23] Rate-spread loans account for about one in six loans made

[21] The Clearing House Association, LLC (2005), "Reply Memorandum in Further Support of Motion for Preliminary and Final Injunction." *Clearing House v. Spitzer* (05-CV-5629), August 19. New York: U.S. District Court, Southern District of New York, p. 17.

[22] Spitzer, "Defendant's Memorandum," pp. 5-6.

[23] We excluded loans purchased by the institutions, as well as 180,018 records without information on applicant income. The Banks have a combined total of 218 operating subsidiaries reporting HMDA information, but one reported five applications with no loans originated.

by the Banks overall, but this proportion varies widely among different groups of borrowers – from a low of one in twenty-five Asian couples (including several other ethnoracial groups that tend to be under-represented among homebuyers and owners) to about two out of five African Americans. If we adjust the proportion of rate-spread loans to compare groups with whites of the same family type, we find the sharpest disparities for non-Hispanic African American couples, who are more than three times as likely to receive high-cost loans. This ratio dips to 2.4 or 2.3 for single African American borrowers. Hispanic couples are almost twice as likely to get high-cost loans as their Anglo white counterparts, while there are no apparent disparities for single Latinas or Latinos.

As Spitzer's response to the Clearing House and OCC suits made clear, racial disparities of this magnitude "would establish a *prima facie* case" under the New York Human Rights Law "as well as federal anti-discrimination law."[24] Parker's negotiations with the Banks in April and May, 2005 were calibrated precisely to investigate the economic and business-necessity rationales (including, *inter alia*, applicant credit scores and underwriting formulas) that are frequently cited to explain minorities' credit outcomes and to maintain legal defenses against claims of discrimination. The Banks' stonewalling after the suits in mid-June has (at least temporarily) denied Spitzer access to the confidential loan file data, and of course the publicly-disclosed HMDA files have very limited underwriting data. Yet it is possible for us to control for several key variables – loan type, purpose, applicant income, the ratio of the loan amount to income. It is also possible to measure certain aspects of applicants' credit histories. Under most circumstances, when lenders reject a loan application they have the option of reporting up to three reasons for their decision. For the Banks under investigation, four-fifths of all rejections include at least one reason for the decision, and the leading justification is credit history (Table 2).

These codes allow us to create an individual measure of credit history for each loan request in the entire dataset. We estimate a logistic regression model to identify the distinctive characteristics of applicants who are rejected *and for whom the underwriter cites credit history as one of the reasons*. The parameters of this bad-credit model (Table 3) may then be used to calculate the similarity of each applicant to those people who are specifically identified by lenders as unacceptable credit risks.[25] This instrumental variable approach is a common econometric tool, but it only became possible to use it in HMDA-based analyses when the second-generation loan-level data became available in the early 1990s.[26] The instrument reduces the 'raw' disparities in rejection rates for Whites, Blacks, and Hispanics, and therefore addresses the problems of omitted-variable bias; yet the instrument does not completely eliminate the gap, and in this regard it closely mirrors the results obtained from the actual credit measures used in the Boston Fed study. The credit history instrument has been subjected to considerable testing and validation in several studies of accept/reject disparities,[27] and in at least one study of market segmentation using Randy Scheessele's HUD

---

[24] Spitzer, "Defendant's Memorandum," p. 6.

[25] To avoid circularity, we estimated the bad-credit model on a random sample of about ten percent of all applications.

[26] Andriana Aboriotes, Sarita Ahuja, Helen Feldman, Carol Johnson, Lekha Subaiya, Nathan Tiller, Julie Urban, and Samuel L. Myers, Jr. (1992), "Disparities in Mortgage Lending in the Upper Midwest: Summary of Results Using 1992 Home Mortgage Disclosure Act Data." Paper presented at the Fannie Mae University Colloquium on Race, Poverty, and Housing Policy. Minneapolis, MN: December 3.

[27] Samuel L. Myers, Jr., and Tze Chan (1995), "Racial Discrimination in Housing Markets: Accounting for Credit Risk." *Social Science Quarterly* 76 (3), 543-561. Steven R. Holloway (1998), "Exploring the Neighborhood Contingency

subprime list.[28] It is only possible now with the new rate-spread disclosures to use the credit instrument to deal with omitted-variable bias when measuring racial disparity in the segmentation of particular borrowers into high-cost loans.

The credit instrument helps to capture many of the justifications offered by the lending industry for observed patterns of market segmentation (Table 3). Higher-income applicants are less likely to be denied for reasons of bad credit, whereas credit problems seem to be more prevalent among borrowers requesting loans with FHA insurance, for mobile homes, or in the home improvement and refinance markets. After controlling for income, African American borrowers are almost twice as likely as whites to be denied for bad credit, while for Hispanics the ratio is 1.4. This instrument is far from perfect, of course: we would like a better model fit, and ideally we would have access to several consistent metrics of creditworthiness reported by underwriters for each application. The lack of credit data, however, is a direct result of the industry's opposition to the inclusion of credit scores in the expanded HMDA reporting requirements. Moreover, the optional and unregulated features of the denial codes effectively give the benefit of the doubt to lenders and underwriters. Credit history is universally recognized as the unassailable legal and political defense against allegations of discrimination; lenders have a powerful incentive to cite this justification, therefore, not only where there are legitimate concerns about credit but also in cases that might involve unintentional bias, disparate impact discrimination, or blatant disparate treatment. Using the credit history instrument as a control variable, therefore, creates a bias against any finding of discrimination.[29]

---

of Race Discrimination in Mortgage Lending in Columbus, Ohio." *Annals of the Association of American Geographers* 88(2), 252-276. Samuel L. Myers, Jr. (2000), *The Effects of Government-Sponsored Enterprise (GSE) Secondary Market Decisions on Racial Disparities in Loan Rejection Rates*. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Steven R. Holloway and Elvin K. Wyly (2001), "The 'Color of Money' Expanded: Geographically Contingent Mortgage Lending in Atlanta, Georgia." *Journal of Housing Research* 12(1), 55-90.

[28] Elvin K. Wyly, Mona Atia, and Daniel J. Hammel (2004), "Has Mortgage Capital Found an Inner-City Spatial Fix?" *Housing Policy Debate* 15(3), 623-685.

[29] It is also worth noting that even the most detailed proprietary measures of consumer credit history cannot be taken as some sort of sacred, objective, or unproblematic mechanism for settling questions of discrimination. No matter how it is measured, credit history is an instrumental variable just like our bad-credit equation. "Although credit projections, not histories, are what should be relevant to the mortgage lending decision, lenders use the applicant's credit history report as an instrument to measure ... commitment to debt repayment. Furthermore, loan underwriters tend to view certain elements of the credit report, such as public records of defaults or bankruptcies, as more important than others." Alicia Munnell, Geoffrey M.B. Tootell, Lynn E. Browne, and James McEneany (1996), "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86, 25-53, quote from p. 28. To further complicate the issue, there is considerable cross-lender variation in marketing and broker referral networks (creating corresponding variations in applicant credit histories) and specific credit thresholds or mitigating factors used in underwriting. In the prime market, secondary purchasing guidelines have standardized credit requirements, but the massive growth in subprime activity has magnified heterogeneity among lenders -- and among their specialized subsidiaries. To capture the effects of this variation, the instrumental variable approach might actually be superior to the detailed payment history data used in the Boston Fed study and several subsequent models based on proprietary data. Our approach ignores the ambiguity in determining which specific factor matters to an underwriter (e.g., recent judgment, late credit card payments, high credit balances, low FICO score, etc.) and simply accepts the final judgment rendered by each underwriter. As such, our instrument models what lenders say they care about, not what theoretical economists say lenders should be doing.

Results

To test whether racial-ethnic and gender disparities in high-cost credit can be fully explained by legitimate underwriting considerations, we compare the results of three models: the first predicts the likelihood that a loan will be rate-spread solely on the basis of applicant protected categories (gender, race/ethnicity); the second adds income, loan type, and other available underwriting factors; and the third adds the instrumental variable measuring bad credit (Table 4).[30] Without considering income or other underwriting factors, applicant race and gender provide a rather poor basis for predicting whether a loan will be high-cost; note the weak model fit diagnostics for Model 1 in Table 4. Nevertheless, the addition of income and loan variables strengthens model fit dramatically while exerting only modest effects on racial disparities (see Model 2, Table 4). Compared to single White Non-Hispanic males, the rate-spread odds ratio for single African American males slips from 3.2 to 2.7, and for females from 3.6 to 2.6. For African American couples and other applicant types, there is no meaningful change in the disparities. Similarly for Latinas and Latinos, controlling for income and loan variables has mixed effects depending on applicant type, with the most notable declines for single Latinas and for couples.

The third model offers the most conservative interpretation, putting full faith and credit in underwriters' judgments of bad credit. As expected, the credit instrument is a strongly significant and positive predictor of rate-spread segmentation: a one-standard deviation increase in the probability of bad credit increases the odds of receiving a rate-spread loan by a factor of 1.25. This result is entirely consistent with the prevailing economic, policy, and legal consensus that credit history plays an important role in subprime market segmentation.[31] This factor is insufficient, however, to explain or justify persistent and pernicious racial and ethnic disparities. To the degree that 'raw' disparities calculated without consideration of any mitigating factors can establish a *prima facie* case for investigations of potential violations of law, our model strengthens the case in three ways.

First, the magnitude of the credit effect falls well short of that observed for many other variables in the model; note that higher odds ratios appear not only for relatively rare events or conditions

---

[30] Multicollinearity tests for these models are encouraging. Tolerance values (one minus the coefficient of determination predicting each independent variable with all of the other model regressors) are typically regarded as robust so long as they exceed 0.20. For Mode 1, twelve of the predictors yield tolerances over 0.85, while none of the remaining three regressors have values below 0.48. Model 2 gives similar results, with the lowest tolerance value (0.47) for other non-Hispanic White applicant types. Model 3, however, reveals the distinctive credit conditions that prevail in the home improvement market: tolerance values fall to 0.17 for the renovation indicator, and to 0.12 for the credit instrument. Both of these values are cause for conce4rn, but as Scott Menard observes, "The good news about collinearity is that it is easy to detect. The bad news is that there are few acceptable remedies for it. Deleting variables involved in collinearity runs the risk of omitted variable bias." Scott Menard (2002), *Applied Logistic Regression Analysis, Second Edition*. Thousand Oaks, CA: Sage Publications, p. 77. My approach is to present cumulative model results side-by-side in Table 4, to illustrate the tradeoffs between collinearity and omitted-variable bias across alternative model specifications.

[31] See, for example, Pennington-Cross et al.: "...the results are quite clear – households that exhibit characteristics of high credit risk are more likely to use subprime lending." Anthony Pennington-Cross, Anthony Yezer, and Joseph Nichols (2000), *Credit Risk and Mortgage Lending: Who Uses Subprime and Why?* Working Paper 00-03. Washington, DC: Research Institute for Housing America, quote from p. 16.

accounting for small proportions of loans,[32] but for every applicant type where the primary borrower is African American. Second, the credit instrument performs robustly in relation to other loan characteristics, capturing the expected broad variations in creditworthiness that distinguish various submarkets. Nearly a quarter of all rate-spread loans involve home improvements, for example, and after accounting for income and other factors, renovations are over ten times more likely than purchase loans to be high-cost (Model 2, Table 4). Part of this gap, however, reflects the riskier credit profile of consumers in the home improvement market, and thus including the credit instrument reduces the odds ratio by almost half (Model 3, Table 4). Controlling for credit also helps to capture some of the rate-spread segmentation in the mobile home market. Still, credit does not explain the overwhelming focus of high-cost lending to lower-income borrowers: adding the credit instrument reduces the income parameter by only 7.4 percent. Third, the most pervasive and historically entrenched racial divisions persist even after accounting for credit. Depending on whether they apply alone or with co-applicants, African Americans are between 2.0 and 3.3 times more likely than single White males to receive high-cost loans. The inclusion of credit certainly does reduce the apparent disparities: adding the instrument reduces the coefficient for African American women by 25 percent, for Black men by 22 percent, and for male-female couples by 18 percent. Yet even after accounting for income as well as a measure of credit risk that may itself include a certain amount of disparate treatment or disparate impact discrimination, the Black-White disparity remains substantial and pernicious: African Americans are *at least* twice as likely to wind up with high-cost loans.

Third, applicant and loan characteristics are inadequate for understanding credit outcomes. Lending industry structure and specialization must also be considered. In their comprehensive re-examination and extension of the Boston Fed study, Ross and Yinger review several studies showing how racial disparities in loan rejection can "virtually disappear" by allowing for different lenders' use of unique combinations of underwriting factors; Ross and Yinger caution that "If this type of variation does not have a legitimate business justification … then the results in these studies demonstrate that lenders can transform disparate-treatment discrimination, which is easily observed as a significant coefficient for the minority status variable, into disparate-impact discrimination, which is buried in the coefficients of the idiosynchratic underwriting variables."[33] My analysis confirms that this analytical danger is even more severe for the case of market segmentation among borrowers who are approved and receive loans. Extending Model 3 to include a vector of dummiesw for each of the 217 operating subsidiaries of the Banks greatly improves our ability to predict rate-spread segmentation. Model pseudo-R-squared jumps from 0.38 to 0.63, while the percent correctly classified goes from 85.8 to 93.8. But just as Ross and Yinger warn, accounting for the idiosynchratic strategies of different subsidiaries has the effect of masking disparate impacts behind seemingly-neutral coefficients; the codes for the varied and specialized operating units of the Banks become, quite literally, dummy variables. Adding the subsidiary vector reduces the odds ratio for African American couples, for example, from 2.7 to

[32] For example, about 6 percent of all rate-spread loans involved home purchase requests for pre-approval, compared with only 2.3 percent of loans below the rate-spread threshold. This difference is quite stark (an odds ratio of 9.92 after controlling for all other effects) but it only applies to about 55 thousand loans out of a total of 1.9 million.

[33] Stephen L. Ross and John Yinger (2002), *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement.* Cambridge, MA: MIT Press, pp. 178-179.

1.3. Accounting for subsidiary differences even undermines the apparent importance of credit, reducing the standardized odds ratio from 1.25 to 1.10.

On purely methodological grounds, this approach is hard to justify. Controlling for industry heterogeneity seems reasonable, but adding more than two hundred variables to a predictive model takes us perilously close to tautology. More than half of the operating subsidiaries of the Banks made no rate-spread loans at all, while high-cost loans comprised at least three-quarters of all originations for 68 subsidiaries; knowing which unit an applicant is dealing with, in other words, is highly predictive of whether the homeowner will end up with a high-cost loan. Nevertheless, the enormous variation across bank subsidiaries is not some obscure econometric bias; it is a central element of the story. Subsidiary structure provides a simple road map to marketing and competitive strategies, and to the subsequent racial and ethnic composition of the subprime market. A few of the relevant connections are apparent if we plot rate-spread specialization as a function of African American specialization (see Figure 1). Note that the vast majority of mortgages originated by the Banks involve a small number of large divisions -- the large circles in the lower-left hand corner of the graph -- engaged in prime business with proportionately few African American customers. The bottom right-hand quarter of the graph is completely barren: among more than two hundred operating units of the Banks, none specialize in prime lending to the African American market. On the other hand, the top of the graph includes a broad continuum of small, specialized units focused almost exclusively on subprime lending. Many of these subsidiaries are oriented to non-Hispanic Whites, while a few others are geared more towards Latinos. But a few dozen focuse heavily on the African American market. Overall, the configuration at first appearance implies a certain equity, with high-cost units targeting the full diversity of American housing markets. But while some Whites end up at small specialized high-cost units, the vast majority of Whites have options at the large, prime divisions. Thirty-eight percent of all African Americans dealing with the Banks received rate-spread loans, compared to fewer than 15 percent of non-Hispanic Whites.

Table 1. Rate-Spread Segmentation for Citigroup, HSBC, JP Morgan Chase, and Wells Fargo.

	Number of originations	Share of group receiving rate-spread loans	Disparity Ratios: Compared with lenders' overall rate-spread business	Disparity Ratios: Compared with White applicants of same type
Single White Males	277,132	14.9	0.92	
Single White Females	211,445	16.7	1.04	
White Couples	507,783	12.0	0.75	
All other Non-Hispanic Whites	149,730	20.9	1.30	
Single African American Males	32,035	35.5	2.21	2.39
Single African American Females	46,945	38.6	2.40	2.31
African American Couples	27,479	37.2	2.31	3.10
All other African Americans	18,014	41.5	2.58	1.99
Single Latinos	39,575	15.5	0.97	1.05
Single Latinas	24,633	20.0	1.24	1.20
Hispanic Couples	42,669	21.9	1.36	1.82
All other Hispanics	24,772	20.6	1.28	0.99
Single Males, American Indian, Asian, or Native Hawaian	25,216	4.9	0.30	0.33
Single Females, American Indian, Asian, or Native Hawaian	16,944	7.6	0.47	0.45
Couples, American Indian, Asian, or Native Hawaian	41,346	3.9	0.24	0.33
All other American Indian, Asian, or Native Hawaian	11,529	9.2	0.57	0.44
Primary Applicant Race or Ethnicity Missing	286,153	17.9	1.11	
Primary Applicant Race or Ethnicity Not Applicable	130,496	7.8	0.49	
Total	1,913,896	16.1		

Notes: Race and Ethnicity categories are defined on the basis of the primary applicant; "All other" categories for each race/ethnicity include same-sex joint applicants, applicants paired with co-applicants of unknown race/ethnicity, and applicants paired with co-applicants of a different race. Race and ethnicity codes are used to create mutually exclusive categories: Single white males, for instance, exclude White applicants identifying their race as Hispanic. Database excludes loans purchased by the institution, and 180,018 requests without information on applicant income.

Table 2. Loan Disposition and Reasons for Denials.

		Share of all Applications
Action Taken		
Loan Originated	2,093,914	46.0
Application Approved but Not Accepted	412,334	9.1
Application Denied by Financial Institution	1,346,271	29.6
Application Withdrawn by Applicant	579,415	12.7
File Closed for Incompleteness	49,806	1.1
Preapproval Request Denied by Financial Institution	17,295	0.4
Preapproval Request Approved but Not Accepted (Optional Reporting)	53,886	1.2
Totals	4,552,921	100.0

Note: Database excludes applications purchased by financial institutions.

		Share of all Denials
Primary Reason for Denial (Optional Reporting)		
Debt-to-Income Ratio	154,489	11.5
Employment History	9,901	0.7
Credit History	328,672	24.4
Collateral	120,548	9.0
Insufficient Cash (downpayment, closing costs)	4,287	0.3
Unverifiable Information	12,929	1.0
Credit Application Incomplete	223,031	16.6
Mortgage Insurance Denied	101	0.01
Other	236,960	17.6
No Primary Denial Reason Cited	255,353	19.0
Totals	1,346,271	100.0
Denials with Secondary Reason Cited	171,337	12.7
Denials with Third Reason Cited	80,014	5.9

Table 3. Logistic Regression Model of Bad-Credit Denials.

	Parameter Estimate	Odds Ratio[1]
Intercept	-2.9189	0.05
Applicant Income	-0.00571	0.53
Loan to income ratio	-0.0562	0.88
FHA, VA, or FSA/RHS Loan[2]	0.3065	1.36
Mobile Home	0.8763	2.40
Preapproval request, purchase application[3]	-2.497	0.08
Home improvement application	1.7066	5.51
Refinance application	0.8666	2.38
Hispanic primary applicant[4]	0.3441	1.41
Primary applicant ethnicity not reported	0.168	1.18
Primary applicant race known, ethnicity not applicable	-0.7343	0.48
African American	0.5884	1.80
Asian	-0.3673	0.69
American Indian, Alaska Native, Native Hawaiian or Other Pacific Islander	0.2799	1.32
Primary applicant race unreported	-0.0455 *	0.96
Primary applicant race not applicable but ethnicity not "n/a"	-0.5727	0.56
Primary applicant race and ethnicity both not applicable	-0.8461	0.43
Number of observations[5]	374,683	
-2 Log Likelihood	199,114	
Chi-Square vs. Null Model	17,072	
Percent concordant	70.9	
Nagelkerke (1991) Max-rescaled Pseudo R-Squared	0.102	

Notes:

1. For continuous measures (applicant income, loan-income ratio), figures report the odds ratio resulting from a one-standard deviation increase in the predictor variable.
2. Reference category is conventional application for a single-family property; database includes no multi-family applications.
3. Reference category is home purchase requests for which pre-approval was not requested.
4. Reference category is White, Non-Hispanic primary applicants.
5. To avoid circularity, model was estimated on a random ten-percent sample of all applications, excluding loans purchased by the institution and applications with no information on applicant income.

* Coefficient not significant at $P< 0.05$; all other coefficients are significant at $P< 0.001$.

Table 4. Logistic Regression Models of Rate-Spread Segmentation.

	Model 1 Race-Ethnicity and Gender Only		Model 2 Add Income and Loan Characteristics		Model 3 Add Credit History	
	Parameter Estimate	Odds Ratio[1]	Parameter Estimate	Odds Ratio[1]	Parameter Estimate	Odds Ratio[1]
Intercept	-1.7452	0.18	-0.5535	0.58	-0.8190	0.44
Single White Females[2]	0.1372	1.15	-0.0969	0.91	-0.0955	0.91
White Couples	-0.2466	0.78	-0.0554	0.95	-0.0536	0.95
All other Non-Hispanic Whites	0.4119	1.51	0.5129	1.67	0.5153	1.67
Single African American Males	1.1493	3.16	0.9932	2.70	0.7701	2.16
Single African American Females	1.2825	3.61	0.9410	2.56	0.7066	2.03
African American Couples	1.2227	3.40	1.2117	3.36	0.9884	2.69
All other African Americans	1.4001	4.06	1.4189	4.13	1.1983	3.31
Single Latinos	0.0524	1.05	0.0355	1.04	-0.1009	0.90
Single Latinas	0.3601	1.43	0.1746	1.19	0.0274 *	1.03
Hispanic Couples	0.4709	1.60	0.2943	1.34	0.1496	1.16
All other Hispanics	0.3983	1.49	0.4400	1.55	0.3000	1.35
Single Males, American Indian, Asian, or Native Hawaian	-1.2243	0.29	-0.6309	0.53	-0.5967	0.55
Single Females, American Indian, Asian, or Native Hawaian	-0.7576	0.47	-0.3977	0.67	-0.3702	0.69
Couples, American Indian, Asian, or Native Hawaian	-1.4575	0.23	-0.6098	0.54	-0.5760	0.56
All other American Indian, Asian, or Native Hawaian	-0.5439	0.58	0.0741 *	1.08	0.1021	1.11
Primary Applicant Race or Ethnicity Missing	0.2221	1.25	0.2426	1.28	0.1900	1.21
Primary Applicant Race or Ethnicity Not Applicable	-0.7212	0.49	-0.3725	0.69	-0.2206	0.80
Applicant Income			-0.0229	0.05	-0.0212	0.06
Loan-to-Income Ratio			-0.4955	0.45	-0.4772	0.46
FHA, VA, or FSA/RHS Loan[3]			-4.0284	0.02	-4.1245	0.02
Mobile Home			0.9324	2.54	0.5826	1.79
Preapproval Request, Purchase Application[4]			2.1432	8.53	2.2943	9.92
Home Improvement Application			2.3526	10.51	1.7173	5.57
Refinance Application			1.6521	5.22	1.4377	4.21
Not Owner-Occupied[5]			-0.4386	0.65	-0.4397	0.64
Owner-Occupied not applicable			0.1241 *	1.13	0.1068 *	1.11
Secured by Subordinate Lien[6]			0.4848	1.62	0.4961	1.64
Not Secured by Lien			-16.7085 *	0.00	-16.5255 *	0.00
Credit History Instrument					4.3717	1.25
Number of observations	1,913,896		1,913,896		1,913,896	
-2 Log Likelihood	1,623,901		1,210,923		1,210,267	
Chi-Square vs. Null Model	65,055		478,033		478,689	
Percent concordant	56.5		85.8		85.8	
Nagelkerke (1991) Max-rescaled Pseudo R-Squared	0.057		0.377		0.378	

Notes:

1. For continuous measures (applicant income, loan-income ratio, credit history instrument), figures report the odds ratio resulting from a one-standard deviation increase in the predictor variable.
2. Reference category is single White Non-Hispanic Male applicant.
3. Reference category is conventional application for a single-family property; database includes no multifamily applications.
4. Reference category is home purchase requests for which pre-approval was not requested.
5. Reference category is owner-occupied applications.
6. Reference category is Secured by First Lien.

* Coefficient not significant at P< 0.05; all other coefficients are significant at P< 0.001.

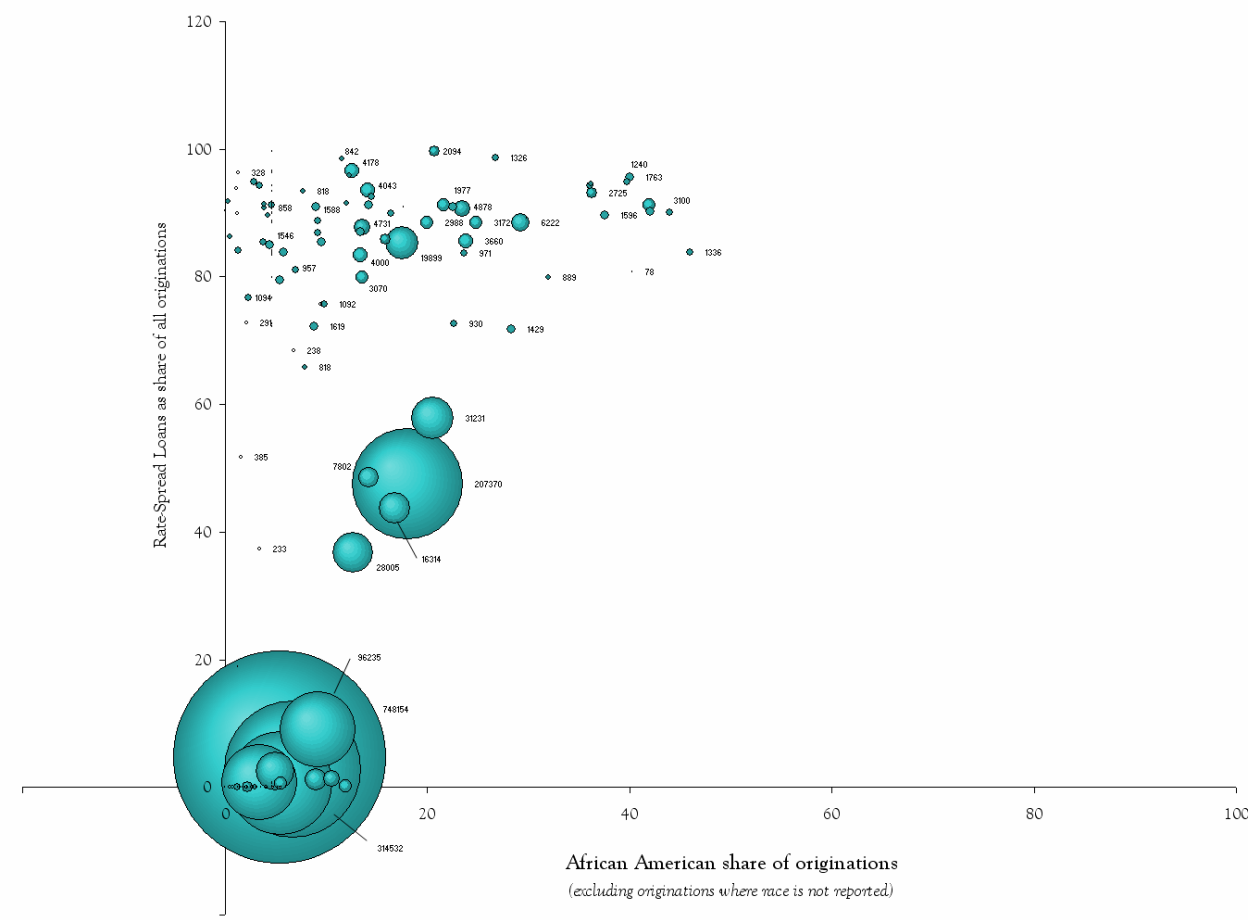

Figure 1. Rate Spread Specialization as a function of African American Share of Lending, Subsidiaries of Four Lenders. Note: African American share is calculated as the percentage of originations with known racial information. Circles are scaled proportional to total originations, but not all subsidiaries are labeled.