

IFN Working Paper No. 811, 2009

# **Reversion to the Racial Mean and Mortgage Discrimination**

Tino Sanandaji



## 1. Introduction

The Home Mortgage Disclosure Act was enacted by Congress in 1975 with the intent of helping regulators enforce fair lending laws and discovering potential discrimination. It was long known that minorities have higher denial rates on mortgages than whites (here and throughout “white” refers to non-Hispanic whites). However, since African-Americans and Hispanics earn less than whites, this fact alone was not deemed sufficient to prove discrimination. Due to a reform of the Home Mortgage Disclosure Act (HMDA) in 1989, detailed data on loan applicant characteristics, including race and income, were available to researchers and journalists for the first time. It was found that high-income minorities were more likely to be turned down for loans than whites with lower income, a fact that led to great controversy when reported in the media (e.g. Seligman, 1991). In 1992 the Federal Reserve Bank of Boston released an influential study (Munnell et al., 1992) on mortgage lending that contained equally incendiary results. The study (henceforth referred to as the Boston Fed study), a later version of which was subsequently published in the American Economic Review (Munnell et al. 1996), demonstrated that minority applicants were denied credit far more often than whites, even after controlling for income, loan size, wealth and other covariates. The Boston Fed study sparked a heated public and academic debate. Proponents of the study argued that it provided irrefutable evidence of discrimination of minorities, which motivated state intervention in the mortgage market. The Boston Fed study was instrumental in changing public policy towards subsidizing minority mortgage lending through public means, since the private sector was seen as inefficient and discriminative. These policy instruments included the activities of Fannie Mae and Freddie Mac, the reinforcement of the Community Reinvestment Act in 1995, and in the federal government private banks to increase minority lending (Liebowitz 2009).

The Boston Fed study was considered groundbreaking since it attempted to correct for the impact of income, and still found differences in rejection rates between racial groups. However, a rational financial institution granting loans should base its assessment not just on the applicant’s income from the most recent year but on the probable future trajectory of the applicant’s income. This paper demonstrates that the relationship between current and future income can be better approximated by a *reversion to the conditional racial mean*, rather than reversion to the mean alone. For this reason, and since races differ in terms of average income, the relationship between single year income and permanent income is different across races. Once this fact is taken into account the difference in denial rates related to race

(or to unobservable variables associated with race) are likely to decrease. Briefly, this article estimates the relationship between 1994 income and average income over 1995-2004 separately for whites, African-Americans and Hispanics. Once the racial mean reverted measure of permanent income is used instead of single year income in a mortgage approval regression, the coefficient for the 'race and ethnicity' dummy (typically interpreted as discrimination) decreases by about one quarter to one fifth in magnitude.

The conclusions of the Boston Fed study were disputed by some economists for a broader set of reasons. The critics focused on the argument that an efficient and profit seeking financial market was unlikely to forgo profit opportunity in this fashion. Instead it was suggested that unobserved characteristics of the applicants other than race could account for the differences observed in the denial rates<sup>1</sup>. Gary Becker dedicated a portion of his Nobel lecture (Becker, 1993) to criticizing policy conclusions from the Boston Fed study, pointing out that "a valid study of discrimination in lending would calculate default rates, late payments, interest rates, and other determinants of the profitability of loans".

If minorities were discriminated against, the marginal white borrowers should be of lower quality and therefore have higher default rate than the marginal African-American borrowers. However, empirical studies have consistently demonstrated that the reverse is true, with African-Americans and Hispanics having higher mortgage default rates than whites (Berkovec et al. [1994, 1998], Anderson and VanderHoff [1999], Cotterman [2002, 2004], Gerardi and Willen [2009]). Han (2002), for example, finds a 4.4% raw default rate for whites, 7% among Hispanics and 7.8% among African-Americans for FHA loans that originated in 1987 and 1988<sup>2</sup>.

These two facts combined present a puzzle. Given the same income, minority applicants are more likely to be rejected, while at the same time minority borrowers are more likely to default. One possible explanation is that unobserved variables other than current-year income explain the higher rejection rates for minority applicants.<sup>3</sup>

---

<sup>1</sup>Critics of the Boston Fed study include Becker (1995), LaCour-Little (1996), Day and Liebowitz (1998), Horne (1997), Harrison (1998), and Stengel and Glennon (1999). Ladd (1998), LaCour-Little (1999) and Ross and Yinger (1999, 2002) contain reviews of the mortgage discrimination literature and of the debate centering on the Boston Fed study. Devaro and Lacker (1995) develop a model and test the empirical implications of errors in variables in the discrimination equation.

<sup>2</sup> Han (2002) also points out that loan performance should be measured using more indicators than default rate alone, and includes other profitability measures relating to the terms of the loan.

<sup>3</sup> Another issue that is not dealt with in the Boston Fed study regards is selection bias and other unobserved variables in deciding to apply for a loan. Borrowers do not randomly apply for loans, and are more likely to

The research following the original release of the 1990 MHDA data (including Munnell et al. [1992, 1996]) largely focused on finding variables other than income which impacted creditworthiness and were associated with race. Such variables include wealth, debt to asset ratio, employment history and credit scores. When including such variables, the coefficient on race for loan rejection was reduced or disappeared (e.g. Schill 1994, Munnell et al. 1996). A shortcoming of this method is that it relies on detailed individual level information that is often hard to come by. Furthermore, it is not clear to what extent the variables in question (such as wealth) have a causal effect on the probability of rejection, or to what extent they act as proxies for creditworthiness. Finally, the fact that the “discrimination coefficient”<sup>4</sup> progressively decreases as new covariates are added raises the possibility that the race variable is significant not because of discrimination, but because race is associated with unobserved determinants of creditworthiness. If, as it is often the case, the discrimination coefficient is reduced by the addition of covariates but remains statistically and economically significant we might suspect that by adding even more controls it would ultimately disappear completely. This hypothesis is hard to test, especially since some important covariates may be unobservable or hard to come by (for example credit scores or job security). Using race specific reversion to the mean to estimate permanent income has the advantage that it does not require the costly collection of covariates. The problem with unobservable characteristics correlated with race is however not solved by this method.

This paper proceeds by discussing the process of racial mean reversion of income. In section 3, empirical evidence of this phenomenon is presented. Section 4 reports the central results of the study; using a measure of expected income estimated through PSID data instead of annual income in the loan application equation, reduces the coefficient for discrimination sizably. Section 5 expands the analysis to other datasets of general mortgage holdings and homeownership, with even starker results. The second part of this paper will attempt to further motivate the first part by demonstrating some policy relevant aspects of mortgage

---

apply when the probability of rejection is low. However we do not know if minority and white applicants are identical in this aspect. For example if minorities are indeed statistically discriminated they might be induced to send out more “marginal” loan applications than whites, and in this way reinforce the higher denial rate. Another possibility is that whites are more often “coached” when making the application (Munnell et al. 1992), and therefore write better applications or don’t apply at all when they are likely to be rejected. Charles and Hurst (2002) estimate mortgage application and minority borrower anticipation of rejection. Here some auxiliary regressions will be run in order to partially account for this issue, but readers are advised to keep the potential problems of selections in mind throughout.

<sup>4</sup> It is common to refer to the coefficient of dummy variables for race or ethnicity, holding other factors constant, as “discrimination coefficient”. For simplicity this convention is followed here, although without making any claims about the causal factor. Differences in outcome could indeed derive from taste based or statistical discrimination, but may also reflect other factors, such as unobservable differences between racial groups.

discrimination. Finally, sections 6 and 7 conclude with a discussion of recent developments in the market for minority mortgages and reports novel estimates of the impact of minority borrowers in the sub-prime market.

## **2. Reversion to the racial mean in income**

More than half a century ago Milton Friedman brought racial differences in permanent and transitory income to attention, writing that “higher average income means that a given measured income corresponds to a higher permanent income for whites than for [African-Americans]” (Friedman, 1957, p. 80). He pointed out that African-Americans with high incomes in a given year appeared to be saving more not because they had a different saving function than whites, but because the mean around which their income fluctuated was lower, so that a higher income contained a larger transitory component. Friedman thus states that “lower savings of whites at each measured income level may reflect simply the inadequacy of measured income as an index of economic status.” (Friedman 1957, p. 80). Since African-Americans earn less than whites, incomes above average in a single year are likely to contain a higher transitory component than a similar observation of a white person *earning the same income*.

This fact has generally been overlooked in the mortgage discrimination literature. One advantage of the explanation put forward in this paper, based on Friedman’s insight, is that it can partially account for the puzzle of higher minority rejection controlling for income and the higher minority default rates, without relying on the use of individual level covariates that are often hard to come by. Furthermore, the framework outlined in this paper can be applied to situations outside of the mortgage market where single year income is used as a control and where outcomes for groups with different permanent income are compared with each other (such categories include race, ethnicity and gender).

Reversion to the racial mean in income is important for the loan approval decision since the income variable typically used in studies of minority rejection rate is a yearly snapshot, whereas lending institutions care about earnings in future years. Individual income can be described as fluctuating around a mean (often with a trend). Since African-Americans and Hispanics earn less on average than whites, at any given income above the mean an African-American is more likely to revert downward to the mean than a white person. As an illustration, the median family income for white-headed households in 2002 and 2004 was

around 63,500 dollars, whereas it was \$34,000 for African-Americans (in 2004 dollars). Using this information, even if nothing else is known about the person, it can be inferred that on average a white person with a family income of \$63,500 is likely to be having a normal year, whereas an African-American earning \$63,500 on average is likely to be having an unusually good year. Therefore, the African-American person in the example is more likely than the white person to experience a reduction of income the following year. Indeed, while only 17% of Whites earning a family income more than \$63,500 dollars in 2002 earned less than that in 2004, 30% of Blacks earning more than \$63,500 earned less than that in 2004. This is partially, but not entirely, due to whites being further from the threshold. For example 62% of Blacks earning between the narrow range of \$60,000-65,000 in 2002 saw a reduction in own income by 2004, compared to 46% of whites earning between \$60,000-65,000.

Moving beyond this example, this paper presents estimations on the relation between single year income and average income in the following ten years (which will be defined as “permanent income”) for whites, Hispanics and African-Americans. It is shown using PSID data that this relationship is different for the different races and, particularly, that minorities with high income are more likely to witness a reduction in their permanent income.

This finding is important, since the political and journalistic reactions following the Boston Fed study were particularly focused on the rejection rates for African-Americans in higher income brackets. The racial reversion to the mean is demonstrated to be particularly important for minorities in these income groups. A striking result is that an African-American earning 75,000 dollars per year in family income in 1994 has the same expected earnings in the next ten years as a white person earning only \$46,000 per year in 1994.

If future ten-year income was the only variable determining lending, a bank without any preferences for racial discrimination would treat African-Americans earning \$75,000 the same as a white borrower with much lower average income. This result is likely to alter the interpretation of a race dummy in a loan application regression that only uses single year income. It should be noted that many commentators would view such characterizations as simply a more sophisticated form of statistical discrimination. It is quite possible that basing lending decisions on such an income model would amount to breaking the law on the part of the bank. In order to decide if taking reversion to the racial mean into account is a form of statistical discrimination or not, we may consider if it is directly important or if it acts as a proxy for unobserved variables (such as creditworthiness). A bank which assumes that

African-Americans with high incomes are likely to see lower income growth and as a result becomes more likely to reject them is guilty of statistical discrimination. A bank that looks at a broader set of variables that correlate with reversion to the mean in income, such as wealth, job security and credit score, might similarly be more likely to deny loans to African-Americans. However in this case if the decision is truly based on individual characteristics other than race (but which correlate with race) the bank may not be guilty of statistical discrimination. In other words, if the mean reverting income generating process is reflected in other covariates of creditworthiness, a bank that ignores race but puts effort into assessing borrower characteristics may nevertheless act *as if* taking race specific mean reversion into account.<sup>5</sup>

Studies of default generally state that higher income is associated with lower default rates (e.g. Cotterman 2004). This is intuitive, but not as obvious as it may sound, as it implies that loans are rationed on the quantity margin, and not through interest rate differentials. Wealth is one of the variables included in typical mortgage approval regressions that constitute a proxy for future income and creditworthiness. Furthermore, wealth is related to permanent income more closely than transitory income, and for this reason (and perhaps others) it differs across minorities even for a given annual income. However wealth has not been proved to be an economically significant predictor of mortgage rejection, as mortgage lenders do not consider individual wealth measures reliable (Munnell et al. 1992).

### **3. An illustration of racial reversion to the mean using 2002-2004 income movements.**

The phenomenon of mean reversion in income is well studied (e.g. Abowd and Card 1989)<sup>6</sup>. This section provides further empirical evidence of the phenomenon of race specific mean reversion. The general idea is that people on average tend to gravitate towards the mean income of their race. Race is of course not unique in this regard; any factor that contains information about variables correlated with earnings will lead to the same phenomenon, such as age, education or sex. Racial reversion to the mean is analyzed since race is the variable that mortgage discrimination studies are preoccupied with.

---

<sup>5</sup> From a policy standpoint it matters whether discrimination is statistical or preferences based. Public policy to ease lending standards to minorities leads to better functioning markets where discrimination is preference based, but may lead to excessive default rates where discrimination is statistical.

<sup>6</sup> Saez et al. (2009) point out the mean reversion can bias estimates of the elasticity of taxable income, as high income earners any given year are more likely to face slower income growth following years, regardless of tax rates.

As an illustration of mean reversion, 84% of whites with incomes above the national median in 2002 have incomes above the median in 2004, compared to 73.4% of blacks and 75.7% of Hispanics. A simple OLS regression also demonstrates racial mean reversion (tables 1, 3 and 4).

The figures are even more striking for the 90<sup>th</sup> percentile. While 64.2% of whites and 60.0% of Hispanics with incomes in the 90<sup>th</sup> percentile in 2002 remained in the 90<sup>th</sup> percentile in 2004, the figure for African-Americans was only 24.9%. Of course, this is partially because African-Americans earn less on average, and African-Americans that are in the highest percentile are on average closer to the cutoff than the corresponding whites and Hispanics. However this is not the whole story. The mean (not dollar weighted) reduction of income from 2002 to 2004 for those in the 90<sup>th</sup> percentile in 2002 was 3.7% for whites, 8.0% for Hispanics and 16.7% for African-Americans as share of 2002 income.

Interestingly, the share of blacks that remain in the *top 90 percentile for black income* for both years is 65.6%, the share of whites that remain in the white top 90 percentile is 62.4%, and the share of Hispanics that remain in the top Hispanic income is 65.9%.

Moving back to the median, the same phenomenon is observed, where within-race income movement behaves similarly, as long as individuals are compared to their racial mean, and not the national mean. 79.1% of Blacks earning above the black median in 2002 were also above the black median 2004, 79.7% of Hispanics in 2004 were above the corresponding Hispanic median income of 2002 and 83.6% of whites above the white median. These results are robust for the years used and for differences between years (implying that there is also income reversion between 1998 and 2002 and so on).

The level of 2002 income where the change in income equals the actual change is \$87,000 for whites (compared to an actual mean income of \$86,100 and median income of \$63,300), \$27,200 for blacks (compared to a mean income of \$42,900, and a median income of \$33,500) and \$46,300 for Hispanics (compared to a mean income of \$49,000 and median income of \$37,400). The simple model of income change thus strongly resembles a process of reversion to racial mean.

It is crucial not to confound race specific reversion to the mean with a general decline in minority income. As seen in table 2 there is no general decline in minority income, if anything Hispanics witness a faster than average increase in income between 2002 and 2004.

The average change in income between 2002 and 2004 was \$1,512 for all (a biannual growth rate of 2%), \$1,515 for whites (1.8%), \$1,327 for African-Americans (3.1%) and \$3,070 for Hispanics (6.3%).

**Table 1**  
**Change in Income between 2002-2004 by Income in 2002\***

Change 2002-2004	Coefficient	Robust Std. Error
Income 2002	-0.198	(0.0456)
Black	-8290	(1752)
Hispanic	-5338	(1705)
Constant	18109	(3419)
R-squared = 0.0501, Number of Observations = 20482		
*Here and throughout variables not statistically significant at 5% level will be reported in italic format.		

**Table 2**  
**Change in Income between 2002-2004 by race**

Change 2002-2004	Coefficient	Robust Std. Error
Black	<i>18</i>	<i>(1328)</i>
Hispanic	<i>1751</i>	<i>(1488)</i>
Constant	<i>1309</i>	<i>(1189)</i>
R-squared = 0.0000, Number of Observations = 20482		

**Table 3**  
**Income in 2004 by race and Income in 2002**

Income 2004	Coefficient	Robust Std. Error
Income 2002	0.802	(0.0456)
Black	-8290	(1752)
Hispanic	-5338	(1705)
Constant	18109	(3419)
R-squared = 0.4738, Number of Observations = 20482		

**Table 4**  
**Log income in 2004 by race and log income in 2002**

Logincome 2004	Coefficient	Robust Std. Error
Logincome 2002	0.680	(0.0307)
Black	-0.230	(0.0327)
Hispanic	-0.062	(0.0255)
Constant	3.493	(0.3397)
R-squared = 0.4877, Number of Observations = 20455		

#### 4. Using PSID race specific income equation with HMDA loan applications

Having demonstrated the principle of race specific reversion to the mean I proceed to use this information to better estimate models of mortgage approval. In order to reduce the transitory proportion of income as much as possible, a ten year average is used.

Reversion to the racial mean does not imply that African-Americans earn less in 1995–2004 than they did 1994. Rather if you did poorly in 1994 you are likely to do better, but if you did unusually well in 1994 you are likely to do worse. *What is central is that “unusually well” is different for blacks and whites.* Knowledge of this fact is crucial for a bank that decides to make a loan looking only at income in 1994. For the large segment earning between \$50,000-80,000 one would have, on average, predicted (and empirically witnessed) a reduction of income for black borrowers but an increase in income for white borrowers.

Tables 5 and 6 illustrate the results already seen for 2002–2004 for the years used to estimate future income, 1994–2004. Note that in Table 6 the income interaction term for African-Americans is not statistically significant. If plotted, the income reversion equation of African-Americans and whites would appear parallel, but with different intercepts.

**Table 5**  
**Average income in 1995-2004 by race and Income in 1994**

Average Income 1995–2004	Coefficient	Robust Std. Error
Income1994	0.606	(0.0500)
Black	-16262	(1610)
Hispanic	-7531	(1893)
Constant	34349	(3394)
R-squared = 0.4221, Number of Observations = 16608		

**Table 6**  
**Average income in 1995-2004 by race, interaction and Income in 1994.**

Average Income 1995-2004	Coefficient	Robust Std. Error
Income1994	0.603	(0.0526)
Black	-15395	(3926)
Hispanic	-15709	(3940)
<i>Black_inc_interaction</i>	-0.026	(0.0673)
<i>Hispanic_inc_interaction</i>	0.123	(0.0602)
Constant	34596	(3586)

R-squared = 0.4226, Number of Observations = 16608

An important point to keep in mind is that the transformation of single year income to future income does not lower minority income on the whole. This is true by construction, since average income 1995–2004 for all races was higher than income observed in 1994. However, income did decline significantly for those who earned above their respective racial mean. Since the family income of potential borrowers in, for example, 2006 is higher than the rest of the population this implies that most borrowers in the sample will have a permanent income lower than their 2006 actual income. A part of the sample, those with low incomes, experience upward reversion to the mean and have predicted permanent incomes higher than their 2006 incomes. In the PSID weighted sample 63.3% of African-Americans had higher 1995–2004 average family income than their 1994 income. The corresponding figure for whites was 59.1%, again reminding us that reversion to the racial mean is a different phenomenon than decline in income.

I use 2006 HMDA data for home loan applications, which covers the overwhelming majority of mortgage loan applications in the United States that year (Avery et al. [2007] discuss issues relating to the use of HMDA data). 2006 is a natural choice as it is the earliest year where data is publicly available. Observations lacking information on income, race or ethnicity have been dropped. Home purchase and refinance loans will be studied, whereas loans for home improvement are dropped. Only loans for one to four family housing are used. Applications that were denied due to incompleteness are coded as denied, whereas applications withdrawn by the applicant are dropped. The results in the paper are robust to these formatting choices. In the HMDA application data loans for homes that are not owner occupied are dropped.

Typically in economics the choice between logit and probit regressions is somewhat arbitrary. In this case however Clarke et al. (2005) show that logit regressions outperform probit in predicting mortgage application outcomes. The difference is small in practice, as is generally the case, with the results being robust to a probit specification. The baseline for the regression is the Midwest, plus a (small) number of applicants whose geographic origin is unknown. Perhaps unsurprisingly, including an interaction term for African-Americans with white co-applicants shows that having a white co-applicant reduces the probability of rejection more for African-Americans than for non-blacks (the marginal effect is roughly a reduction of the rejection probability by 2.5 percentage points). Using an interaction effect for Blacks in the south however does not produce economically significant results. Including these variables does not have any substantial effect on the magnitude of change in Black and Hispanic discrimination from using single year or permanent income. The general results are also robust to the inclusion of a dummy for refinanced loans (which increases the probability of rejection); the difference in the coefficient for African-American between regressions diminished somewhat in size.

The construction of a proxy for “permanent” income is based on comparing 1994 income and the average of 1995–2004 real income.<sup>7</sup> This relation, which is somewhat different for each race, is then used to create a new variable that maps 2006 income into expected mean income in 2007-2016. In order to improve the forecast, real income growth between 1994 and 2006 has been taken into consideration and real income for 2006 is deflated by a factor of approximately 0.86. At a first glance this transformation may not appear completely fitting since income growth rates differ for different socioeconomic groups. But on the other hand the problem is partially mitigated as the borrowers in the sample are mostly higher- or middle-income individuals, and in both groups income growth was close to the average. Furthermore, the size of the loan is deflated by the same factor as income. Therefore the transformation is ultimately justified by the fact that the constructed variable is a proxy for permanent income; it is a variable constructed for use in a regression of discrimination, and not as something whose intrinsic value is of interest.

Tables 7 and 8 summarize the most important results of this paper. The marginal effect of racial dummies falls by roughly one quarter when a measure of permanent income is used rather than 2006 income, holding some covariates fixed. For African-Americans this

---

<sup>7</sup> Income and loan amount data is inflation-adjusted using 2004 dollars as baseline, since this is the base year used in the PSID sample.

represents a decline of the discrimination dummy from 12.8% to 9.6%, and for Hispanics from 6.9% to 5.3%. This should be contrasted with the denial rates without any covariates, where the marginal effect of the African-American discrimination dummy is 14.4%. The impact of using permanent income rather than single year income is thus roughly as important as that of adding all other covariates in Tables 7–10, including loan size and income itself. Tables 9 and 10 present the equations of Tables 7 and 8 in a linear probability model, producing similar results.

**Table 7**  
**Mortgage application denial by race, income in 1994 and controls<sup>8</sup>**

<b>Logistic Regression, 2006 Income</b>		
<u>Denied</u>	Coefficient	Robust Std. Error
Black	0.5799	(0.00190)
Hispanic	0.3242	(0.00174)
Female	-0.0575	(0.00128)
Fsa/Rhs	0.6528	(0.01630)
Va	-1.4508	(0.00912)
Fhainsured	-1.0460	(0.00426)
WhiteCo-applicant	-0.2034	(0.00150)
BlackCo-applicant	0.0633	(0.00340)
OtherCo-applicant	-0.0531	(0.00236)
Logamount	0.0775	(0.00079)
West	-0.1467	(0.00174)
South	-0.1810	(0.00159)
Northeast	-0.0217	(0.00196)
Logincome	-0.3621	(0.00118)
Constant	2.0512	(0.01178)

Pseudo R-squared = 0.0266, Number of Observations = 16046771

<b>Marginal effects of Logistic regression, Average marginal effects on Prob(denied) after logit</b>		
Denied	dy/dx	Std. Error
Black	0.1280	(0.00045)
Hispanic	0.0678	(0.00039)
Female	0.0117	(0.00026)
Fsa/Rhs	-0.1101	(0.00225)
Va	-0.1939	(0.00069)
Fhainsured	-0.1580	(0.00045)
WhiteCo-applicant	-0.0376	(0.00027)
BlackCo-applicant	0.0127	(0.00069)
OtherCo-applicant	-0.0104	(0.00046)
Logamount	0.0143	(0.00015)
West	-0.0273	(0.00032)
South	-0.0340	(0.00029)
Northeast	-0.0042	(0.00038)
Logincome	-0.0669	(0.00021)

<sup>8</sup> A note on reporting: for Table 7 and onwards due to the large sample size and the nature of the problem, all coefficients are statistically significant. See the Data Appendix for explanations of the variables.

**Table 8**  
**Mortgage application denial by race, predicted income 1995–2004 and controls**

<b>Logistic Regression, “Permanent” Income</b>		
<u>Denied</u>	Coefficient	Robust Std. Error
Black	0.4380	(0.00203)
Hispanic	0.2525	(0.00177)
Female	0.0701	(0.00127)
Fsa/Rhs	−0.5813	(0.01632)
Va	−1.4308	(0.00913)
Fhainsured	−1.0136	(0.00426)
WhiteCo-applicant	−0.2318	(0.00150)
BlackCo-applicant	0.0450	(0.00340)
OtherCo-applicant	−0.0722	(0.00236)
Logamount	0.0495	(0.00078)
West	−0.1675	(0.00174)
South	−0.1903	(0.00159)
Northeast	−0.0393	(0.00196)
Logpermanent_inc	−0.4811	(0.00212)
Constant	3.8122	(0.02113)

Pseudo R-squared = 0.0241, Number of Observations = 16046771

<b>Marginal effects of Logistic regression, Average marginal effects on Prob(denied) after logit</b>		
Denied	dy/dx	Std. Error
Black	0.0956	(0.00047)
Hispanic	0.0531	(0.00039)
Female	0.0143	(0.00026)
Fsa/Rhs	−0.1000	(0.00236)
Va	−0.1923	(0.00070)
Fhainsured	−0.1545	(0.00046)
WhiteCo-applicant	−0.0431	(0.00026)
BlackCo-applicant	0.0091	(0.00069)
OtherCo-applicant	−0.0141	(0.00045)
Logamount	0.0092	(0.00014)
West	−0.0313	(0.00031)
South	−0.0359	(0.00029)
Northeast	−0.0077	(0.00038)
Logpermanent_inc	−0.0892	(0.00039)

Just as adding additional covariates to the regression depresses the discrimination dummies, accounting for mean reversion reduces the discrimination dummies. The effect of taking racial mean reversion into account is clearly sufficiently large to motivate its inclusion in this regression. The impact is arguably large enough to motivate the use of permanent income as opposed to annual income in more general settings in the discrimination literature where racial dummies are used and single year income is controlled for (two other applications of this will be presented later).

A bank in possession of a richer dataset could of course take into account education, age, and impression from the borrower, local knowledge about the geographic and other characteristics that are not reported in the HMDA database, which would perhaps reduce the extent of reversion to the mean. More crucially, these data would not appear in a regression using the dataset provided by the HMDA. Therefore regressions using the HMDA data are “blind” to a richer set of variables that can explain reversion to the mean. In an effort to reduce the extent of bias as much as possible, a proxy for permanent income can be introduced, even when a complete model of the causes of mean reversion is missing. In some sense, it is irrelevant what the mechanism driving reversion to the mean is (age, measurement error, random chance, or economic factors) as long as the same or similar mechanisms are at work in the world that mortgage lenders inhabit. For example, measurement error of income in the PSID leads to reversion to the mean, and measurement error of income in mortgage data will have the same effect.<sup>9</sup> The main mechanism is likely to be the effect of transitory income, as already identified by Friedman (1957). It should be noted that the observed patterns are suggestive of forward looking behavior by banks and lending agencies (if patterns of mean reversion in income are taken into account when making lending decisions).

Table 11 presents some descriptive statistics about the HMDA sample. As can be seen, applying permanent income reduces the average income of African-Americans, and raises that of whites. This is a reflection of the income range overrepresented in loan applications (recall that since permanent income is based on 1995–2004 data, and since real income increased in this period, for a full sample of the population average permanent income should not be less than annual income). This reduction of relative income for middle class African-

---

<sup>9</sup> There are some potential mechanisms at work in the HMDA loan application data that are not present in the PSID, such as an incentive to intentionally exaggerate income. This would lead to stronger downward reversion to the mean in HMDA rather than PSID data.

Americans is one of the driving forces behind the reduction of the discrimination coefficient, and is in line with the general arguments of this paper.<sup>10</sup>

**Table 9**  
**Mortgage application denial by race, income 1994 and controls (LPM)**

<b>Linear Probability Model, 2006 Income</b>		
<u>Denied</u>	Coefficient	Robust Std. Error
Black	0.1209	(0.00042)
Hispanic	0.0618	(0.00034)
Female	0.0114	(0.00024)
Fsa/Rhs	-0.1136	(0.00237)
Va	-0.1895	(0.00070)
Fhainsured	-0.1621	(0.00049)
WhiteCo-applicant	-0.0345	(0.00026)
BlackCo-applicant	0.0091	(0.00076)
OtherCo-applicant	-0.0105	(0.00044)
Logamount	0.0144	(0.00014)
West	-0.0277	(0.00032)
South	-0.0336	(0.00030)
Northeast	-0.0048	(0.00037)
Logincome	-0.0664	(0.00021)
Constant	0.8294	(0.00212)

R-squared = 0.0296, Number of Observations = 16046771

<sup>10</sup> Some attempts were made to make use of the extensive information available in the HMDA data. Instead of calculating permanent income nationally by race as done in the PSID, it was calculated by race, by marital status, by race of spouse and by and region. The results of this analysis however are indistinguishable from the method used in Tables 7–10. One possibility may be the small sample size in PSID precludes us from estimating the relation between income and permanent income for subgroups with any precision. Since this exercise adds complexity but does not change the results, the simpler method of estimating permanent income is used. A specification with loan size to income as a single ratio gives similar results, with a smaller decline in the race dummy. This approach tends to generate less explanatory power than allowing both loan size and income to vary.

**Table 10**  
**Mortgage application denial by race, predicted income 1995–2004 and controls (LPM)**

<b>Linear Probability Model, “Permanent” Income</b>		
<u>Denied</u>	Coefficient	Robust Std. Error
Black	0.0965	(0.00044)
Hispanic	0.0498	(0.00035)
Female	0.0138	(0.00024)
Fsa/Rhs	-0.0994	(0.00237)
Va	-0.1851	(0.00070)
Fhainsured	-0.1556	(0.00049)
WhiteCo-applicant	-0.0398	(0.00025)
BlackCo-applicant	0.0054	(0.00076)
OtherCo-applicant	-0.0142	(0.00044)
Logamount	0.0090	(0.00014)
West	-0.0317	(0.00032)
South	-0.0355	(0.00030)
Northeast	-0.0082	(0.00037)
Logpermanent_inc	-0.0856	(0.00036)
Constant	1.1247	(0.00354)

R-squared = 0.0267, Number of Observations = 16046771

**Table 11**  
**Sample characteristics for HMDA 2006 loan application data**

<b>Income (adjusted nominally and downward for trend to reflect 1994 levels)</b>		
	Mean	Median
Single Year Income		
All	74590	57920
White	75526	57116
African-American	61673	49876
Hispanic	73097	61138
Permanent Income		
All	76658	66501
White	80417	80417
African -American	54809	54809
Hispanic	71968	71968
Mortgage Denial Rate		
All	25.6%	
White	22.6%	
African-American	37.2%	
Hispanic	29.7%	

## **5. Mortgage Holdings, Homeownership and Racial Reversion to the Mean in the General Population**

The analysis performed so far in this paper is based on the probability of rejection, conditional on an application having been made. It is implicitly assumed that the application process is independent of the probability of rejection, and that it is the same across all races. Furthermore the relationship between current and future income is based on the 1994–2005 period, whereas the mortgage data is from 2006. In order to test the robustness of the results against these problems, supplementary analysis is performed using the 1995 American Housing Survey. The biannual American Housing Survey is a representative national panel dataset, not of households, but of housing units. While the data cannot be used to evaluate discrimination in lending directly, it represents the best measure of mortgage and housing market outcomes. The dataset is smaller than the HMDA application data, and less suited for analyzing discrimination in loan approval rates as the mortgage approval process is not observed. However it has some advantages; chiefly that it allows for the observation of the ultimate outcomes of mortgage applications, and in this it gets around the selection problem in race differences in applications. Having a mortgage is the dependent variable, with race, income and some covariates as the independent variables. As before single year income will be compared with permanent income. Again it should be emphasized that using permanent income instead of one year income does not change African-American or Hispanic incomes on average. This result is even clearer in the representative American Housing Survey (in contrast to the HMDA mortgage application data where some self-selection based on income has already taken place). Average family income in 1994 dollars<sup>11</sup> for the sample was \$35,200 for African-Americans, whereas average permanent income was \$39,500. For Hispanics the corresponding figures are \$39,300 and \$47,400 respectively.

Using permanent income instead of 1995 income completely removes (even reverses the sign) of the discrimination coefficient of African-Americans, and reduces the discrimination coefficient by more than half for Hispanics.

---

<sup>11</sup> In order to be consistent with the transformation of the HMDA 2006 income data the real 1995 incomes have been adjusted downward slightly to account for real growth in family income 1994-1995 as well as being inflation-adjusted. In both cases this is in line with the PSID based transformation of real 1994 to real 1995–2004 income. The transformation has a limited impact on the results presented here.

**Table 12**  
**Have Mortgage by race, 1995 income and controls<sup>12</sup>**

<b>Marginal Effects of Logistic Regression, 1995 Income</b>		
<u>Average marginal effects on Prob(havemortgage = 1) after logit</u>		
Have Mortgage	dy/dx	Std. Error
Black	-0.1160	(0.00700)
Hispanic	-0.1772	(0.00691)
Female	-0.0162	(0.00575)
Adults	0.0341	(0.00358)
Age	0.0338	(0.00089)
Age <sup>2</sup>	-0.0003	(0.00001)
Married	0.1383	(0.00645)
Children	0.0218	(0.00214)
Logincome	0.1482	(0.00420)

Pseudo R-squared = 0.2717, Number of Observations = 33525

---

<sup>12</sup> Adult refers to number of adults in household; children refer to the number of children.

**Table 13**  
**Have Mortgage by race, predicted 1995–2004 income and controls**

<b>Marginal Effects of Logistic Regression, “Permanent” Income</b>		
<u>Average marginal effects on Prob(havemortgage=1) after logit</u>		
Have Mortgage	dy/dx	Std. Error
Black	0.0522	(0.00907)
Hispanic	−0.0739	(0.00906)
Female	−0.0153	(0.00586)
Adults	0.0306	(0.00356)
Age	0.0314	(0.00082)
Age <sup>2</sup>	−0.0003	(0.00001)
Married	0.1191	(0.00624)
Children	0.0241	(0.00211)
Logpermanent_inc	0.4209	(0.00683)
Pseudo R-squared = 0.2924, Number of Observations = 33525		

Finally the same analysis will be made for homeownership. There is a relatively large literature that attempts to account for lower homeownership rates of minorities (e.g. Colling and Robert 2001). As is the case with the analysis of mortgage holdings, many studies rely on using a combination of racial dummy variables, single year income and other covariates.<sup>13</sup> Hilber and Liu (2008) manage to explain differences between African-American and white homeownership by accounting for covariates such as wealth differences. By simply using the permanent income variable (and without the need for a richer set of covariates such as wealth data) it is shown here that the different patterns of homeownership for African-Americans can be entirely accounted for.<sup>14</sup> This exercise is not meant to be interpreted as a comprehensive study of homeownership, rather it illustrates that applying racial reversion to the mean has broader uses than mortgage applications alone. Arguably, most situations where racial differences are explained by single year income data could be improved by recalling

<sup>13</sup> When analyzing mortgage holdings or homeownership the racial dummies are not as naturally interpreted as discrimination as in the mortgage approval case.

<sup>14</sup> The general patterns of results from regressions of current and permanent income on mortgage holding and home ownership for African-Americans is robust to a restriction of the sample to the top half of the income distribution. In the full sample equations the coefficient for log of permanent income on mortgage holding/homeownership is substantially larger than the coefficient for log of income. This difference is reduced if the regression is restricted to the top half of the income distribution.

Milton Friedman's (1957) insights and including a measure of permanent income or mean reversion.

**Table 14**  
**Have Mortgage by race, 1995 income and controls**

<b>Marginal Effects of Logistic Regression, 1995 Income</b>		
<u>Average marginal effects on Prob(havemortgage=1) after logit</u>		
Ownhome	dy/dx	Std. Error
Black	-0.1436	(0.00679)
Hispanic	-0.2047	(0.00751)
Female	-0.0124	(0.00548)
Adults	0.0305	(0.00316)
Age	0.0251	(0.00069)
Age <sup>2</sup>	-0.0002	(0.00001)
Married	0.1316	(0.00496)
Children	0.0189	(0.00195)
Logincome	0.0997	(0.00293)

Pseudo R-squared = 0.2391, Number of Observations = 44753

**Table 15**  
**Have Mortgage by race, predicted 1995–2004 income and controls**

<b>Marginal Effects of Logistic Regression, "Permanent" Income</b>		
<u>Average marginal effects on Prob(havemortgage = 1) after logit</u>		
Ownhome	dy/dx	Std. Error
Black	0.0202	(0.00780)
Hispanic	-0.0874	(0.00879)
Female	-0.0103	(0.00550)
Adults	0.0253	(0.00314)
Age	0.0233	(0.00067)
Age <sup>2</sup>	-0.0001	(0.00001)
Married	0.1117	(0.00486)
Children	0.0208	(0.00193)
Logpermanent_inc	0.3425	(0.00624)

Pseudo R-squared = 0.2572, Number of Observations = 44753

## **6. The importance of race and ethnicity in the subprime mortgage market**

The financial crisis of 2008 was to a large extent caused by unexpected losses in securities based on subprime mortgages, coupled with low solidity among banks and financial institutions.<sup>15</sup> It has been suggested that the expansion of the subprime market was linked to public policy intending to ease access to credit for minorities (Wallison 2009, Liebowitz 2009). Public involvement in the credit market includes the activities of Fannie Mae and Freddie Mac, the Community Reinvestment Act, and softer measures in encouraging banks to lower lending standards employed by both the Clinton and Bush administrations. Needless to say, the extent of potential mortgage discrimination is a crucial factor in deciding the extent and nature of intervention in the mortgage market. Furthermore, the claim that the US market was plagued by extensive (and irrational) discrimination of minorities constituted the basis of large scale lobbying campaigns by non-profit organizations intent on expanding subprime lending to minorities (Schmidt and Tamman 2009).

This paper makes no claim of demonstrating a direct link between public policy and the subprime crisis. However, using HMDA data, a hitherto unpublished fact about the aggregate size of subprime lending will be presented. Under the Home Mortgage Disclosure Act (HMDA), most mortgage originators are obligated to report basic attributes of mortgage applications to the Federal Financial Institutions Examinations Council. Furthermore, in recent years pricing data on loans is also to be reported for those on which the Annual Percentage Rate (APR) exceeds the yield for Treasury securities of comparable maturity by 3 percentage points for first-lien loans, and 5 percentage points for subordinate-lien loans. As previously noted the HMDA database constitutes the most comprehensive national source of mortgage data, and covers an estimated 80 percent of all home loans nationwide, and an even higher share in metropolitan areas. Thus the data is considered likely to give a representative picture of home lending in the whole of U.S. (Avery et al., 2007). Instead of using definitions such as “subprime” and near prime /Alt A (which are mortgage industry categories used to distinguish the creditworthiness of loan takers) the Federal Reserve Board of Governors opted to use the risk premium as the measure of loan riskiness. Of course the line between prime, near prime and subprime are all amorphous, as is any cutoff point in terms of interest rates. In practice in the HMDA the “higher price” category has generally been considered a proxy for subprime, and the terms will be used interchangeably here.

---

<sup>15</sup> Subprime mortgages refer to mortgage loans made to individuals with low credit ratings.

The idea that lending to minorities could have been at the core of the bubble has been dismissed with the argument that minorities constitute too small a fraction of the aggregate market given their fewer numbers and smaller incomes (e.g. Gross 2008). Since there are no national statistics on subprime lending, no racial breakdown of the subprime market has been available to test this claim. Instead, following Federal Reserve guidelines, the risk premium of loans will be used here as a proxy for subprime loans<sup>16</sup>. It is shown that during the height of the mortgage bubble roughly half of all subprime loans were made to ethnic minorities, an overrepresentation of 217%. Race and ethnicity is thus linked to the subprime bubble more than previously thought<sup>17</sup>. These figures do not by themselves explain the subprime bubble; however the HMDA data gives us a clearer picture of the mortgage market, and is suggestive in terms of guiding research focus. Any model that attempts to explain the subprime bubble would be more credible if it manages to account for the aggregate importance of minorities and their role in the expansion of the market while taking strong per capita minority overrepresentation into account.

---

<sup>16</sup> In this case the proxy is perhaps more economically meaningful than the variable itself. Subprime only refers to the broad categorization of loans by lenders, whereas the price differential gives us a measure of market assessment of risk.

<sup>17</sup> It should be forcefully emphasized that the purpose of this paper is not to put any “blame” on any ethnic group; this is strictly a positive study. Since the topic is sensitive it should also be noted that the author does not believe that anyone except policymakers has any blame for the mortgage meltdown, individuals are expected to take advantage of economic opportunities, including relaxed lending standards. The personal view of the author is furthermore that minority home ownership is a laudable goal, but that it should be pursued through direct means, such as subsidies and human capital formation.

**Table 16**  
**Subprime borrowing by race and ethnicity, 2004–2007**

<b>High cost/ "subprime" mortgages by race and ethnicity</b>	Share of total subprime	<b>Sum subprime, billion dollars, nominal 2004-2007</b>	2004	2005	2006	2007	Share of adult popul.	Share of aggregate national income	Subprime loans taken in relation to aggregate income
<b>Total US*<sup>18</sup></b>		<b>1787</b>	<b>210</b>	<b>511</b>	<b>533</b>	<b>533</b>			
Asian	4.8%	86	8	26	26	26			
Other**	3.3%	59	8	18	17	17			
Hispanic***	26.4%	472	43	135	147	147	13.2%	8.7%	208%
African-American	17.0%	303	36	84	92	92	12.0%	8.1%	301%
All minorities, including Hispanic	<b>49.6%</b>	878	92	253	271	271	31.0%	23.2%	<b>217%</b>
White Non-Hispanic	<b>50.4%</b>	900	118	257	262	262	69.0%	76.8%	<b>66%</b>

In light of the financial crash of 2008, it is easy to portray subprime mortgages in negative terms. However it should be kept in mind that these changes in the mortgage market represent a form of innovation, they are new instruments that improve credit access and allocation. The fact that initial attempts with these innovation failed does not necessarily suggest that future designs cannot succeed (Mulligan, 2008). Gerardi et al. (2007) demonstrate using PSID data that more flexibility in mortgage markets leads to a stronger link between future income and homeownership, and financial innovation in this case was efficiency- and welfare-enhancing.

More generally the housing bubble and subprime crisis can be viewed as results of the credit cycle. One possible way to explain such cycles is that financial institutions tighten credit standards when defaults go up in recessions, but in good times they gradually lose institutional memory and relax lending standards, with the cycle coming to an end when the next period of high defaults occurs. Calomiris (2008) suggests that bubbles occur when high

---

<sup>18</sup> \* Unknown race or ethnicity is assumed to be randomly distributed. In practice unknown race is more common for minority borrowers, for example a disproportionate share of missing data is from high minority census tracts. In practice this means that the data underestimates the share of loans taken by minorities to some degree.

\*\* "Others" include American Indian/Alaska Native, Nat Hawaiian/Other Pacific Islanders, those with two or more minority race, and joint races (White/Minority Race).

\*\*\*Hispanic and "Others" do not add up to 100%, among other since individual with African-Americans or Native Americans race can also have Hispanic ethnicity. Income and adult population shares come from U.S. Census, 2004–2007.

liquidity (and low interest rates) interacts with new innovations, such as tech-stocks, subprime lending to new markets or securitization of new assets. In a similar way it may be the case that crises become particularly severe when the expansive period of a credit cycle interacts with new financial innovations or the opening of new markets. Data presented in this section and later in this paper point to expansion of lending to low-income minorities both in per capita and aggregate terms as one of the main sources of financial expansion preceding the 2008 crisis. One question that would be suitable for future research is to examine to what extent the expansion of minority, and particularly Hispanic, borrowing was caused by general price increases in states heavily populated by Hispanics, and to what extent the bubble was caused by influx of immigrants and higher Hispanic borrowing.

## **7. Changing structures of the minority mortgage market**

In this paper the lending data used was dated 2006 while the relationship between annual income and future income was measured using 1994–2005 income data. Income levels have been scaled down to account for inflation and real growth in family income in these years. Nevertheless it is possible that 2006 is an unrepresentative year, especially since the results are compared to a study of early 1990s (Munnell et al., 1992). Therefore for auxiliary analysis the paper will rely on another publicly available dataset, the American Housing Survey. The results indicate that between the early 1990s and 2007, the structure of the mortgage market for minorities changed, largely as a result of the expansion of high-price (i.e. subprime) mortgages. The change is one of degree, and not of character. Nevertheless the results obtained indicate that in the future it might be valuable to obtain new data and extend the analysis done on 2006 HMDA data to earlier years.

The mortgage market accounts for heterogeneity in borrower creditworthiness through a mix of price differentials (charging higher interest) and quantitative rationing (denying loans)<sup>19</sup>. As a result of the expansion of higher-priced mortgage instruments, lending to minorities increased, while these minorities faced higher interest rates as a result. African-Americans and Hispanics differ from whites in home ownership (i.e. having a mortgage in the first place) and also differ from whites in the average mortgage size given ownership. Their mortgages tend to be smaller or at most the same size, with the exception of Hispanics whose

---

<sup>19</sup> Other differences in the terms of the loan beside the interest rate can also be used to differentiate borrowers.

average mortgage size given ownership started to surpass those of whites during the height of the housing bubble in 2005–2007.

During 1993–2007, the period covered by the data, the relative income of African-Americans and Hispanics compared to whites stayed more or less constant. Meanwhile homeownership and the associated probability of having a mortgage increased dramatically for minorities, especially Hispanics, and especially between the mid 1990s and 2005 which are the years in which the mortgage industry has changed. This increase in mortgage holding and mortgage size was accompanied by a larger spread between the interest rate paid by minorities and whites. As an example, in 1997 minorities on average paid 1.4 percentage points above LIBOR, and whites paid 1.1 percentage points above LIBOR, a 26% difference. By 2007 this differential had risen to 34%. Meanwhile during this period the negative probabilities associated with both owning a mortgage and mortgage size given ownership for minorities decreased.<sup>20</sup>

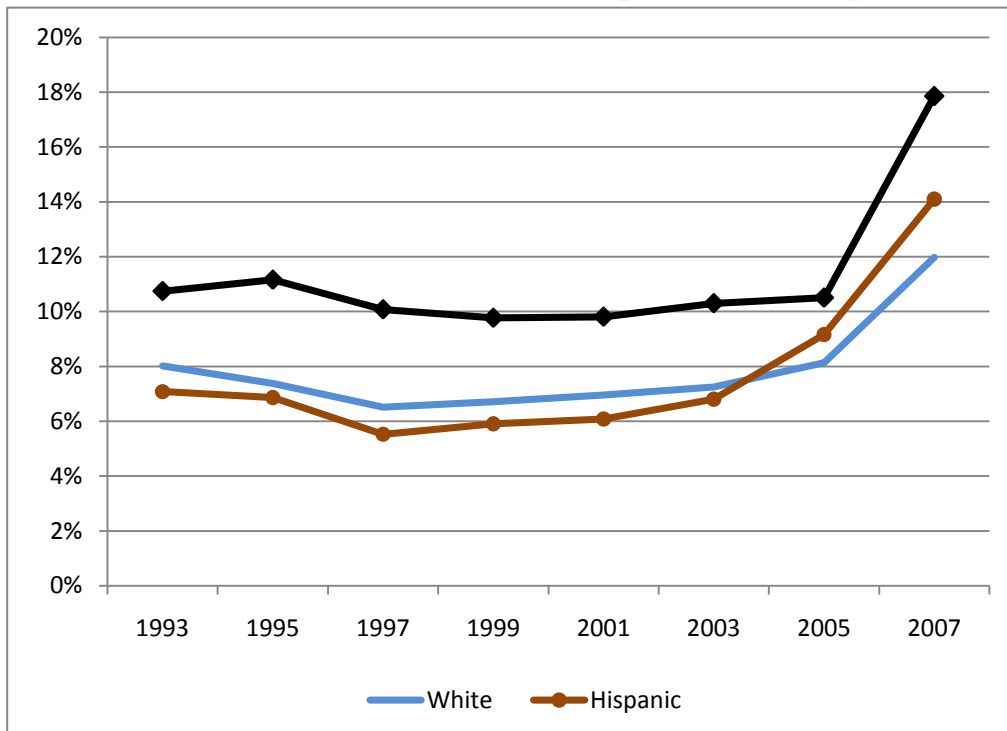
Between 1993–2007 per capita mortgage holdings in per capita terms increased by 128% for whites, 197% African-Americans and 218% Hispanics, far outpacing increases in median or mean family income (Figure 3). During this period mortgage size as a share of family income increased by 24% for whites, 47% for African-Americans and 45% for Hispanics.

Lending standards also became increasingly more permissive over time. For instance in 1997 about 7% of Hispanic households reported purchasing their homes with no down payment, a figure that had doubled to 14% by 2007, now surpassing. Both regarding interest rate differentials and increases in mortgage holdings the effect is particularly strong for Hispanics, although African-Americans follow the general pattern. Note that this development in quantity and price is consistent with a story of reduced rationing, but not consistent with a general convergence of minority economic standards.

---

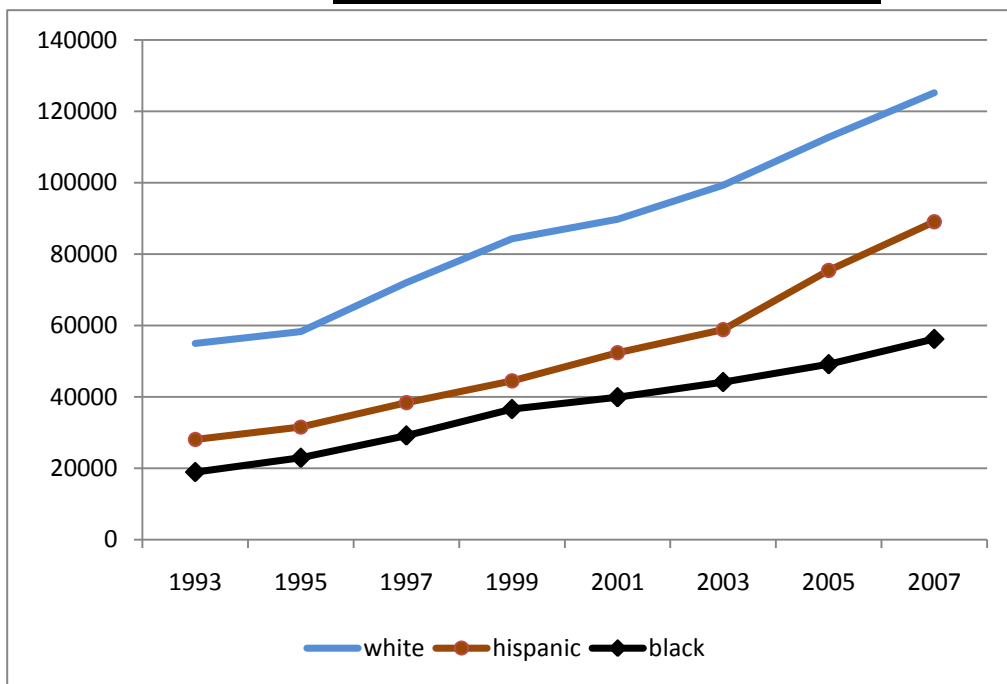
<sup>20</sup> As measured by the race coefficient in a Tobit model of mortgage ownership. This accounts for both the probability of having a mortgage and the mortgage size given ownership. The drawback of the Tobit model is that it relies on strong assumptions regarding the structure of the data

**Figure 1**  
Share of buyers with no down payment on home purchases



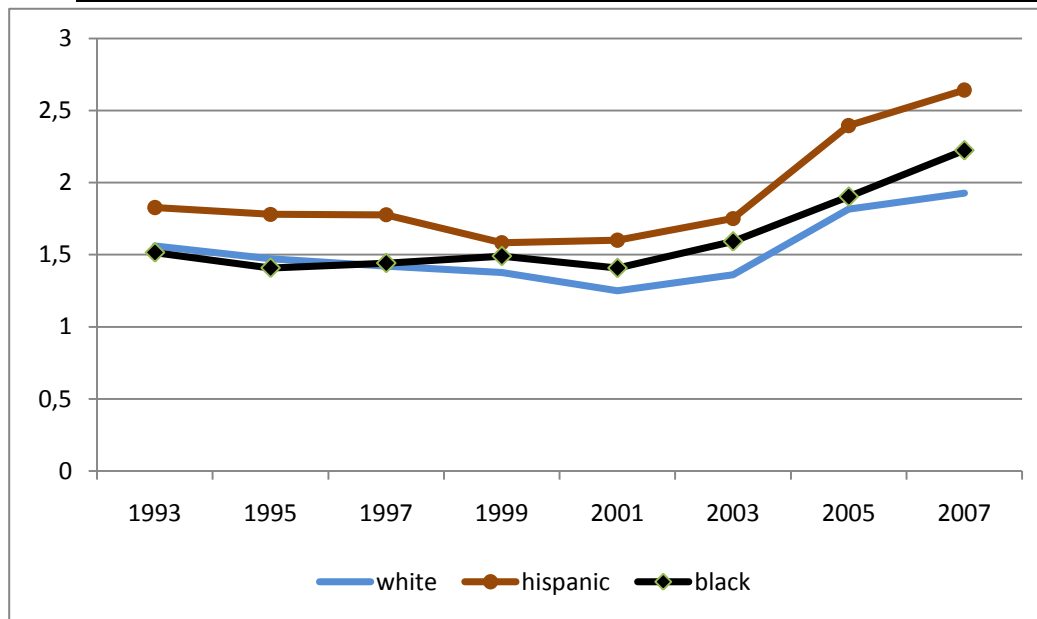
Source: American Housing Survey, 1993-2007.

**Figure 2**  
Per capita mortgage in real 2007 dollars



Source: American Housing Survey, 1993-2007.

**Figure 3**  
**Per capita mortgage size as share of per capita family income for homeowners**



*Source:* American Housing Survey, 1993-2007.

Another manifestation of looser lending standards is credit given to immigrants. Between 2001 (the first year for which data is available) and 2007 per capita real mortgage holdings of immigrants in relation to income increased by 72% for non-Hispanic immigrants and by 83% for Hispanic immigrants. Another indication of ease of credit availability is that between 2001 and 2007 the share of immigrants that had lived less than 10 years in the United States and who held mortgages increased by 2.4 percentage points for non-Hispanic immigrants and 1.2 percentage point for Hispanic immigrants. Meanwhile mortgage size for these two groups increased (for those having a mortgage) by 52% and 57% respectively. These figures can be interpreted both optimistically, as a sign of improved functioning of credit markets, negatively, as excessive lending.

Table 19 uses data from 2001 and 2007, showing that that the negative impact of being an immigrant on the probability of holding a mortgage decreased between 2001 and 2007 by about 4 percentage points. Another suggestive result (Tables 20 and 21) is that the importance of the numbers of years lived in the United States for having a home mortgage declined between 2001 and 2007 for immigrants. In 2001 each additional year in the United States was associated with a 1.09% increase in the probability of having a home mortgage. In 2007 the importance of length of stay was one fifth smaller, with each additional year leading to a 0.89% increase in the probability of having a mortgage. Meanwhile the importance of

income for having a mortgage was higher in 2007 compared to 2001<sup>21</sup>. This result is consistent with the idea that in periods of expansive credit, financial markets are faster to link income to ownership and credit, reducing the requirement on immigrants to build up creditworthiness and ties to the labor market through a longer stay in the country.

**Table 17**  
**Have Mortgage 2001 and 2007 by current income, race, immigration status and controls**

<b>Marginal Effects After Logit 2001 and 2007</b>		
Have Mortgage	dy/dx	Std. Error
Black	-0.1752	(0.00507)
Hispanic	-0.1535	(0.00514)
Immigrant	-0.2454	(0.00604)
Immigrant2007interaction	0.0400	(0.00680)
Female	0.0066	(0.00313)
Adults	0.0858	(0.00178)
Age	0.0075	(0.00044)
Age <sup>2</sup>	-0.0001	(0.00000)
Married	0.1400	(0.00360)
Children	0.0377	(0.00144)
Logpermanent_inc	0.1250	(0.00180)
Pseudo R-squared = 0.1857, Number of Observations = 121903		

<sup>21</sup> Note that immigrant mortgage holdings were higher in 2007 than 2001, and the average mortgage size was larger. The result that the importance of years decreased from 2001 to 2007 is confirmed and statistically significant when taking conditional size of mortgages into account using a pair of Tobit regressions. It is open to discussion whether the difference between 2001 and 2007 is economically significant or not.

**Table 18**  
**Have Mortgage 2001 by current income, race, immigration status and controls**

<b>Marginal Effects After Logit 2001 for Immigrants</b>		
Have Mortgage	dy/dx	Std. Error
Black	-0.04048	(0.02148)
Hispanic	-0.13000	(0.01363)
Female	0.05042	(0.01103)
Adults	0.05251	(0.00558)
Age	0.01054	(0.00200)
Age <sup>2</sup>	-0.00010	(0.00002)
Married	0.05877	(0.01284)
Children	0.04340	(0.00492)
Logincome	0.17846	(0.01554)
Years	0.01086	(0.00059)

Pseudo R-squared = 0.1952, Number of Observations = 9778

**Table 19**  
**Have Mortgage 2007 by current income, race, immigration status and controls**

<b>Marginal Effects After Logit 2007 for Immigrants</b>		
Have Mortgage	dy/dx	Std. Error
Black	-0.11290	(0.01912)
Hispanic	-0.09212	(0.01383)
Female	0.04343	(0.01061)
Adults	0.03299	(0.00577)
Age	0.01133	(0.00198)
Age <sup>2</sup>	-0.00011	(0.00002)
Married	0.06625	(0.01279)
Children	0.04475	(0.00473)
Logincome	0.18647	(0.01716)
Years	0.00894	(0.00054)

Pseudo R-squared = 0.17474, Number of Observations = 10525

## Summary

By relying on Milton Friedman's five decades old insight, it is recognized that given a certain income level, the proportion of transitory income for individuals from different races or ethnic groups is different. When a simple measure of permanent income is used instead of a single year income the coefficient for discrimination in mortgage lending is reduced or disappears completely. Thus, it is likely that previous research, which did not take race specific reversion to the mean in income into account, overestimated the level of discrimination in the mortgage market. An analysis of home ownership, producing similar results, indicates that racial reversion to the mean has applications in areas broader than just mortgage approvals.

Finally, a descriptive overview of the mortgage market over the last decade or so strongly suggests that the role of race and ethnicity is important in understanding recent events such as the subprime mortgage crisis. In light of this it might be suggested that future research look into whether the government and advocacy groups overestimated the extent of racial discrimination in the mortgage market. If this was the case it is possible that both public policy and the direction of the general debate over the last few years led to "overlending", in a mistaken response to perceived underlending.

## References

- Abowd, John and David Card., "On the Covariance Structure of Earnings and Hours Changes." *Econometrica* 57, no.2 (1987): 411–445.
- Anderson, Richard and James VanderHoff. "Mortgage Default Rates and Borrower Race." *Journal of Real Estate Research* 18, no.2 (1999): 279–290.
- Avery, Robert, Kenneth Brevoort, and Glenn Canner. "Opportunities and Issues In Using HMDA Data." *Journal of Real Estate Research* 29, no.4 (2007): 351–379.
- Becker, Gary. "Nobel lecture: The economic way of looking at behavior." *Journal of Political Economy*, 101, no.3 (1993): 385–409.
- Becker, Gary. "How Bad Studies Get Turned Into Bad Policies." *Business Week*, 26 June 1995, page 26.
- Berkovec, James, Glenn Canner, Stuart Gabriel and Timothy Hannan. "Race, Redlining, and Residential Mortgage Loan Performance." *Journal of Real Estate Finance and Economics* 9, no. 3 (1994): 263–294.
- Berkovec, James, Glenn Canner, Stuart Gabriel and Timothy. Hannan. "Discrimination, Competition, and Loan Performance in Fha Mortgage Lending." *The Review of Economics and Statistics* 80, no.2 (1998): 241–250.
- Calomiris, Charles. "The Subprime Turmoil: What's Old, What's New, and What's Next." *The Journal of Structured Finance* 15, no. 1 (2009): 6–52
- Charles, Kerwin and Erik Hurst. "The Transition to Home Ownership and The Black-White Wealth Gap." *The Review of Economics and Statistics* 84, no.2 (2002): 281–297.
- Collins, William and Robert Margo. "Race and Home Ownership: A Century-Long View." *Explorations in Economic History* 38, no.1 (2001): 68–92.
- Clarke, Judith, Marsha Courchane and Nilanjana Roy. "On the Robustness of Racial Discrimination Findings in Mortgage Lending Studies." *Econometrics Working Papers No.0516*, Department of Economics, University of Victoria 2005
- Cotterman, Robert 2002. "New Evidence on the Relationship between Race and Mortgage Default: The Importance of Credit History Data." Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Fair Housing and Housing Finance Publication.
- Cotterman, Robert 2004. "Analysis of FHA Single-Family Default and Loss Rates." Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research, *Economic Development Publications*, No. 39123, March 2004.
- DeVaro, Jed, and Jeffrey Lacker. "Errors in Variables and Lending Discrimination." *Economic Quarterly* 81, no.3 (1998):19–32.
- Friedman, Milton. *A Theory of the Consumption Function*. Princeton, NJ. Princeton University Press, 1957.

- Gerardi, Kristopher, Harvey Rosen and Paul Willen. "Do Households Benefit from Financial Deregulation and Innovation? The Case of the Mortgage Market," NBER Working Papers No.12967, 2007.
- Gerardi, Kristopher and Paul Willen. "Subprime Mortgages, Foreclosures, and Urban Neighborhoods," Federal Reserve Bank of Boston Public Policy Discussion Paper No. 08-6, 2009
- Gross, Daniel. "Subprime Suspects." *Newsweek* October 7, 2008, online edition.
- Hilber, Christian and Yingchun Liu. "Explaining the Black-White Homeownership Gap: The Role of Own Wealth, Parental Externalities and Locational Preferences" *Journal of Housing Economics* 17, no.2 (2008): 152–174.
- Han, Song 2002. "On the Economics of Discrimination in Credit Markets." Finance and Economics Discussion Series of the Board of Governors of the Federal Reserve System, No.2002-2, 2002
- Harrison, Glenn. "Mortgage lending in Boston: A reconsideration of the evidence" *Economic Inquiry* 36, no. 1: 29–38.
- Horne, David. "Mortgage Lending, Race and Model Specification" *Journal of Financial Services Research*, 11, no. 1: 43–68.
- LaCour-Little, Michael. "Application of Reverse Regression to Boston Federal Reserve Data Refutes Claims of Discrimination." *Journal of Real Estate Research* 11, no.1 (1996): 1–12.
- LaCour-Little, Michael. "Discrimination in Mortgage Lending: A Critical Review of the Literature." *Journal of Real Estate Literature* 7, no. 1: 15–49.
- Ladd, Helen. "Evidence on Discrimination in Mortgage Lending" *Journal of Economic Perspectives* 12, no.2 (1998): 41–62.
- Liebowitz, Stan. "Anatomy of a Train Wreck: Causes of the Mortgage Meltdown." In *Housing America: Building out of a Crisis*, eds. Benjamin Powell and Randall Holcomb, Oakland, CA.: The Independent Institute, 2009.
- Mulligan, Casey and Luke Threinen. "Market Responses to the Panic of 2008." NBER Working Papers No.14446, 2008.
- Munnell, Alicia, Lynn Browne, James McEneaney and Geoffrey Tootell. "Mortgage Lending in Boston: Interpreting HMDA Data," Federal Reserve Bank of Boston Working Paper WP-92-7, 1992
- Munnell, Alicia, Geoffrey Tootell, Lynn Brown and James McEneaney. "Mortgage Lending in Boston: Interpreting HMDA Data," *American Economic Review* 86, no.1 (1996): 25–54.
- Ross, Stephen and John Yinger. "Does Discrimination Exist? The Boston Fed Study and its Critics," In *Mortgage Lending Discrimination: A Review of Existing Evidence*, eds. M.A. Turner and F. Skidmore, Washington, D.C.: The Urban Institute, 1999: 42–84.

Ross, Stephen and John Yinger. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*. Cambridge, MA.: MIT Press, 2002.

Saez, Emmanuel, Joel Slemrod and Seth Giertz “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review” NBER Working Papers No.15012, 2008.

Schill, Michael and Susan Wachter 1994. “Borrower and Neighborhood Racial and Income Characteristics and Financial Institution Mortgage Application Screening.” *The Journal of Real Estate Finance and Economics* 9, no.3 (1994): 223–39.

Schmidt, Susan and Maurice Tamman , “Housing Push for Hispanics Spawns Wave of Foreclosures,” *Wall Street Journal sec. A1*, January 9, 2009, morning edition.

Seligman, Dan. “Mortgage Moonshine.” *Fortune Magazine*, December 2, 1991, 183–184

Stengel, Mitchell and Dennis Glennon. “Evaluating statistical models of mortgage lending discrimination: A bank-specific analysis,” *Real Estate Economics* 27, no.2 (1999): 299–334.

Wallison, Peter. “The True Origins of This Financial Crisis.” *The American Spectator* 42, no.1 (2009): 22–27.

## Data Appendix

### Variables in Tables 7-10

**Black** is a dummy variable that takes the value 1 if the main applicant is African-American, and zero otherwise.

**Hispanic** is a dummy variable that takes the value 1 if the main applicant is of Hispanic origin, and zero otherwise. The variable takes the value 0 if the applicant is both of Hispanic origin and Black.

**FSA/RHS** is a dummy variable that takes the value 1 if the loan is guaranteed by the Farm Service Agency or Rural Housing Service and 0 otherwise.

**FHA** is a dummy variable that takes the value 1 if the loan is insured by the Federal Housing Administration and 0 otherwise.

**VA** is a dummy variable that takes the value 1 if the loan is guaranteed by the Veterans Administration and 0 otherwise.

**Female** takes the value 1 if the main applicant is of Female and zero otherwise.

**WhiteCo-applicant** takes the value 1 if the co-applicant is white, and zero otherwise (for most white applicant the co-applicant is also white).

**BlackCo-applicant** takes the value 1 if the co-applicant is African-American, and zero otherwise.

**OtherCo-applicant** takes the value 1 if the co-applicant is neither white nor African-American. The largest group consist of Hispanic and Asian co-applicants.

**Logamount** is the log of the size of the loan applied for.

**West** takes the value 1 if the home for which the application is made is in census region West.

**South** takes the value 1 if the home for which the application is made is in census region South.

**Northeast** takes the value 1 if the home for which the application is made is in census region Northeast.

**Logincome** is the log of the family income of the main applicant plus \$1.

**Logpermanent\_inc** is the log of the constructed family income of the main applicant plus \$1.

### Creation of Permanent income variables

The transformation used to translate 2006 and 1995 income to “permanent income” is first to set those years single income in 2004 real terms, and then compute a constant plus income term for each race. As noted in the paper this comes directly from the PSID. The average transformation (not in fact used) for all races is  $\text{permanent income} = 31301.73 + \text{income} * 0.6160529$ .

For non Hispanic whites it is  $\text{permanent income} = \text{income} * 0.5988562 + 35187.69$ ,

for African-Americans  $\text{permanent income} = 19200.8 + \text{income} * 0.5773632$ ,

for Hispanics =  $18886.81 + \text{income} * 0.7261857$  and

for other races =  $16417.97 + \text{income} * 0.8020087$ .

### **Construction of racial dummies and income variables in the PSID**

The race or Hispanic variable is constructed by using race and Hispanic origin in year 2005, and if missing race and Hispanic origin in 1989 and (again if missing) 1985, 1997, 2001. The years and ordering are chosen with regards to consistency in PSID question format. Hispanic whites have been coded as Hispanic, whereas Hispanic blacks are coded as blacks. Weights used are “CORE/IMMIGRANT FAMILY” from 2005, if missing in 2005 from 1997, and if missing in 1997 then 2001. Race and ethnicity of co-applicant will be used from 1995, to correspond most strongly with income in 1994 (gathered in 1995). For location the state of residence in 1993 will be used, the closest year before 1995 when data is available.

Only observations where at least two income years between 1994 and 2004 exist have been kept. A variable with average income 1995–2004 is used as a proxy for permanent income. Observations with negative average income in these years have been dropped. When one or more years are missing the average variable excludes them and uses the average of the existing years.

The PSID data always use the male as the household head, whereas the applicant data use the main loan applicant as head. Occasionally the main applicant is female, and the co-applicant male. In these cases a dummy for female is included. Where the race (and existence) of a co-applicant is included in the relation between current and future family income a white female main applicant married to an African-American male is assumed to have the same relationship as a black household head married to a white (and so on). Note that the assumption is only made regarding the form of the relationship to future income, the actual family incomes are given by the data.

The co-applicant will in practice be coded as white if race and/or Hispanic origin is unknown. For the main applicant however all observations where income, race, loan amount and ethnicity are unknown is dropped. For the regression that includes geographic data observations the baseline will be the Midwest, plus regions outside the 50 states and D.C. The estimates are robust for different measures of denial rate (loan applications that were denied due to incompleteness are included, and so are withdrawn loan applications).