Household Debt and Defaults from 2000 to 2010:
The Credit Supply View
Online Appendix

Atif Mian
Princeton University and NBER

Amir Sufi
University of Chicago Booth School of Business and NBER

May 2, 2016
1 Individual Level Equifax Data

1.1 Data

Our main data set is based on individual-level credit bureau data from Equifax. This is the same data set used in Mian and Sufi (2011).\(^1\) The initial random sample of individuals was drawn for the year 1997 from a group of 4,025 zip codes with Fiserv Case Shiller Weiss data available. The initial sample contains 320,295 individuals.

We limit the sample to the 288,042 individuals that have credit information available from 1997 to 1999. We make this restriction because there is a large amount of attrition in the initial two years, driven by individuals with very few accounts. The attrition rate is 6% in the first two years, but then is reduced to a constant 2% afterward. We include all individuals that have data available for the first three years of our sample. We have data for these individuals through 2010.

We isolated the sample to people living in zip codes covered by Fiserv Case Shiller Weiss because our original research using the data required zip level house price indices. These 4,025 zip codes contain 25% of U.S. population and 40% of total debt. As has been discussed in our previous research (Mian and Sufi (2009), Mian and Sufi (2011)), the main difference between zip codes in the sample and not in the sample is population density. Only zip codes with a large number of households and a large number of housing transactions generate enough data to construct a zip code level house price index. These are mostly urban areas or suburban areas close to urban areas.

There are two potential sources of sample bias from our data. First, we miss new individuals that first enter the Equifax credit bureau system after 1997. Fortunately, this is likely not a major concern given that new entrants typically do not take on large sums of debt, and they therefore are unlikely to change the conclusions of our analysis focusing on aggregate debt patterns. Second, we only have individuals that resided in zip codes in 1997 with FCSW data. It is harder to assess whether this selection would materially change any results. However, Figure A1 gives us comfort that the FCSW criteria is not a major issue. It shows debt growth according the Federal Reserve Flow of Funds and according to our sample. As Figure A1 shows, debt growth from the flow of funds and from our data match closely.

\(^{1}\)In Mian and Sufi (2011), we were required to sort individuals into groups of five. After the publication of that study, we were granted permission to use the individual level data. The records do not contain any information to identify individuals.
The credit bureau data has excellent information on debt and defaults, but it only has limited information on individual characteristics. The primary measure we have is the Vantage Score, which is a credit score based on creditworthiness of the individual. We will discuss the Vantage Score in more detail below. We also have age for 90% of the sample.

Given the lack of data on income or home values, we supplement the credit bureau data using zip level information on income and home values from other data sources. More specifically, we match each individual to their zip-level average adjusted gross income per tax return and zip-level average home value. Zip level average home value comes from taking the average house price from Zillow in 2000, and then growing the house price by zip-level price indices from CoreLogic.

1.2 Credit scores

As mentioned above, the credit score we have in the credit bureau data is known as the Vantage Score. Like FICO and other credit scores, it is meant to measure the creditworthiness of a borrower. The Vantage Score varies between 550 and 990, as opposed to 300 to 850 for FICO scores. There is no specific cutoff in the Vantage Score data that indicates a subprime borrower, but 700 is a cut-off widely used to indicate a low credit quality borrower. In our random sample of individuals, about 35% of the sample has a Vantage Score below 700 in 1997.

In the analysis below, we split the sample based on the individual’s Vantage Score in 1997. Given that 35% of the sample has a Vantage Score below 700, the subprime cutoff is at the high end of our second quintile. It is important to note that the credit score here applies to borrowers, not to the product used by the borrowers. People with low credit scores may obtain prime mortgages, and people with high credit scores may obtain subprime mortgages. We are not interested in evaluating the rise in debt across different products, but rather across different people.

In all of the analysis below, we group individuals by their initial credit score as of 1997 and track them over time. Why are credit scores the most natural characteristic on which to group individuals? The most obvious reason is practical: it is one of the only individual level characteristics we have in the credit bureau data. However, there are deeper economic justifications for grouping individuals in this manner. Credit scores are one of the primary variables used in credit origination decisions, and individuals with low credit scores have been shown in a number of studies to have high denial rates and a high marginal propensity to borrow. Further, credit scores are designed to
predict default, which is itself an important outcome to evaluate.

We group individuals based on initial credit scores in 1997; an alternative approach would be a dynamic sort, in which individuals are grouped into credit score bins every year. We prefer the static sort for a number of reasons. First, our goal is to evaluate which individuals contributed to the rise in debt and defaults. A dynamic sort would have different individuals in different groups every year, which makes answering our primary question of which individuals drove the rise in debt more difficult.

Second, and more importantly, credit scores become endogenous to credit outcomes and house price growth over time. For example, Mian and Sufi (2011) show that low credit score homeowners in inelastic housing supply cities see a decline in default rates during the housing boom relative to low credit score homeowners in elastic cities. The main reason is that low credit score homeowners seeing high house price growth refinance their way out of defaults – i.e., they do a cash-out refinancing if they cannot make a mortgage payment. As a result, low credit score homeowners in high house price growth areas will see a relative improvement in credit scores that is not driven by fundamental improvements in credit quality, but instead by the housing boom. We discuss this in more detail in Section 6 of the main text.

As another example, an individual that starts in a high credit score group but then sees their credit score drop almost assuredly experienced some kind of financial distress. As a result, we should not be surprised that individuals that enter the low credit score group from a high credit score group see lower debt growth. We want to purge our credit score classification from such endogenous determinants. We use the initial credit score as of 1997, and individuals remain in their group throughout the sample. One disadvantage to this approach is mechanical mean reversion – if low credit score individuals start with low debt, we should expect faster growth. We address this concern in detail in the results below.

1.3 Summary statistics

Table A1 contains summary statistics. Total debt per individual at the beginning of the sample is $49 thousand on average, and mortgage or home equity related debt accounts for $40 thousand. Some of the debt on individual credit reports is held jointly with a spouse, and so debt held by an individual in credit report data is best viewed as something between debt held by an individual
and a household. Table A1 also contains summary statistics on zip-level average adjusted gross income, and zip-level house price growth from 2000 to 2006.

In Table A2, we present summary statistics by credit score quintile. Individuals with lower credit scores have lower debt balances in 1997 and are less likely to have housing-related debt. However, the relation between debt balance and credit score quintile is not monotonic – those in the fourth quintile actually have less debt than those in the third quintile. Lower credit score individuals are younger, and live in zip codes with lower adjusted gross income per tax return. Conditional on having housing debt, the lowest 60% of the credit score distribution have debt to value ratios around 71%. Individuals with lower credit scores live in zip codes in 2000 that subsequently experience higher house price growth, which is consistent with the zip code level evidence presented in Mian and Sufi (2009).

2 Other Data Sources

In order to analyze the household-level American Community Survey (ACS) data, we modify data sets provided by IPUMS-USA. Since the ACS is conducted not on specific people but on specific addresses, the data sets include multiple people living in the surveyed addresses. Moreover, while we are interested in household-level demographics, people residing in group quarters such as institutions rather than in households are also included in the original data sets. Therefore, we first restrict the sample to heads of households.

We then use variables for the race and family income prepared in the data sets to obtain the nominal income and ethnicity of each household. To calculate real income, we use the Personal Consumption Expenditure deflator in 2000 dollar provided by the Bureau of Economic Analysis. Finally, we distinguish recent homebuyers with a mortgage based on indicator variables for the housing tenure, mortgage status, and period since moving-in which are pre-defined in the data sets. When collapsing the final sample to annual data, we weight the sample by the number of households that each observation represents in the total U.S. population.

DataQuick by CoreLogic covers transaction level data. DataQuick does not cover all counties in all years. We focus in the analysis on 2,446 zip codes that have two conditions met: (1) they are in the Mian and Sufi (2009) sample of zip codes for which FCSW data are available, and (2) they
are located in a county for which DataQuick has transaction data from 1998 to 2010. The original Mian and Sufi (2009) sample contains 3,011 zip codes.

3 Homeownership Rate: Various Measures

From 2003 to 2005, the number of total occupied housing units saw a rapid increase, and the increase was driven mostly by households that owned homes rather than renting. As Table A3 reports, nearly 80% of the rise in the total occupied housing units is driven by the increase in owner occupied housing units. Similarly, when comparing the growth rate of total, owner, and renter occupied housing units over the last fifteen years, one can confirm that the growth of total occupied housing units during 2003-2005 coincides with the growth of owner occupied housing units (Figure A2).

However, the homeownership rate from 2003 to 2005 increased by only 0.4% (from 68.6% to 69%, Figure A3). In this section of the appendix, we examine the validity of the homeownership rate as a housing demand index, and we argue that the homeownership rate can potentially be misleading due to intrinsic limitations of its definition. Furthermore, we suggest an alternative measure of housing demand under which the owner occupied housing units are scaled by the adult population.

The first limitation of the homeownership rate, defined as \( \frac{u_o}{u_o + u_r} \), where \( u_o \) is owner occupied housing units and \( u_r \) is renter occupied housing units, is its endogeneity: By definition, the change in owner occupied housing units influences both the numerator and denominator of the formula. As a result, the homeownership rate does not change in proportion to the change in owner occupied housing units, underestimating housing demand.

Second, the change in the homeownership rate is dependent on the current level of owner and renter occupied housing units. That is,

\[
\frac{\partial}{\partial u_0} \left( \frac{u_o}{u_o + u_r} \right) = \frac{u_r}{(u_o + u_r)^2}
\]

(1)

As a result, the greater the owner and renter occupied housing units are, the less susceptible the homeownership rate would be to the change in owner occupied housing units. With an already
high level of total occupied housing units in 2003 (106.5 million), the change in owner occupied housing units during 2003-2005 might have weakly influenced the homeownership rate despite its substantial magnitude (3 million).

Lastly, the homeownership rate is affected not only by the change in owner occupied housing units but also by the change in renter occupied housing units. To examine this issue more closely, let us assume that the owner and renter occupied housing units increase at rates of \( r_o \) and \( r_r \), respectively. Then, the homeownership rate would change by

\[
\frac{u_o(1 + r_o)}{u_o(1 + r_o) + u_r(1 + r_r)} - \frac{u_o}{u_o + u_r} \propto r_o - r_r
\]

implying that no matter how many new households are formed, the homeownership rate would remain unchanged if owner and renter occupied housing units grow at the same rate. Therefore, the sluggish movement of the homeownership rate in 2003-2005 despite a rapid increase in owner occupied housing units (4.14\%) during the same period can be partly attributed to a concurrent increase in renter household units (2.24\%). The relationship in equation (2) is meaningful especially because it suggests that the homeownership rate is not a good indicator to capture the housing demand caused by marginal home buyers and it can possibly be misleading during the period when credit supply leads to housing demand.

On the other hand, our new measure of housing demand, defined as \( \frac{u_o}{\text{adult population}} \), may suffer less from those limitations. In other words, it is not endogenous in that the change in owner occupied housing units affects only the numerator of the formula. Moreover, since the population growth has been stable over decades regardless of housing markets, our new measure varies in proportion to the change in owner occupied housing units. That is, assuming that the population grows at a rate of \( r_p \), our new measure would change by

\[
\frac{u_o(1 + r_o)}{u_o(1 + r_o)} - \frac{u_o}{\text{population}} \propto r_o - r_p
\]

This implies that our alternative measure can be more responsive to the change in owner occupied housing units than the existing homeownership rate by virtue of stable population growth.

However, we admit that our new measure is still far from perfect because population is in
general much greater than owner occupied housing units, making our new measure less sensitive to the change in owner occupied housing units. On top of that, it also has difficulties in interpretation. While the homeownership rate portrays what percentage of households live in their own house, having a value between 0 and 1, our new measure has a meaning only when it is compared to historical data due to the absence of reference point: it is hard to say whether a certain value of our new measure is high or low by itself.
References


Figure A1: Debt Growth: Sample Matches Aggregate

This figure plots the growth in household debt in the Federal Reserve Flow of Funds and growth in total debt for our sample of individual credit reports.
Table A1: Summary Statistics
This table presents summary statistics for our sample of 288,042 individuals. Housing debt to home value is measured only for individuals with some housing-related debt outstanding as of 2000.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>10&lt;sup&gt;th&lt;/sup&gt;</th>
<th>90&lt;sup&gt;th&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total debt, 1997, thousands</td>
<td>288042</td>
<td>49.81</td>
<td>94.24</td>
<td>0.00</td>
<td>148.94</td>
</tr>
<tr>
<td>Housing debt, 1997, thousands</td>
<td>288042</td>
<td>40.43</td>
<td>88.31</td>
<td>0.00</td>
<td>134.00</td>
</tr>
<tr>
<td>Has housing debt, 1997</td>
<td>288042</td>
<td>0.36</td>
<td>0.48</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Credit score (Vantage), 1997</td>
<td>288042</td>
<td>747.99</td>
<td>113.83</td>
<td>586.00</td>
<td>888.00</td>
</tr>
<tr>
<td>Age, 1997</td>
<td>257338</td>
<td>45.58</td>
<td>15.85</td>
<td>26.00</td>
<td>70.00</td>
</tr>
<tr>
<td>Zip average AGI, 1998, thousands</td>
<td>286799</td>
<td>49.80</td>
<td>30.35</td>
<td>26.97</td>
<td>77.18</td>
</tr>
<tr>
<td>Debt to income ratio, 1998</td>
<td>284164</td>
<td>1.13</td>
<td>1.77</td>
<td>0.00</td>
<td>3.39</td>
</tr>
<tr>
<td>Zip average home value, 2000</td>
<td>237358</td>
<td>183.07</td>
<td>106.17</td>
<td>87.70</td>
<td>309.30</td>
</tr>
<tr>
<td>Housing debt to home value, 2000</td>
<td>93995</td>
<td>0.68</td>
<td>0.50</td>
<td>0.16</td>
<td>1.23</td>
</tr>
<tr>
<td>Zip house price growth (%), 2000 to 2006</td>
<td>239707</td>
<td>89.73</td>
<td>50.23</td>
<td>19.79</td>
<td>157.97</td>
</tr>
</tbody>
</table>
Table A2: Averages by 1997 Credit Score Quintile
This table presents averages by 1997 credit score quintile. Each quintile contains 20% of the sample. Housing debt to home value is measured only for individuals with some housing-related debt outstanding as of 2000.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit score (Vantage), 1997</td>
<td>581.0</td>
<td>678.8</td>
<td>756.7</td>
<td>830.5</td>
<td>894.5</td>
</tr>
<tr>
<td>Total debt, 1997, thousands</td>
<td>25.44</td>
<td>42.02</td>
<td>54.04</td>
<td>51.23</td>
<td>76.60</td>
</tr>
<tr>
<td>Housing debt, 1997, thousands</td>
<td>14.49</td>
<td>28.77</td>
<td>43.23</td>
<td>44.22</td>
<td>71.76</td>
</tr>
<tr>
<td>Has housing debt, 1997</td>
<td>0.164</td>
<td>0.299</td>
<td>0.401</td>
<td>0.392</td>
<td>0.544</td>
</tr>
<tr>
<td>Age, 1997</td>
<td>37.23</td>
<td>40.53</td>
<td>44.05</td>
<td>50.23</td>
<td>55.35</td>
</tr>
<tr>
<td>Zip average AGI, 1998, thousands</td>
<td>41.83</td>
<td>45.76</td>
<td>49.98</td>
<td>53.17</td>
<td>58.28</td>
</tr>
<tr>
<td>Debt to income ratio, 1998</td>
<td>0.646</td>
<td>1.152</td>
<td>1.341</td>
<td>1.146</td>
<td>1.355</td>
</tr>
<tr>
<td>Zip average home value, 2000</td>
<td>158.5</td>
<td>171.9</td>
<td>184.3</td>
<td>189.7</td>
<td>210.5</td>
</tr>
<tr>
<td>Housing debt to home value, 2000</td>
<td>0.708</td>
<td>0.732</td>
<td>0.715</td>
<td>0.666</td>
<td>0.626</td>
</tr>
<tr>
<td>Zip house price growth (%), 2000 to 2006</td>
<td>96.91</td>
<td>93.18</td>
<td>88.93</td>
<td>84.78</td>
<td>85.02</td>
</tr>
</tbody>
</table>
Table A3: Occupied housing units and their changes, 2003-2005 ('000s)

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Owner</th>
<th>Renter</th>
<th></th>
<th>Total</th>
<th>Owner</th>
<th>Renter</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>106,505</td>
<td>73,091</td>
<td>33,414</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>110,281</td>
<td>76,119</td>
<td>34,162</td>
<td>3,776 (3.55%)</td>
<td>3,028 (4.14%)</td>
<td>748 (2.24%)</td>
<td></td>
</tr>
</tbody>
</table>
Figure A2: Total, owner and renter occupied housing units
Figure A3: Homeownership Rate