Online Appendix to “Why Do Borrowers Default on Mortgages?”

Peter Ganong and Pascal Noel

Contents

A Figures and Tables 1
B Data Appendix 36
C Econometric Assumptions, Proofs, and Simulations 38
D Empirical Appendix 48
E Stigma Cost of Default 52
A Figures and Tables

A.1 Figures

Figure A-1: Income Change as Share of Initial Income by Home Equity

Notes: This figure replicates Figure 1 using a dependent variable of the change in monthly income divided by the average of the monthly income in months -12, -11, and -10 prior to default.
Figure A-2: Payment Due And Payment Made Prior to Default

(a) Payment Due

Change from $t = -12$

(b) Mortgage Payments by Home Equity

Mortgage payment change from $t = -12$ (as share of monthly payment due)

Notes: The top panel shows mortgage payment due, average income, and mortgage payment made in the year prior to default in the JPMCI data. The bottom panel shows mortgage payment made as a share of payment due in the year prior to default in the JPMCI data.
Figure A-3: Distribution of Income Change Prior to Mortgage Default

Notes: This figure shows the cumulative distribution function for the change in income, divided by average initial payment due. Average initial payment due is computed one year prior to mortgage default and is computed separately for underwater and above water borrowers. This figure provides an alternative visualization of the histogram in Figure 2, and includes the change for all underwater borrowers. The distribution of the change in income is truncated at -8 and 8 to improve readability.

Figure A-4: Change in Percentiles of Income Prior to Mortgage Default by Home Equity

Notes: This figure shows the change in different percentiles of the income distribution in the year prior to mortgage default. The percentiles are calculated separately for above water and underwater borrowers. See Section 4.3 for details.
Figure A-5: Income in Year Prior to Mortgage Default by Mortgage Type and Home equity

(a) Fixed Rate Mortgages

(b) Adjustable Rate Mortgages

Notes: This figure replicates Figure 1 separately for fixed rate and adjustable rate mortgages.
Figure A-6: Income Prior to Default in Non-Recourse States

Notes: This figure replicates Figure 1 from the JPMCI data for the subset of states that do not allow mortgage lenders to sue to recover non-mortgage assets. We use the classification of non-recourse states from Ghent and Kudlyak (2011).
Figure A-7: Evolution of Income by Home Equity Prior to Foreclosure

Notes: This figure replicates Figure 1 defining the date of default as the date of foreclosure initiation.
Figure A-8: Income by Alternative Missed Payment Thresholds and Home Equity

(a) One Month Past Due
(b) Two Months Past Due
(c) Four Months Past Due
(d) Five Months Past Due

Notes: This figure replicates Figure 1 for alternative months past due thresholds.
Figure A-9: Distribution of Defaulters by LTV

Notes: This figure shows the distribution of defaulters in the Chase analysis sample by LTV.
Figure A-10: Estimate of Measurement Error in Observed Loan-to-Value Ratio

(a) All Sales  
(b) Foreclosure Sales  

(c) All Sales  
(d) Foreclosure Sales  

Source: Corelogic Home Price Indexes and Deed data.  
Notes: This figure provides supporting analysis to the adjustment of $\hat{\alpha}_{\text{life event}}$ for measurement error in observed LTV described in Section 4.3 and shown in Table 6a. Our method for constructing the error in observed home prices largely follows Giacoletti (2021). See Appendix B.3 for details.  
The top panels compare the true distribution of home price errors to a Cauchy distribution. The true distribution of errors is shown in gray bars. Home sale price errors are $\frac{\text{PriorSalesPrice} \times \Delta \text{HomePriceIndex}}{\text{ActualSalesPrice}} - 1$. The teal line approximates the non-parametric estimate using a Cauchy distribution, which is truncated from below at -100 percent. We estimate the location and scale parameters of this distribution by minimizing the squared distance between the actual median and interquartile range and the simulated median and interquartile range.  
In the bottom panels, we use this parametric distribution to compute the probability that a borrower is actually underwater for a range of observed LTV values, again in teal. In the text, we refer to this function as $P(G^* = 1 | LTV)$.  

PriorSalesPrice x ΔHomePriceIndex ActualSalesPrice
Figure A-11: Subsamples with (Relatively) More Strategic Default

(a) Income Before Default (Consecutive Missed Payments)

(b) Income Before Default (Subprime Borrowers)

Notes: This figure replicates Figure 1 from the JPMCI data for the subset of borrowers who miss three consecutive payments and the subset of borrowers with subprime loans. Borrowers who miss three consecutive payments are 58 percent of underwater defaults and 44 percent of above water defaults.
Figure A-12: Share of Mortgage Defaults with Consecutive Missed Payments

Notes: This figure extends the analysis in Keys et al. (2013) using the CRISM data. That paper measures the share of mortgage defaults that transition straight from 60 days past due to 180 days past due in four months, while remaining otherwise current on all non-HELOC revolving debt. We refer to such defaults as “straight and otherwise current”. The average share of defaults that meet this definition is 15.9 percent of defaults for underwater borrowers and 9.7 percent of defaults for above water borrowers. Thus, the excess share of straight and otherwise current defaults for underwater borrowers is 6.2 percent.
Figure A-13: Cyclicality Instrument First Stage

Notes: This figure shows a binned scatter plot of the first stage relationship between the cyclicality instrument and LTV in the Chase sample. This corresponds to equation (14). Both the instrument and LTV are residualized against all fixed effects and controls.
Figure A-14: Alternative Measures of Strategic Default

(a) Share of Defaults

Notes: This figure compares measures of mortgage affordability by home equity in the Panel Study of Income Dynamics (PSID) and the bank account data. Gerardi et al. (2018) measures mortgage affordability using income $y$, mortgage payment $m$, and non-housing consumption $c$. That paper classifies a borrower as can-pay if she can afford the mortgage without cutting consumption ($y - m - c \geq 0$) and as subsist-and-pay if she can afford a subsistence consumption level and pay her mortgage ($y - m - c_{subsistence} \geq 0$). See Section 5 for details on these definitions. Panel (a) reports the share of defaults that are classified as strategic using the can-pay and subsist-and-pay criteria. We also replicate the can-pay criteria in the Chase data using current income minus lagged expenses (both housing and non-housing). Panel (b) reproduces the PSID analysis from panel (a), classifying defaults by whether the borrower’s LTV is above 90, which is the LTV cutoff used in Gerardi et al. (2018).
Figure A-15: Alternative Measures of Strategic Default – Distributions

(a) Available Resources Using Subsistence Measure

(b) Available Resources Using Loan-to-value (LTV) Cutoff of 90

Notes: This figure reports two robustness checks on the PSID data in Figure 6a, which uses $y - m - c_{predefault}$ as the x-variable and constructs home equity groups using an LTV cutoff of 100. Panel (a) uses an alternative x-variable $y - m - c_{subsistence}$, where $c_{subsistence}$ is a measure of the expenditure required to achieve a subsistence level of spending on non-housing consumption goods. Panel (b) uses an alternative LTV cutoff of 90, which is the cutoff used for PSID data in Gerardi et al. (2018). See Section 5 for details.
Figure A-16: Campbell and Cocco (2015) Structural Model Replication

Notes: This figure shows that we can replicate the summary statistics in Table 2 of Campbell and Cocco (2015).
Figure A-17: Prevalence of Above Water Mortgage Default

Notes: This figure shows the distribution of the loan-to-value (LTV) ratio at default in the Credit Risk Insight Servicing McDash (CRISM) data. Default is defined as three missed payments.
Notes: This plot shows conditional means for vingtiles of the borrower cost of default ($\varepsilon$) under two different simulation scenarios. The left panel shows the probability of a life event, the middle panel shows the probability of default in the full simulation model, and the left panel shows the probability of default in the absence of life events. See Appendix C.3 for details.
Figure A-19: Cause of Default by DGP and Set of Included Defaults

Notes: This figure shows the fraction of defaults in three groups: obviously strategic ($Y(0,1) = 1$), coincide with life event but not caused by life event ($Y(1,1) = 1, Y(0,1) = 1$) and coincide with & caused by life event ($Y(1,1) = 1, Y(0,1) = 0$) for the simulation described in Appendix C.3.

Figure A-20: Cause of Default by DGP and (Unobserved) Cost of Default

Notes: This figure disaggregates the types of default among the “all defaults” panel of Figure A-19 by the borrower cost of default $\varepsilon$. See Appendix C.3 for details.
Figure A-21: Lagged Default Status by Scenario and (Unobserved) Cost of Default

Notes: This figure disaggregates lagged default status among the “all defaults” panel of Figure A-19 by the borrower cost of default $\varepsilon$. See Appendix C.3 for details.

Figure A-22: Bias in Estimated Share of Defaults Caused by Life Events ($\alpha$) Scenarios

Notes: This figure shows the bias ($\hat{\alpha} - \alpha$) when studying all defaults and when narrowing the sample to just the transition to default. See Appendix C.3 for details.
Figure A-23: Bias-Variance Trade-Off

Notes: This figure shows the bias-variance trade-off between the standard “back-of-the-envelope” method and the reverse regression (or “Bayes”) method for causal attribution that we use in the paper. The figure reports estimates of \( \alpha \) and a 95 percent confidence interval using both approaches within the context of a simulation. In the simulation the true \( \alpha \) is 60 percent, denoted by the dashed horizontal line. The left panel shows the baseline case, which has minimal measurement error in treatment. In this scenario, both the “back-of-the-envelope” and “Bayes” methods are unbiased and precise. The right panel shows the noisy case, where treatment is measured with substantial error. In this scenario, attenuation bias causes the estimate from the “back-of-the-envelope” method to be biased towards zero (but still precise), whereas the “Bayes” method is unbiased (but less precise). See Appendix C.4 for details.
Figure A-24: Income After Three Missed Payments by Subsequent Payment Behavior

Notes: This figure analyzes the evolution of income both prior to and after default (defined as three missed payments). Define $\Delta Y_t = 1$ as a deterioration in delinquency status and $\Delta Y_t = 0$ as no deterioration or an improvement. With $t$ indexing the date of default, the top-left panel shows borrowers with $\Delta Y_{t+1} = 0$, the bottom-left panel shows borrowers with $\Delta Y_{t+1} = 1$, the top-middle panel shows borrowers with $\Delta Y_{t+1} = 0, \Delta Y_{t+2} = 0$, the bottom-middle panel shows borrowers with $\Delta Y_{t+1} = 1, \Delta Y_{t+2} = 1$, the top-right panel shows borrowers with $\Delta Y_{t+s} = 0, s \in \{1 \ldots 6\}$, and the bottom-right panel shows borrowers with $\Delta Y_{t+s} = 1, s \in \{1 \ldots 6\}$. 

Income change (as share of monthly payment due)

-12 -6 Default 6 -12 -6 Default 6

-100% -50% 0% -100% -50% 0%

Above water  Underwater
Figure A-25: Income After One or Two Missed Payments by Subsequent Payment Behavior

(a) After Two Payments

(b) After One Payment

Notes: This figure replicates Figure A-24 for alternative missed payments thresholds.
Figure A-26: Income After Default: Pooling Across Payment Histories

(a) Miss One Pmt
(b) Miss Two Pmts
(c) Miss Three Pmts

Notes: This figure shows the evolution of income both prior to and after default by alternative definitions of default, pooling across all payment paths subsequent to the default date.
Figure A-27: Income Around Mortgage Modification

(a) Above vs Underwater

Income change (as share of monthly payment due)

(b) By Date of Default

Notes: This figure shows the dynamics of bank account income around mortgage modification. The top panel separates borrowers by home equity. The bottom panel reports additional heterogeneity by date of default relative to modification, grouping borrowers into three equally-sized bins. The bottom panel omits borrowers who do not default in the 18 months prior to modification (this group accounts for 3% of the sample). Monthly payment due is measured 18 months prior to modification.
### A.2 Tables

Table A-1: Income Drop From Unemployment by Home Equity

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Change in income (as share of mortgage payment due)</td>
</tr>
<tr>
<td>Post UI receipt</td>
<td>−0.251 (0.014)</td>
</tr>
<tr>
<td>Post UI receipt * underwater</td>
<td>0.026 (0.026)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>394,374</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table reports a regression of the income drop after unemployment by home equity. The regression is analogous to equation (9). We compare the income change in the three months after the start of unemployment (measured by the receipt of unemployment insurance (UI) benefits, as in Ganong and Noel 2019) to the income in a three-month pre-period one year before the start of unemployment (i.e. $t = \{-12, -11, -10\}$). As in equation (9), the income change is normalized by the mortgage payment due. Standard errors are clustered by mortgage.
Table A-2: Loan Characteristics versus Benchmarks

(a) Origination

<table>
<thead>
<tr>
<th>Sample</th>
<th>Benchmark</th>
<th>JPMCI</th>
<th>CRISM</th>
<th>McDash</th>
<th>MBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>All mortgages</td>
<td>Share investor</td>
<td>6.8%</td>
<td>4.0%</td>
<td>5.6%</td>
<td></td>
</tr>
<tr>
<td>All mortgages</td>
<td>Share primary occupant</td>
<td>89%</td>
<td>93%</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>All mortgages</td>
<td>Share subprime</td>
<td>4.8%</td>
<td>3.0%</td>
<td>3.0%</td>
<td>8.6%</td>
</tr>
<tr>
<td>All mortgages</td>
<td>Origination year (25th percentile)</td>
<td>2003</td>
<td>2004</td>
<td>2004</td>
<td></td>
</tr>
<tr>
<td>All mortgages</td>
<td>Origination year (50th percentile)</td>
<td>2006</td>
<td>2007</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>All mortgages</td>
<td>Origination year (75th percentile)</td>
<td>2009</td>
<td>2009</td>
<td>2009</td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Share investor</td>
<td>6.4%</td>
<td>4.3%</td>
<td>5.8%</td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Share primary occupant</td>
<td>90%</td>
<td>94%</td>
<td>92%</td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Share subprime</td>
<td>21%</td>
<td>16%</td>
<td>14%</td>
<td>30%</td>
</tr>
<tr>
<td>Defaulters</td>
<td>Origination year (25th percentile)</td>
<td>2005</td>
<td>2005</td>
<td>2005</td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Origination year (50th percentile)</td>
<td>2006</td>
<td>2006</td>
<td>2006</td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Origination year (75th percentile)</td>
<td>2007</td>
<td>2007</td>
<td>2007</td>
<td></td>
</tr>
</tbody>
</table>

(b) Performance

<table>
<thead>
<tr>
<th>Sample</th>
<th>Benchmark</th>
<th>JPMCI</th>
<th>CRISM</th>
<th>McDash</th>
<th>MBA</th>
</tr>
</thead>
<tbody>
<tr>
<td>All mortgages</td>
<td>90 day delinquency rate</td>
<td>3.2%</td>
<td>3.3%</td>
<td>3.8%</td>
<td>3.6%</td>
</tr>
<tr>
<td>All mortgages</td>
<td>90 day delinquency rate on subprime loans</td>
<td>13%</td>
<td>18%</td>
<td>18%</td>
<td>13%</td>
</tr>
<tr>
<td>All mortgages</td>
<td>90 day delinquency rate on non-subprime loans</td>
<td>2.6%</td>
<td>2.8%</td>
<td>3.4%</td>
<td>2.8%</td>
</tr>
<tr>
<td>All mortgages</td>
<td>Share underwater</td>
<td>19%</td>
<td>22%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Share underwater</td>
<td>50%</td>
<td>58%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Share with foreclosure within year (above water)</td>
<td>40%</td>
<td>50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defaulters</td>
<td>Share with foreclosure within year (underwater)</td>
<td>45%</td>
<td>51%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table compares summary statistics regarding the matched mortgage-bank account dataset from Chase to three datasets: Credit Risk Insight Servicing McDash (CRISM), McDash, and the Mortgage Bankers’ Association (MBA) National Delinquency Survey in 2011. The CRISM dataset is constructed by linking credit bureau records from Equifax with mortgage servicing records from McDash. The MBA dataset covers a broader set of loans (roughly 85-88 percent of the residential mortgage market), but has fewer fields. Positive and negative equity status is only observed in the linked CRISM dataset and not in McDash because it requires total mortgage debt calculated from the credit bureau data. We use 2011 as the comparison year because this is the year when U.S. house prices reached their nadir. Investor and primary occupant are reported by borrowers at mortgage origination. “Foreclosure” indicates that the mortgage servicer initiated foreclosure proceedings.
Table A-3: Distribution of Age versus PSID Benchmark

<table>
<thead>
<tr>
<th>Percentile</th>
<th>JPMCI</th>
<th>PSID</th>
</tr>
</thead>
<tbody>
<tr>
<td>25th</td>
<td>44</td>
<td>38</td>
</tr>
<tr>
<td>50th</td>
<td>52</td>
<td>47</td>
</tr>
<tr>
<td>75th</td>
<td>61</td>
<td>56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentile</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td>25th</td>
<td>40</td>
<td>38</td>
<td>38</td>
</tr>
<tr>
<td>50th</td>
<td>48</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>75th</td>
<td>56</td>
<td>53</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table compares the distribution of age for mortgage borrowers as of 2011 in Chase and in the Panel Study of Income Dynamics (PSID).

Table A-4: Income Change for All Underwater Borrowers

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Change in income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post</td>
<td>−0.009 (0.003)</td>
</tr>
</tbody>
</table>

| N mortgages | 1,891,046 |
| Observations| 11,346,276 |

Notes: This table estimates the average income change for all underwater borrowers in months \( t = \{-2, -1, 0\} \) from Figure 1. Section 4.1 provides details on how this series is constructed. The dependent variable is the ratio of monthly income to average monthly payment due in the pre-period (months \( t = \{-12, -11, -10\} \)). The regression specification is \( \frac{Income}{Payment_{pre}} = \lambda + \phi_1(t = -2, -1, 0) + \varepsilon \). The table reports estimates for \( \hat{\phi} \). Standard errors are clustered by mortgage.
Table A-5: Income Drop at Default – Straight Default and Foreclosure

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Change in income from one year before default</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Straight default</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Date of default</td>
<td>−1.300</td>
<td>−0.819</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Date of default * underwater</td>
<td>0.166</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td>N mortgages</td>
<td>69,343</td>
<td>96,844</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>416,058</td>
<td>581,064</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table re-estimates Column (1) of Table 5 focusing on the subsample that misses three straight mortgage payments, or using foreclosure as the alternative definition of default.
Table A-6: Share of Defaults Causally Attributable to Life Events (\(\hat{\alpha}_\text{life event}\))
Using Alternative Missed Payments Cutoffs

<table>
<thead>
<tr>
<th>Months past due</th>
<th>(\hat{\alpha}_\text{life event}) (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.109 (0.024)</td>
</tr>
<tr>
<td>2</td>
<td>0.900 (0.013)</td>
</tr>
<tr>
<td>4</td>
<td>0.959 (0.016)</td>
</tr>
<tr>
<td>5</td>
<td>1.004 (0.019)</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of \(\hat{\alpha}_\text{life event}\), which is the share of defaults causally attributable to life events, in the JPMCI data using alternative months past due cutoffs. \(\hat{\alpha}_\text{life event}\) is constructed using equation (7), adjusting the pre-period and default period to match the number of missed payments (i.e. pre-period as \(t = -12\) and default period as \(t = 0\) for one missed payment, pre-period as \(t = \{-12, -11\}\) and default period as \(t = \{-1, 0\}\) for two missed payments, etc).
Table A-7: Impact of Negative Equity on Default – Retiree Subsample

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underwater</td>
<td>0.122</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
<td>(0.087)</td>
</tr>
<tr>
<td>LTV fitted residuals</td>
<td>2.378</td>
<td>2.183</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.189)</td>
</tr>
<tr>
<td>$\hat{\alpha}_{negative\ equity}$</td>
<td>0.115</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>(0.078)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Region-Year FEs</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>CBSA FEs</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Borrower and loan characteristics</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CBSA controls</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Origination year FEs</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Instrument</td>
<td></td>
<td></td>
</tr>
<tr>
<td>First stage partial F-Stat</td>
<td>57.23</td>
<td>16.02</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-10,564</td>
<td>-10,514</td>
</tr>
<tr>
<td>Observations</td>
<td>114,535</td>
<td>114,535</td>
</tr>
</tbody>
</table>

Notes: This table replicates the default hazard model estimates from columns (3) and (4) of Table 7a for the subsample of retired borrowers, defined as those age 62 or older who receive a Social Security payment in at least three months of the calendar year.
### Table A-8: Correlation of Instrument with Home Equity and Income

**(a) All Borrowers**

<table>
<thead>
<tr>
<th></th>
<th>Loan-to-Value</th>
<th>Current Income</th>
<th>Future Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>-0.782***</td>
<td>411.221</td>
<td>-77.530</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(338.178)</td>
<td>(323.516)</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>0.67</td>
<td>5944.95</td>
<td>6075.25</td>
</tr>
<tr>
<td>Effect of 1σ increase in instrument (%) of dependent variable mean</td>
<td>-8.13%</td>
<td>0.49%</td>
<td>-0.09%</td>
</tr>
<tr>
<td>Observations</td>
<td>13,477,225</td>
<td>13,477,225</td>
<td>9,099,746</td>
</tr>
</tbody>
</table>

**(b) Retirees**

<table>
<thead>
<tr>
<th></th>
<th>Loan-to-Value</th>
<th>Current Income</th>
<th>Future Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>-0.530***</td>
<td>-857.533</td>
<td>-828.956</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(710.918)</td>
<td>(744.853)</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>0.55</td>
<td>5818.46</td>
<td>5787.8</td>
</tr>
<tr>
<td>Effect of 1σ increase in instrument (%) of dependent variable mean</td>
<td>-7.04%</td>
<td>-1.08%</td>
<td>-1.05%</td>
</tr>
<tr>
<td>Observations</td>
<td>1,146,733</td>
<td>1,146,733</td>
<td>778,499</td>
</tr>
</tbody>
</table>

Notes: this table reports estimates of equation (14) for different dependent variables. The unit of observation is a borrower-month. Panel (a) uses all borrowers and Panel (b) uses retirees, defined as those age 62 or older who receive a Social Security payment in at least three months of the calendar year. Future income is income one year in the future. Standard errors are clustered at the CBSA level.

*p < 0.1; **p < 0.05; ***p < 0.01
Table A-9: Impact of Home Equity on Default

(a) Chase Sample

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTV ≤ 0.85</td>
<td>−0.587</td>
<td>−0.480</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>LTV &gt; 0.95</td>
<td>0.216</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>LTV fitted residuals</td>
<td>1.697</td>
<td>1.644</td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.076)</td>
</tr>
</tbody>
</table>

α_{negative equity} 0.194 0.18 (0.023) (0.017)

Region-Year FEs Y N
CBSA FEs Y Y
Borrower and loan characteristics Y Y
CBSA controls Y Y
Origination year FEs N Y
Instrument Cyclicality-HPI Cyclicality-Month
First stage partial F-Stat 81.64 16.95
Log Likelihood -333,569 -332,744
Observations 1,432,248 1,432,248

(b) CRISM Sample

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTV ≤ 0.85</td>
<td>−0.523</td>
<td>−0.497</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>LTV &gt; 0.95</td>
<td>0.167</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>LTV fitted residuals</td>
<td>1.289</td>
<td>1.236</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.053)</td>
</tr>
</tbody>
</table>

α_{negative equity} 0.154 0.141 (0.022) (0.020)

Region-Year FEs Y N
CBSA FEs Y Y
Borrower and loan characteristics Y Y
CBSA controls Y Y
Origination year FEs N Y
Instrument Cyclicality-HPI Cyclicality-Month
First stage partial F-Stat 667.76 127.7
Log Likelihood -454,238 -454,059
Observations 1456127 1456127

Notes: This table replicates the default hazard model estimates from columns (3) and (4) of Table 7 using three LTV groups rather than a binary above water versus underwater comparison. Denoting the coefficient on LTV > 95 as \( \hat{\delta} \), the table also reports \( \hat{\alpha}_{negative\ equity} = 1 - exp(-\hat{\delta}) \). This captures the thought experiment of moving all borrowers above this LTV to an LTV of 90 (the average LTV of the omitted group).
Table A-10: Distribution of Checking Account Balances of Defaulters

<table>
<thead>
<tr>
<th>LTV p10</th>
<th>p25</th>
<th>p50</th>
<th>p75</th>
<th>p90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above water</td>
<td>0.5</td>
<td>31.2</td>
<td>344.8</td>
<td>1,294.8</td>
</tr>
<tr>
<td>Underwater</td>
<td>3.0</td>
<td>46.9</td>
<td>417.5</td>
<td>1,463.6</td>
</tr>
</tbody>
</table>

Notes: This table shows the distribution of checking account balances in dollars at the date of default for the primary analysis sample in the JPMCI data. To avoid disclosing information for any single household, the table reports pseudo-medians based on cells of at least 10 observations. Note that this table describes borrowers at the date of default, which is different from Table 4 in the main text, which describes borrowers six months before default.

Table A-11: Income and Assets of Defaulters by Loan-to-Value

<table>
<thead>
<tr>
<th>LTV Drop</th>
<th>As share of income</th>
<th>Drop as share of mortgage payment due</th>
<th>Checking Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;100</td>
<td>-0.238</td>
<td>-0.928</td>
<td>1,115</td>
</tr>
<tr>
<td>100-120</td>
<td>-0.253</td>
<td>-0.860</td>
<td>1,203</td>
</tr>
<tr>
<td>120-140</td>
<td>-0.253</td>
<td>-0.856</td>
<td>1,250</td>
</tr>
<tr>
<td>140-160</td>
<td>-0.257</td>
<td>-0.896</td>
<td>1,304</td>
</tr>
<tr>
<td>160+</td>
<td>-0.261</td>
<td>-0.947</td>
<td>1,290</td>
</tr>
</tbody>
</table>

Notes: This table measures economic conditions at the time of default by loan-to-value (LTV) bin. The first two columns show measures of the average income drop from one year prior to default to the month of default and the third column shows mean checking account balances at the date of default. Note that this table describes borrowers at the date of default, which is different from Table 4 in the main text, which describes borrowers six months before default.

Table A-12: Distribution of Home Equity and Default versus Benchmarks in 2009

<table>
<thead>
<tr>
<th>LTV bin</th>
<th>Default rate</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>JPMCI</td>
<td>CRISM</td>
</tr>
<tr>
<td>LTV &gt; 100</td>
<td>10.9%</td>
<td>9.7%</td>
</tr>
<tr>
<td>80 &lt; LTV ≤100</td>
<td>3.8%</td>
<td>3.8%</td>
</tr>
<tr>
<td>LTV ≤80</td>
<td>1.3%</td>
<td>1.3%</td>
</tr>
</tbody>
</table>

Notes: This table compares the distribution of home equity and default for mortgage borrowers in Chase to the Credit Risk Insight Servicing McDash (CRISM) dataset, and the Panel Study of Income Dynamics (PSID) in 2009.
Table A-13: Efficacy of Methods of Estimating $\alpha$ in Simulated Data

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Group/Formula</th>
<th>Independent</th>
<th>Correlated</th>
</tr>
</thead>
<tbody>
<tr>
<td>'All defaults' method for estimating $\alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share with life event $P(T^*)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All underwater [1]</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>Underwater: defaulters [2]</td>
<td>56%</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>Above water: defaulters [3]</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Share of defaults caused by life events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha} = (2 - 1)/(3-1)$</td>
<td>0.11</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.11</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha} - \alpha$</td>
<td>0.01</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>'Transition to default' method for estimating $\alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share with life event $P(T^*)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All underwater [1]</td>
<td>40%</td>
<td>46%</td>
<td></td>
</tr>
<tr>
<td>Underwater: defaulters [2]</td>
<td>59%</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Above water: defaulters [3]</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Share of defaults caused by life events</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha} = (2 - 1)/(3-1)$</td>
<td>0.19</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.17</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>$\hat{\alpha} - \alpha$</td>
<td>0.01</td>
<td>0.09</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table uses simulations to explore estimates of $\alpha$ (the share of defaults caused by negative life events) under different assumptions about the data-generating process. In the “Independent” scenario, we assume that the probability of life events is independent of the borrower’s default cost. In the “Correlated” scenario, we assume that the probability of a life event is correlated with the borrower’s default cost. The top panel shows estimates where we study the correlates of all defaults and the bottom panel shows estimates where we study the correlates of transitions to default. See Appendix C.3 for details.

Table A-14: Income Drop at Default by Home Equity in Actual Data

<table>
<thead>
<tr>
<th>Dependent variable: change in income</th>
<th>All underwater borrowers</th>
<th>Above water defaulters</th>
<th>Underwater defaulters</th>
<th>$\hat{\alpha}$</th>
<th>Default definition</th>
<th>Transition to default</th>
<th>All periods in default</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.009</td>
<td>-0.928</td>
<td>-0.871</td>
<td>0.938</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>-0.245</td>
<td>-0.238</td>
<td>0.974</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table shows the change in income from one year before default to the date of default as a share of the mortgage payment due in the JPMCI analysis sample. See Appendix C.3 for details.
Table A-15: Distribution of Payment Behavior After Three Missed Payments by Home Equity

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Share of above water</th>
<th>Share of underwater</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Made next 1 payment</td>
<td>0.390</td>
<td>0.277</td>
<td>0.113</td>
</tr>
<tr>
<td>Made next 2 payments</td>
<td>0.237</td>
<td>0.163</td>
<td>0.073</td>
</tr>
<tr>
<td>Made next 6 payments</td>
<td>0.094</td>
<td>0.068</td>
<td>0.026</td>
</tr>
<tr>
<td>Missed next 1 payment</td>
<td>0.604</td>
<td>0.722</td>
<td>-0.118</td>
</tr>
<tr>
<td>Missed next 2 payments</td>
<td>0.484</td>
<td>0.619</td>
<td>-0.135</td>
</tr>
<tr>
<td>Missed next 6 payments</td>
<td>0.266</td>
<td>0.381</td>
<td>-0.115</td>
</tr>
</tbody>
</table>

Note: This table shows the share of borrowers for each group in Figure A-24

Table A-16: Measures of Unemployment Among Mortgagors by Home Equity

<table>
<thead>
<tr>
<th></th>
<th>Unemployed</th>
<th>Receive UI by direct deposit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above water</td>
<td>7.3%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Underwater</td>
<td>12.2%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Ratio underwater to above water</td>
<td>1.67</td>
<td>1.68</td>
</tr>
<tr>
<td>Data source</td>
<td>PSID</td>
<td>JPMCI</td>
</tr>
</tbody>
</table>

Note: PSID definition is head of household or spouse is unemployed at time of survey, using the 2009 and 2011 waves of the PSID. JPMCI definition is the share of borrowers who receive unemployment insurance by direct deposit in any given three month period from 2008 to 2015.
B Data Appendix

B.1 JPMCI Data

To be included in the analysis sample, we require that the household have an open checking account (to measure income) from one year before default through the date of default. The checking account data are available beginning in January 2007. The mortgage servicing data are available through August 2015. We define the date of default as the first date a default appears in the matched checking account-servicing dataset after January 2007. To meet the requirement of having income history for the year before default, we study defaults from January 2008 through August 2015.

The unit of observation in this study is a first lien mortgage. There are 139,212 mortgages which meet this definition of default, have reliable data on payments made, have non-missing loan-to-value ratios, and have income data available for one year prior to default. There are 133,997 unique households associated with these 139,212 mortgages; this situation arises because there are a very small number of households that default on multiple first lien mortgages that are serviced by Chase.

Our primary analysis sample uses borrowers who reach delinquency of 90 days past due for the first time and have checking account data available in the 12 months prior to reaching 90 days past due. In some robustness analyses, we use alternative definitions of default: 30 days past due, 60 days past due, 120 days past due, 150 days past due, or a foreclosure initiation. For these alternative definitions of default, we rebuild the sample such that it meets the balanced panel requirement of checking account data available 12 months prior to the date of default.

After building the sample, we also take further steps to clean the data. The following variables are winsorized to the 95th percentile of positive values: monthly income, end-of-month checking account balance, monthly payment due, monthly payment made, and property value. LTV is defined as the median of non-missing LTVs between 9 and 3 months before default.

In some cases, a customer will have more than one checking account with Chase. In this case, we define income and balances summing across accounts. The crosswalk from customers to accounts is only available from 2009 forward; for income and balances prior to 2009, we use the 2009 crosswalk.

B.2 CRISM Data

The CRISM data is composed of two datasets: a mortgage servicing dataset from McDash and a credit bureau dataset from Equifax. The datasets are linked by the availability of a mortgage ID key in the credit data. For our benchmarking analysis, we take a 1 percent sample of these consumers for computational reasons.

In analyzing the CRISM data, we broadly follow the data-cleaning choices in Beraja et al. (2019). For each consumer, we consider their loans in both Equifax and McDash. We restrict our attention to those consumers that are first observed with less than three first liens inEquifax, and less than three of any type of second lien. Equifax only reports separate features (such as origination date, outstanding balance, origination balance) for the largest two loans, and also reports variables that reflect aggregate totals for all loans. Restricting
to those who start the panel with two or fewer loans allows us to track a third loan through changes to the total variables.

We convert the Equifax data to a loan-level panel, identifying a loan by its origination date, origination amount, and lien type, and match loans in Equifax to those observed in McDash. This serves as a check on the quality of the Equifax match, and gives us more comprehensive information on second liens. We consider an Equifax loan/McDash loan pairing a match if the origination date of the Equifax loan is within one month and the origination amount is within $10,000 of the McDash loan. If more than one McDash loan is matched, we use the date of origination, origination amount, and date of termination as tie breakers. We allow multiple Equifax loans to be matched to a single McDash loan, since occasionally loan information is revised in Equifax.

For each loan in Equifax, we consider second liens (closed-end seconds and HELOCs) that are (i) from the same consumer ID, (ii) opened alongside or after the first lien (allowing for 3 months delay in reporting), and (iii) lower in origination balance than the first lien. If a second lien is plausibly assigned to multiple first liens, it gets assigned to any first lien that has the maximum balance at origination among those first liens.

We use default information and primary balance from McDash and merge it onto second lien information as established above. This allows us to observe a measure of the cumulative unpaid balance at default. For each loan, we also observe a ZIP code and CBSA in McDash. We use this geographic information to inflate appraisals at origination by CoreLogic house price indexes. We use the CBSA index where the ZIP code index is unavailable. After estimating prices and the unpaid balance, we have estimates of a cumulative loan-to-value at default. Additional information from the McDash occupancy field is used to classify borrowers into investors or primary occupants. Additional information from Equifax is used to observe whether borrowers are current on lines of revolving credit.

In calculating rates of foreclosure within 12 months, we consider the universe of loans that newly transition into default at a given date. We consider the share in foreclosure as the proportion of these loans that are observed in foreclosure at any time from the first 90 day default to 12 months following the first 90 day default.

For the analysis in Figure A-12, we use the Equifax side of the data to measure the concurrence of default on non-HELOC revolving debt and on the primary mortgage lien. This deviates slightly from the analysis in the remaining exhibits, which only uses the McDash default date. We measure the concurrence of default in a single dataset to avoid any potential issues with differences in timing between the Equifax and McDash data.

### B.3 Quantifying Measurement Error in Observed LTVs

We use two components of the CoreLogic data from 1989 to 2019: the house price indexes (to measure predicted sale values) and the deed data (to measure the change in prices using 12 million home sales).

In analyzing the CoreLogic data, we broadly follow the data-cleaning choices of Giacoletti (2021), with a few exceptions. First, we use a ZIP-level CoreLogic price index instead of the Zillow index that Giacoletti uses. Second, Giacoletti requires properties to also appear in the tax data; we do not, because the tax data is not required for our analysis. Third, we do not remove observations in the top and bottom 2.5 percent of errors in the main
analysis. We want to avoid understating the mass in the tails of the error distribution since this is where misclassification is most likely. Instead, we follow Kotova and Zhang (2020) in removing anomalous transactions that have aggregate appreciation or depreciation of more than 50 percent per year over their holding period. Fourth, we do not restrict our sample geographically and take any valid pair of transactions in the deed data as our sample. Finally, although we maintain Giacoletti’s sample restriction regarding the earliest date of transactions to avoid issues with data quality of earlier records, we do not restrict our sample to final transactions taking place in or before 2013. This extends our sample through to the beginning of 2019.

C  Econometric Assumptions, Proofs, and Simulations

C.1  Assumption 1 – Prevalence of Above Water Default and Foreclosure

There are two frictions that make above water default quite common. First, there are substantial frictions to accessing home equity for borrowers in financial distress (Boar, Gorea and Midrigan Forthcoming; DeFusco and Mondragon 2020). Underwriting for refinancing and second liens requires a good credit history and a documented “source of repayment” (Office of the Comptroller of the Currency 2005), which usually means proof of income or proof of substantial liquid assets (Fannie Mae 2011). In the Fannie Mae underwriting guide, unemployment insurance is not an acceptable source of income. An unemployed homeowner who needs a loan to cover her current mortgage payments would not meet the prevailing underwriting standard during our sample period. Second, borrowers may also choose to sell their home, but there are frictions in this process as well (Gilbukh and Goldsmith-Pinkham 2021). For example, Guren (2018) documents that less than half of listed homes were sold within three months.

These two frictions mean that above water default is quite common. In Figure A-17, we corroborate Low (2018)’s finding that above water default is ubiquitous. The figure shows that even at the peak of the housing crisis, 40 percent of defaults were by above water borrowers. Furthermore, not only is above water default ubiquitous, but the economic risks from such default are substantial. Borrowers face an immediate credit score impact. The credit score decline from falling behind by three months on a mortgage is more than 80 percent as large as the decline from foreclosure and almost 60 percent as large as the decline from bankruptcy (Christie 2010). Perhaps the bigger risk, however, is foreclosure.

Above water foreclosures are common too; in fact, the rate of foreclosure initiations among borrowers who have missed three payments is similar and high for above water and underwater borrowers. In our sample, we calculate that 40 percent of above water borrowers who fall behind by three months have a foreclosure initiation within one year (Table A-2). This is only slightly below the foreclosure initiation rate of 45 percent for underwater borrowers.

Why would a lender foreclose when the value of the collateral exceeds the value of the loan? The high propensity to foreclose even on above water borrowers has both institutional and economic roots. Many mortgage servicers are not even allowed to consider home equity in the foreclosure decision because of rules made by the government-sponsored enterprises
Although the lender would prefer that an above water defaulter sell their home and repay the loan in full, they cannot instantaneously force a sale; instead, the foreclosure process is the legal mechanism by which the lender attempts to trigger a sale. Further, Low (2018) shows that above water foreclosures are an equilibrium outcome in a quantitative model with matching frictions in the home sale market that make the time-to-sell and resale value uncertain. The GSE practice of ignoring home equity sets the standard for the industry. Even for non-GSE-owned loans, servicers who do not follow the industry standard face increased litigation risk.

As a matter of economics, it is not obvious whether lenders should foreclose more quickly on above water or underwater homes whose mortgages are in default. On the one hand, the return to the lender from foreclosing on an above water home is higher because the lender will likely recoup the full balance outstanding on the loan (which is unlikely to occur for an underwater loan). On the other hand, an additional month of waiting for an above water borrower to sell their own home may yield a higher sale price than a resale through foreclosure and also enables the lender to avoid the upfront administrative costs associated with foreclosure.

### C.2 Proof of Proposition 1

\[
\alpha_{\text{life event}} = \frac{E(Y|G = 1) - E(Y(0, 1)|G = 1)}{E(Y|G = 1)}
\]

\[
= 1 - \frac{E(Y(0, 1)|G = 1, T^* = 0)}{E(Y|G = 1)}
\]

\[
= 1 - \frac{P(Y = 1|T^* = 0, G = 1)}{P(Y = 1|G = 1)}
\]

\[
= 1 - \frac{P(T^* = 0|Y = 1, G = 1)}{P(T^* = 0|G = 1)}
\]

where the first step uses assumption 3 (random assignment of \(T^*\)), the second step uses that \(Y\) is binary, and the third step uses Bayes rule. We first analyze the numerator \(P(T^* = 0|Y = 1, G = 1)\) and then analyze the denominator \(P(T^* = 0|G = 1)\). Although neither the numerator nor the denominator are identified without further assumptions, the ratio of the two is identified using Assumptions 1-4.

The law of iterated expectations implies that

\[
E(T|Y = 1, G = 1) = P(T^* = 0|Y = 1, G = 1)E(T(0)|T^* = 0, Y = 1, G = 1) + (1 - P(T^* = 0|Y = 1, G = 1))E(T(1)|T^* = 1, Y = 1, G = 1)
\]
where $T(T^* G, Y) = T(T^*)$ from Assumption 4a. Re-arranging terms gives:

$$
P(T^* = 0|Y = 1, G = 1) = \frac{E(T(1)|T^* = 1, Y = 1, G = 1) - E(T|Y = 1, G = 1)}{E(T(1)|T^* = 1, Y = 1, G = 1) - E(T(0)|T^* = 0, Y = 1, G = 1)}
$$

$$
= \frac{E(T(1)) - E(T|Y = 1, G = 1)}{E(T(1)) - E(T(0))}
$$

(16)

where the second equality follows from Assumption 4a. This object exists because $E(T(1)) - E(T(0)) \neq 0$ by Assumption 4b. We can identify $E(T(1))$ because

$$
E(T(1)) = E(T|Y = 1, G = 0, T^* = 1)P(T^* = 1|Y = 1, G = 0)
$$

$$
= E(T|Y = 1, G = 0)
$$

(17)

where $P(T^* = 1|Y = 1, G = 0) = 1$ by Assumption 1. Substitute equation (17) into the numerator of equation (16) to get

$$
P(T^* = 0|Y = 1, G = 1) = \frac{E(T|Y = 1, G = 0) - E(T|Y = 1, G = 1)}{E(T(1)) - E(T(0))}
$$

(18)

This expression captures the numerator of the ratio in equation (15). Applying the same logic to the denominator in the ratio of equation (15) gives

$$
P(T^* = 0|G = 1) = \frac{E(T(1)|T^* = 1, G = 1) - E(T|G = 1)}{E(T(1)|T^* = 1, G = 1) - E(T(0)|T^* = 0, G = 1)}
$$

$$
= \frac{E(T(1)) - E(T|G = 1)}{E(T(1)) - E(T(0))}
$$

$$
= \frac{E(T|Y = 1, G = 0) - E(T|G = 1)}{E(T(1)) - E(T(0))}
$$

(19)

where $E(T|G = 1)$ includes both underwater defaulters and non-defaulters. We take the ratio of equations (18) and (19). The denominators $(E(T(1)) - E(T(0)))$ cancel, so

$$
\frac{P(T^* = 0|Y = 1, G = 1)}{P(T^* = 0|G = 1)} = \frac{E(T|Y = 1, G = 0) - E(T|Y = 1, G = 1)}{E(T|Y = 1, G = 0) - E(T|G = 1)}
$$

Plugging this ratio into equation (15) gives

$$
\alpha_{\text{life event}} = 1 - \frac{P(T^* = 0|Y = 1, G = 1)}{P(T^* = 0|G = 1)} = \frac{E(T|Y = 1, G = 1) - E(T|G = 1)}{E(T|Y = 1, G = 0) - E(T|G = 1)}
$$

Note that $E(T(0))$ cancels when computing $\alpha$ and so knowledge of $E(T(0))$ is not necessary for identifying $\alpha$. This is why is it possible to identify the causal object $\alpha$ even though both the treatment effect and the probability of treatment are unknown.
C.3 Relaxing Assumption 3

C.3.1 Simulate data

We posit a statistical model of default behavior which is designed to capture the three possible theories of default: negative life events (cash-flow), negative equity (strategic), and double-trigger. Assume a panel of borrowers, each observed for $S$ periods. In each period, borrower default is a function of three borrower-specific variables: $T_s^*$ is a life event, $\eta_s$ is a purely temporary default shifter which governs the excess default motivations of underwater borrowers, and $\varepsilon$ is a permanent attribute, where a high $\varepsilon$ indicates a low cost of default. $\varepsilon$ is intended to capture long-run attributes that affect the probability of default such as the “moral” or “stigma” cost of default. The key assumption in the simulation is that $\varepsilon$ reflects a cost of default that is stable within the time horizon studied in the simulation. We assume that the three variables enter the default equation additively

$$Y_s = 1(\varepsilon + \eta_s + \beta T_s^* > a)$$

which allows for the possibility of interactions between the different forces (i.e., double-trigger behavior). We further assume the following distributions for the primitives:

$$\varepsilon \sim Unif[0, \bar{\varepsilon}]$$
$$\eta_s \sim Unif[0, \bar{\eta}]$$

We consider two polar cases for the data-generating process for negative life events:

$$T_s^* \sim \begin{cases} 
\text{Bernoulli}(p) & \text{Independent} \\
\text{Bernoulli}(p\frac{2\varepsilon}{\bar{\varepsilon}}) & \text{Correlated}
\end{cases}$$

Finally, relative to the framework in the paper, we treat $T^*$ as observed and do not model $T$ because the purpose of this simulation is to focus on causal inference regarding the role of $T^*$.

The goal of the simulation is to enable us to understand the impact of relaxing Assumption 3 on the paper’s estimates. In the “Independent” scenario, Assumption 3 is satisfied because life events are random. We also consider an alternative, less stringent version of Assumption 3, which is:

**Assumption 3**: (conditional exogeneity): $\{Y_s(0,1,\varepsilon), Y_s(1,0,\varepsilon), Y_s(1,1,\varepsilon)\} \perp T_s^*|G, \varepsilon$

In the “Correlated” scenario, Assumption 3 is not satisfied because there is a correlation between the individual cost of default $\varepsilon$ and the probability of a life event, but Assumption 3 is satisfied.

The prior literature provides no obvious guidance on the sign of the omitted variable bias (whether $\varepsilon$ is positively or negatively correlated with the probability of a life event $T_s^*$).³ We focus in our simulations on the negative correlation case because this paper’s primary

³One strand of papers posits that strategic default is most common among borrowers who are investors and have low attachment to the home in question, and therefore may have low private costs of default (Albanesi, De Giorgi and Nosal, 2017). These borrowers, who previously had enough resources to buy properties beyond their primary residence, may be less likely to have negative life events than the average
conclusion is that almost no defaults are strategic and that conclusion might change when Assumption 3 is relaxed to allow for a negative correlation (i.e., a negative correlation is what would lead us to understate the share of strategic default). We allow for a severe form of omitted variable bias: the probability of a life event ranges from 100 percent for the borrowers with the lowest cost of default to nearly 0 percent for the borrowers with the highest cost of default. We normalize the probability of a life event by \( 2/\bar{\varepsilon} \) so that the overall probability of a life event is similar under the two assumptions about the data-generating process for \( T^* \). Furthermore, we adopt parameters such that the bias in our estimate is maximized, which happens when the share of defaults causally attributable to life events is small. We show that, even with these two extreme assumptions, the overall bias in our estimator is small. Furthermore, when we adopt more realistic assumptions, the bias is fully mitigated.

C.3.2 Analyze simulated data

Analysis of Using All Periods in Default to Estimate \( \alpha \) When we use our method to analyze every period in which a borrower is in default, our methodology does well when Assumption 3 holds but does poorly in some parameterizations when Assumption 3 is relaxed. In this subsection, we deviate from the paper by defining \( Y_s = 1 \) as any period with a default. We provide this example to illustrate how omitted variable bias can lead to misleading conclusions about a causal relationship, but note that this is not the method we use in the paper. We discuss bias from the actual method in the next section.

In the case where Assumption 3 holds \( (Y_s(0,1) \perp T^*_s|G) \), the probability of a life event \( T^*_s \) is independent of the cost of default \( \varepsilon \).\(^4\) We simulate a parameterization of the model where \( \bar{\varepsilon} = 5, \bar{\eta} = 4, p = 0.5, \beta = 0.5, a = 5 \) with 10,000 borrowers each of whom live for 10 periods. A sample size of 10,000 borrowers for the simulation is sufficiently large that we can ignore issues of sampling variation and focus on the key question of causal inference. We show robustness to alternative parameterizations in Section C.3.2. We assume that the probability of a life event is independent of the other model parameters \( (T^*_s \sim \text{Bernoulli}(p)) \) and call this the “Independent” scenario.

Figure A-18 shows in the first panel that, consistent with this assumption, the average probability of a life event does not vary with the cost of default in the Independent scenario. It also shows in the second panel that the probability of default is highest for those with the lowest cost of default (those with high \( \varepsilon \)) and in the third panel that the default rate would change very little if there were no life events. We estimate equation 3 (substituting \( T^* \) for \( T \)). Table A-13 shows that we estimate \( \hat{\alpha}_{xsec} = 0.11 \), which means that 11 percent of defaults are caused by life events. This matches the true parameter value in the simulation \( (1 - E(Y_s(0,1)|Y_s = 1, G = 1)) \). This is to be expected, because we proved that this set of borrower. If that was true, then we might expect a positive correlation between default costs and probability of a negative life event. A second strand posits that strategic default is most common among subprime borrowers who have low attachment to credit markets overall, and therefore may also have low private costs of default since their credit score is already low (Mayer et al., 2014). These borrowers may be more likely to have negative life events than the average borrower if low credit scores are more common for borrowers with high latent unemployment risk (perhaps arising from past spells of unemployment). If this was true, then we might expect a negative correlation between default costs and the probability of a life event.

\(^4\)To be precise, Assumption 3 is that \( Y(0,1) \perp T^*|G \), while here we add a subscript \( s \) to capture the panel dimension of the simulation. We otherwise ignore the panel dimension of the data in this section.
conditional expectations identifies the causal object of interest in Proposition 1.

Next, we consider a case where we relax Assumption 3 by introducing substantial omitted variable bias and show that the estimator applied to all periods with a default does poorly. Specifically, we assume that the probability of a life event is higher for people with a low stigma cost of default \( T_s^* \sim \text{Bernoulli}(p_{\epsilon}) \) and call this the “Correlated” scenario. Figure A-18 shows that in this scenario, the probability of a life event varies from 100 percent for people with the lowest cost of default to 0 percent for people with the highest cost of default. By design, the extent of omitted variable bias is extreme in this situation: the correlation of \( \epsilon \) and \( T_s^* \) is greater than 0.5. In this case, we estimate in Table A-13 that \( \hat{\alpha}_{xsec} = 0.54 \), when in fact the true \( \alpha \) is 0.12.

The estimator does poorly because of the standard intuition of how omitted variable bias can lead an analyst to overstate the strength of a causal relationship. The data in this simulation feature borrower life events and borrower default that coincide, but life events have little causal impact on default. To clarify how this affects our estimates of \( \alpha \), Figure A-19’s left panel shows that the “Correlated” scenario shows excess mass in the green bar relative to the “Independent” scenario. This means that there is a substantial share of defaults that have a life event but are not caused by a life event. A higher rate of life events among defaulters than among the general population leads to an (incorrect) inference that life events are causing many defaults.

**Analysis of Transitions to Default to Estimate \( \alpha \)** We redo the analysis from Section C.3.2, limiting attention to the subsample where \( Y_{t-1} = 0 \). This is the methodology we use in Section 4 of the paper. In the language of the model in the paper, which does not have a time dimension, \( Y = 1 \) is the transition to default. This is the natural outcome of interest in our setting. Every loan is current at the time of origination, and so every default requires a transition from not defaulting to defaulting. Understanding borrower default therefore requires understanding why borrowers transition into default. This is why the prior literature on mortgage default often evaluates default as time-to-failure through the lens of a hazard model.

This estimand exactly identifies the parameter of interest when Assumption 3 holds and shows substantial improvement in the scenario where Assumption 3 is relaxed. We analyze the same simulated data as in the previous section. In the “Independent” scenario, Table A-13 shows that the simulation estimate matches the true parameter value. Again, this is to be expected because of Proposition 1. In the “Correlated” scenario, the magnitude of the bias is substantially reduced; it is one-fifth as large as when we analyze all defaults.

Focusing on the transition to default mitigates the potential bias from relaxing Assumption 3 in our simulation because it drops borrower-periods where permanent heterogeneity causes misleading conclusions. These are periods after the initial transition to default for the borrowers with a low cost of default (regardless of whether they have experienced a life event) and a high probability of negative life events. Figure A-20’s left panel shows that the defaults that coincide with a life event but would have happened without it are concentrated among borrowers with a low permanent cost of default. Figure A-21 shows that narrowing the sample to just examine transitions to default drops most of the misleading default-periods. Although the extent of the bias could differ under alternative data-generating processes, we note that we have deliberately selected a data-generating process where omitted variable
bias is severe (because the correlation between the cost of default \( \varepsilon \) and the probability of a life event \( T^* \) is quite strong).

Fixed effects models offer a helpful, albeit inexact, analogy for why analyzing transitions to default leads to minimal bias even in the presence of unobserved heterogeneity. Had \( Y_s \) been linear in the latent confounder \( \varepsilon \), even after we relax Assumption 3, a first-difference specification would have enabled us to exactly identify \( \alpha \). Because \( Y_s \) is binary, the identification results from the continuous case do not hold. However, the intuition that focusing on transitions differences out the latent type carries over to this setting.

We conduct additional simulations to verify the robustness of our conclusions about bias across the support of the parameter of interest. Specifically, relative to the base simulations described above, we progressively reduce the importance of permanent heterogeneity (by decreasing \( \bar{\psi} \) to 2.5) while simultaneously increasing the effect of life events on default (increasing \( \beta \) to 3).\(^5\) These changes drive \( \alpha \) up to about 0.80. Finally, to raise \( \alpha \) even further, we lower the probability of life events \( p \) to 0.1 and lower \( \bar{\eta} \) to 2.75, which raises \( \alpha \) to 0.94.

The results are shown in Figure A-22 and contain three lessons. First, defining \( Y = 1 \) as the transition to default greatly reduces (but does not eliminate) bias at most values of \( \alpha \). Second, the estimates are similar from using all periods with a default and using transitions to default when \( \alpha \) is close to 1. To understand why the estimates are similar, see the panel labeled “Realistic” in Figure A-21, which shows that most defaults were not preceded by a default in the previous period. The two samples (all defaults and transitions to default) therefore study similar samples and yield similar conclusions. Third, there is no evidence of bias in the simulation when \( \alpha \) is close to 1.

The results in Figure A-22 suggest a further empirical test using actual data. The figure shows that the estimates using all defaults and using transitions to default are similar when either of two conditions is satisfied: (i) when Assumption 3 is satisfied or (ii) when \( \alpha \) is close to 1. We therefore re-implement our empirical methodology using every month in which a borrower is in default, instead of just using the first month and show the results in Table A-14. The estimate of \( \hat{\alpha} \) is 0.974 when we study all periods where a borrower is in default as compared to 0.938 when we study only the transition to default. Interpreted through the lens of the simulations, the similarity of the two estimates suggests that the bias from relaxing Assumption 3 is limited.

Based on both the simulations which relax Assumption 3 and the alternative empirical estimate that uses all defaults, it appears that our conclusions about the prevalence of strategic default are robust to relaxing Assumption 3.

\(^5\)Suppose we modify equation 20 such that \( Y_s = \varepsilon + \eta_s + \beta T^*_s \). Note that in the “Correlated” simulation scenario above, a conditional version of exogeneity holds \( (Y_s(0,1,\varepsilon) \perp T^*_s | G, \varepsilon) \). First-differencing gives \( E(Y_s - Y_{s-1}|\varepsilon, \eta_1 \ldots \eta_T) = \beta(T^*_s - T^*_{s-1}). \) In a model with homogeneous treatment effects, \( \beta \) gives the average causal effect of \( T^*_s \) on \( Y \); together with the probability of treatment \( P(T^*_s) \), this is sufficient to identify \( \alpha \).

\(^6\)Our identification method requires the presence of some above water defaults. In the simulation, this requires that \( \bar{\psi} + \beta > a \). Had we decreased \( \bar{\psi} \) with no change to another parameter, then this condition would not be fulfilled and no above water borrower would ever default. We therefore include offsetting increases to \( \beta \) so as to ensure that some above water default persists in the simulation, as it does in the data.
C.4 Bias-Variance Trade-off

This section uses a simple simulation to explore the bias-variance trade-off for the reverse regression method of causal attribution proposed in Section 2.3.2 relative to the standard method described in Section 2.3.1. We are interested in measuring the fraction of underwater mortgage default causally attributable to life event treatment \( T^* \). Assume that all borrowers have a 50 percent probability of binary treatment, i.e., \( T^* \sim Bernoulli(0.5) \). Assume that 25 percent of underwater borrowers (those with \( G = 1 \)) receive a binary strategic default shifter \( S \), so that

\[
S = \begin{cases} 
0 & \text{if } G = 0 \\
Bernoulli(0.25) & \text{if } G = 1 
\end{cases}
\] (21)

Borrowers default if they receive a negative life event, a strategic default shock, or both, i.e.

\[
Y = \mathbb{1}(T^* + S \geq 1).
\]

Given these parameters, 100 percent of underwater borrowers with a negative life event default, 25 percent of underwater borrowers without a negative life event default (due to the strategic motive), and 62.5 percent of all underwater borrowers default. Because only 25 percent of all underwater borrowers would default in the absence of life event treatment, the true \( \alpha_{\text{life event}} \) for underwater borrowers from equation (2) is

\[
\alpha_{\text{life event}} = 0.625 - 0.25 = 0.6.
\]

In other words, life events are a necessary condition for 60 percent of underwater defaults in this setup.

Next, we introduce a measure of treatment that is noisy. In practice, we often do not observe the treatment itself (e.g., the life event \( T^* \)) but rather a noisy proxy (e.g., the change in income, which we denote by \( T \)). In the simulation, we assume that the observed income change is centered at zero, falling on average for those with negative life events and rising on average for those without, but is observed with a normally distributed measurement error \( \varepsilon \sim N(0, \sigma^2) \), which gives:

\[
T = 0.5 - T^* + \varepsilon.
\]

We consider two measurement error scenarios. In the baseline case, we assume only a small amount of measurement error with a standard deviation of 0.05, i.e., \( \varepsilon_{0.0025} \sim N(0, 0.0025) \). In the noisy scenario, we assume a higher degree of measurement error, i.e., \( \varepsilon_{0.25} \sim N(0, 0.25) \). We simulate each parameterization of the model with 10,000 above water borrowers and 10,000 underwater borrowers and assume that the econometrician can observe \((Y, T, G)\) for all borrowers in the simulation, but not \( T^* \). In each case, we can compare the standard “back-of-the-envelope” method for estimating \( \alpha_{\text{life event}} \) from Section 2.3.1 (which we denote here as \( \tilde{\alpha}_{\text{life event}} \)) to the reverse regression or “Bayes” method proposed in Section 2.3.2 (which we denote here as \( \hat{\alpha}_{\text{life event}} \)). The results are summarized in Figure A-23.

C.4.1 Baseline scenario

The standard “back-of-the-envelope” method of estimating \( \alpha_{\text{life event}} \) from equation (5) requires three inputs: an estimate of the average treatment effect on default, the probability
of treatment, and the probability of default. Because treatment is randomly assigned, to obtain the average treatment effect the researcher would simply regress default on the noisy measure of treatment, i.e.,

\[ Y = \lambda + \beta T \]  

(22)

and use \(-\hat{\beta}\) as the average treatment effect (since treatment leads to an observed income decrease). In the baseline simulation, we calculate \(-\hat{\beta} = 0.743\) for underwater borrowers, close to the true causal impact of treatment for underwater borrowers of 0.75. There is minimal attenuation bias in this estimate because the noise in \(T\) is minimal. Furthermore, the estimate of the causal impact is precise, with a standard error of 0.006. The researcher infers the probability of treatment from the share of underwater borrowers with an observed income decline, which is 50.5 percent \((\hat{P}(T^*) = 0.505)\), and observes an underwater default rate \(E(\tilde{Y}) = 0.628\). Plugging these in to equation (2) gives

\[ \hat{\alpha}_{\text{life event}}^{\text{baseline}} = \frac{(0.743)(0.505)}{0.628} = 0.597, \]

which, as with the causal impact, is close to the true \(\alpha_{\text{life event}}\) because measurement error is minimal. We calculate a standard error of 1 percent using the delta method.

We can also implement the “Bayes” method for estimating \(\alpha_{\text{life event}}\) from Section 2.3.2. Implementing this method using equation (6) also requires three inputs: the average observed income change for underwater defaulters \((E(T|Y = 1, G = 1))\), the average observed income change for above water defaulters \((E(T|Y = 1, G = 0))\), and the average observed income change for all underwater borrowers \((E(T|G = 1))\). In the simulation, the average change in income for underwater defaulters is -0.304, the average change for above water defaulters is -0.501 (all these borrowers experience a negative life event, which leads on average to an observed income decrease of 0.5), and the average change of income for all underwater borrowers is -0.005 (since just over half receive a negative life event, which leads to an average income change of -0.5 and just under half receive no negative life event, which leads to an average income change of 0.5). Plugging these values into equation (6) gives

\[ \hat{\alpha}_{\text{life event}}^{\text{baseline}} = \frac{(-0.304) - (-0.005)}{(-0.501) - (-0.005)} = 0.603, \]

which is also nearly identical to the true \(\alpha_{\text{life event}}\). We calculate a standard error of 1.1 percent using the delta method.

The point estimates and confidence intervals for both \(\hat{\alpha}_{\text{life event}}^{\text{baseline}}\) and \(\hat{\alpha}_{\text{life event}}\) are depicted visually in the left panel of Figure A-23.

C.4.2 Noisy scenario

We implement the same approaches for measuring \(\hat{\alpha}_{\text{life event}}^{\text{baseline}}\) and \(\hat{\alpha}_{\text{life event}}\) in the noisy scenario. Because \(T\) is measured with significant noise, there is now substantial attenuation bias in \(\hat{\beta}\) from equation (22). We estimate \(-\hat{\beta} = 0.381\), substantially below the true causal impact of 0.75. However, despite the bias in the estimate, because the outcome in equation (22) is measured just as precisely as in the baseline scenario, the precision of the estimate is just as good: the standard error on the estimate is again 0.006. The increased bias but
similar precision translate to an estimate of $\hat{\alpha}_{\text{noisy}}^{\text{life event}} = 0.301$, about half as large as the true $\alpha_{\text{life event}}$.

This trade-off is reversed with the Bayes method. The Bayes method estimates the income changes, which are now much noisier. However, they are unbiased. Thus, we find $\hat{\alpha}_{\text{noisy}}^{\text{life event}} = 0.607$, close to the true $\alpha_{\text{life event}}$, but measured much less precisely because the variance of each measured income change is now much larger. The standard error for the $\hat{\alpha}_{\text{noisy}}^{\text{life event}}$ nearly doubles, to 1.9 percent.

The point estimates and confidence intervals for both $\hat{\alpha}_{\text{noisy}}^{\text{life event}}$ and $\hat{\alpha}_{\text{noisy}}^{\text{life event}}$ are depicted visually in the right panel of Figure A-23. Attenuation bias from measurement error in treatment causes the estimate from the “back-of-the-envelope” approach to be biased towards zero (but still precise), whereas the “Bayes” approach is unbiased (but less precise).

C.5 Potential Outcomes Model with Instrument for Negative Equity

This section generalizes the model from Section 2 to allow for an instrument for the effect of negative equity.

Environment (continued)

We maintain the environment from Section 2.1. Instead of a potential outcome function $Y(T^*, G)$, we allow for a general instrument $Z$:

$$Y(T^*(G(Z), Z), G(Z), Z)$$

$Z$ refers to a cyclicality instrument conditional on controls and we discuss it in the text of the paper. In principle, the instrument $Z$ can affect default $Y$ in three ways: by changing the probability of a life event $T^*$, by changing the probability of negative equity $G$, and directly by affecting $Y$ holding constant $T^*$ and $G$.

Assumption 3′ ($Z$ is as good as random):

$$Y(T^*(G(Z), Z), G(Z), Z) \perp Z$$

$T^*(.), G(Z) \perp Z$

This assumption rules out omitted variable bias such that a correlation between $Z$ and $Y$ could exist without $Z$ also having a causal impact on $Y$.

Assumption 4′ (exclusion restriction, $Z$ affects $Y$ and $T$ only through its effects on $G$):

$$Y(T^*(G(Z), Z), G(Z), Z) = Y(T^*(G(Z)), G(Z)) \text{ and } T^*(G(Z), Z) = T^*(G(Z)).$$

This restriction rules out the possibility that $Z$ affects $T^*$ directly or $Y$ directly. We can expand the $T^*$ and $G$ arguments of the potential outcome function and restate both assumptions jointly as:

$$Y(1, 1), Y(1, 0), Y(0, 1), Y(0, 0), T^*(1), T^*(0), G(Z) \perp Z$$

We note that Assumption 4′ allows for the possibility that negative equity (induced by the instrument $Z$) increases the probability of a life event. See footnote 31 for additional details.
D Empirical Appendix

D.1 Income After Default

This appendix investigates the evolution of income after default. We find that payment and income dynamics after default are closely linked. Defaulters who resume making payments also have an income increase back to pre-default levels, while defaulters who do not resume making payments (and therefore fall behind further) continue to have depressed incomes.

We define additional notation for this analysis. Let \( t \) index dates and \( Y_t \) index delinquency. Let \( Y_t \in \{0, 30, 60, 90, 120, 150, 180, 210, 240, 270, \text{Foreclose}\} \) where the numeric values denote the number of days that the loan is past due and if the lender begins foreclosure proceedings, then the number of days is censored and instead \( Y_t \) is recorded as “Foreclose”. Finally, define

\[
\Delta Y_t \equiv \begin{cases} 
1 & Y_t > Y_{t-1} \text{ or } Y_t = \text{Foreclose}, Y_{t-1} \neq \text{Foreclose} \\
0 & Y_t \leq Y_{t-1}
\end{cases}
\]

Following the logic of our main empirical design (which conditions on payment behavior), we begin by conditioning on payment behavior for the first payment due after default. Figure A-24 shows the evolution of income for borrowers after default, using our baseline definition of three missed payments \((Y_t = 90, Y_t' \leq 60 \quad \forall t' < t)\). The date of default is therefore \( t \). In the top left panel we analyze borrowers whose delinquency does not deteriorate further at time \( t+1 \) \((Y_{t+1} \leq 90)\) and who therefore made at least one month’s worth of payments at date \( t+1 \). We refer to this group as having \( \Delta Y_{t+1} = 0 \). Income for the remaining borrowers, who miss the next payment due \((\Delta Y_{t+1} = 1)\), is shown in the bottom-left panel. These two groups partition the full set of possible payment outcomes in the month after default. After default, we find that borrowers who make a payment also experience an average income recovery, while the average income of borrowers who miss the next payment remains depressed. These patterns do not differ by home equity.

Similar patterns appear as we study additional payment history beyond the first payment due after default: borrowers who continue to make payments experience an income recovery while borrowers who continue to miss payments have incomes that remain depressed. The middle column of Figure A-24 shows borrowers who made the next 2 payments \((\Delta Y_{t+1} = 0, \Delta Y_{t+2} = 0)\) or missed the next two payments \((\Delta Y_{t+1} = 1, \Delta Y_{t+2} = 1)\). The right column shows borrowers who made the next 6 payments \((\Delta Y_{t+s} = 0, s \in \{1\ldots6\})\) or missed the next 6 payments \((\Delta Y_{t+s} = 1, s \in \{1\ldots6\})\). Unlike the left column, these columns with “made/missed next \( x \) payments” are no longer a full partition of payment outcomes, but nevertheless document income in scenarios of particular interest.

Two lessons emerge from Figure A-24. First, the basic “neutrality” result—that the income drop is similar for above water and underwater defaulters—holds when we look at specific payment histories after default. Second, income changes track payment histories, providing further evidence of the tight link between income and default.

Both the neutrality result and also the conclusion that income changes track payment histories extend to other definitions of default. Figure A-25 repeats the same analysis as
Figure A-24, but instead defines default as two missed payments ($Y_t = 60, Y_{t'} \leq 30 \forall t' < t$) and one missed payment ($Y_t = 30, Y_{t'} = 0 \forall t' < t$).

Against the backdrop of the neutrality results described above, we note that there are two well-known empirical patterns of differences between underwater and above water borrowers. The first pattern is that underwater borrowers are unconditionally more likely to experience negative life events in any period. This is the central finding in Bhutta, Dokko and Shan (2017). This finding also holds when using one common measure of life events (unemployment) in both datasets analyzed in this paper. We find that unemployment is almost twice as common for underwater than for above water borrowers, both in the PSID and in Chase (Table A-16).

The second pattern is that the rate of transition from less severe delinquency to more severe delinquency (“the roll rate”) is higher for underwater borrowers ($P(\Delta Y_1 = 1|G = 1) > P(\Delta Y_1 = 1|G = 0)$). This pattern is apparent in prior empirical research (see, e.g., Mayer et al. 2014), and also holds in our sample. Table A-15 shows the shares for each group in Figure A-24. After 90-day default, 72 percent of underwater borrowers miss the next payment while 60 percent of above water borrowers miss the next payment.

Another interesting way to describe the data is to document the path of average income after default, pooling across subsequent payment histories. Given the neutrality result shown in Figure A-24 ($E(T_{PD}|\Delta Y_1 = 1, G = 1) \approx E(T_{PD}|\Delta Y_1 = 1, G = 0)$ and $E(T_{PD}|\Delta Y_1 = 0, G = 1) \approx E(T_{PD}|\Delta Y_1 = 0, G = 0)$ where $T_{PD}$ is the change in income “post default” relative to 12 months prior to default) and the difference in roll rates by home equity, we expect that income after default will be more depressed for underwater borrowers than for above water borrowers. We expect that this will hold because of the law of iterated expectations

$$E(T_{PD}|G) = E(T_{PD}|\Delta Y_1 = 1, G) = P(\Delta Y_1 = 1|G) + E(T_{PD}|\Delta Y_1 = 0, G) = P(\Delta Y_1 = 0|G)$$

which implies that $E(T_{PD}|G = 1) < E(T_{PD}|G = 0)$. Figure A-26 confirms that, as expected, the pooled income after default is lower for underwater borrowers.

Figures A-24 and A-26 suggest that the higher rate of negative life events for underwater borrowers (pattern 1 above) could be the cause of the higher roll rates for underwater borrowers (pattern 2 above). Figure A-26 shows a difference by home equity in income after default (not conditioning on payment behavior), while Figure A-24 shows no difference in income after default (when conditioning on payment behavior). If controlling for differences in payment behavior eliminates differences in observed cash flows by home equity, a natural explanation is that the differences in cash flows were causing the differences in payment behavior.

### D.2 Mortgage Modification

This appendix section asks whether the availability of mortgage modifications contaminates our estimates of strategic default. Some mortgage modifications in the Great Recession

---

7This is also consistent with evidence in Bernstein (2021) and Gopalan et al. (2021), who find that borrowers with negative equity are more likely to suffer income declines because of constrained mobility and financial distress.
sought to set a borrower’s revised mortgage payment after modification as a certain share of their income. Herkenhoff and Ohanian (2011) and Mulligan (2009) pointed out that this created an implicit tax on labor income. To the extent that borrowers are aware of this implicit tax and able to reduce their labor supply, the availability of unrealized mortgage modifications would hold back borrower incomes before modification.

For our estimates to be contaminated by this channel, it would require differential effects by home equity, yet such differential effects are inconsistent both with theory and with empirical evidence. From a theoretical perspective, it would be surprising to see differential effects because it would need to be the case not only that underwater borrowers awaiting a modification were artificially depressing their labor supply but also that underwater borrowers were doing so more than above water borrowers. Because income-contingent modifications were available to both borrowers with positive equity and borrowers with negative equity, there is no theoretical reason why the channel would specifically distort the income of negative equity defaulters. From an empirical perspective, Figure A-27a plots income around mortgage modification for the universe of mortgages that received a modification in the JPMCI data. The figure shows no difference in the path of income for above water versus underwater defaulters, which is inconsistent with the view that our measure of income for underwater borrowers is contaminated by the availability of mortgage modifications.

Beyond this direct test, two other types of empirical evidence suggest that the availability of modifications is not driving labor supply for distressed borrowers: timing and heterogeneity by date of default. First, if a borrower were reducing labor supply on purpose so as to qualify for a more generous modification, we would expect to see incomes fall right before the modification date and recover after the modification was in place. Instead, average income begins declining 15 months prior to modification, bottoms out at 6 months before modification, gradually recovers close to its pre-modification level, and then is stable from 3 months before modification to 3 months after modification. Because income is constant around modification, we find no evidence that borrowers decrease their labor supply immediately before and then increase their labor supply immediately after modification.

Second, Figure A-27a masks meaningful heterogeneity in income by the date of default relative to modification. One potentially puzzling aspect of Figure A-27a is that the decline in income prior to modification is only about 20% of the mortgage payment due, which is much smaller than the decline in income prior to default documented in Figure 1 even though nearly all borrowers who received a modification defaulted prior to receiving one. We therefore disaggregate the series into three equally-sized bins by the number of months between default and modification in Figure A-27b. The left panel shows that borrowers do indeed have a large decline in income prior to their actual default date followed by a gradual recovery in income. It therefore appears that the date of default is driven by the timing of income declines, rather than income declines being driven by the timing of modification. Because Figure A-27a pools observations across different horizons between default and modification, it masks the declines in income which occur prior to default.

Further analysis shows even more clearly that default appears to respond to income rather than income responding endogenously to modification availability. One shortcoming of the left panel in Figure A-27b is that measuring the change in income relative to the first month of the plot is potentially confusing in this context. The problem is that the group with a long delay between default and modification may have already experienced a negative income
shock more than 18 months prior to modification, so their baseline income may already be depressed. To ease comparability across groups with different delays, the middle panel of Figure A-27b shows income in dollars without any normalization. It appears that there is a similar decline in income around default regardless of the time from default to modification. One ambiguity remaining in the middle panel is that each group’s average income declines over a 9 month period and it is not possible to tell whether the plot masks sharper negative income shocks to individuals. We find that it does. The right panel re-centers the income data by date of default and shows that there is a sharp decline in income prior to default, especially for people who experience long delays between default and modification. There is little evidence of heterogeneity in the recovery in income after default by time from default to modification. The centrality of default—and not modification—in explaining income dynamics suggests that the availability of modifications does not contaminate our estimates of strategic default. Borrowers appear to respond to income declines by defaulting. They don’t appear to cause income declines to coincide with a modification.
E Stigma Cost of Default

We are interested in finding the change in average per-period consumption ($\Delta c$) that would cause a change in utility equal to a one-time stigma cost:

$$u(\bar{c} + \Delta c) = u(\bar{c}) + stigma.$$  

We know that utility from consumption takes the following form:

$$u(c_t) = \sum_{t=1}^{T} \beta^{t-1} \frac{c_t^{1-\gamma}}{1-\gamma}$$

Assuming constant consumption ($\bar{c}$) across all periods, and letting $\gamma = 2$, as in Campbell and Cocco, we get

$$u(\bar{c}) = \frac{1}{\bar{c}} \sum_{t=1}^{T} -\beta^{t-1} = \frac{1}{\bar{c}} \lambda$$

where $\lambda = -\frac{1-\beta^T}{1-\beta}$. To do so we invert the utility function to get consumption as a function of utility

$$\bar{c}(u) = \frac{\lambda}{u}$$

We then use this formula to find the percent change in $\bar{c}$ that would come from a change in average utility equal to stigma. In doing so, $\lambda$ cancels out and we are left with the percent change in consumption as a function of average expected lifetime utility and stigma:

$$\% change = \frac{\bar{c}(u + stigma) - \bar{c}(u)}{\bar{c}(u)}$$

$$= \frac{\lambda}{u + stigma} - \frac{\lambda}{u}$$

$$= \frac{-stigma}{u + stigma}$$

Plugging in the average expected lifetime utility in the period when the mortgage is originated (-2.63), this formula replicates the Campbell and Cocco result ($stigma = -0.05 \implies % change = -1.9\%$) and gives $% change = -25.5\%$ at our estimated value of $stigma = -0.90$ that best fits the JPMCI data.

We therefore calculate that a borrower whose behavior matches the Campbell and Cocco model estimated to fit the JPMCI data would give up about $100,000 in consumption to avoid default. Ignoring bequests, non-housing consumption is equal to income minus housing expenditures. Non-housing consumption in the Campbell and Cocco model is income minus housing expenditures. Average income is $43,300 in Campbell and Cocco’s Table II panel (d) and housing expenditures are 40 percent of income at the time of default, so non-housing consumption is $26,000 per year. The consumption equivalent the borrower would give up is therefore $6,600 per year. Discounting at 3 percent and summing over 20 years, this gives a present value of $100,000.