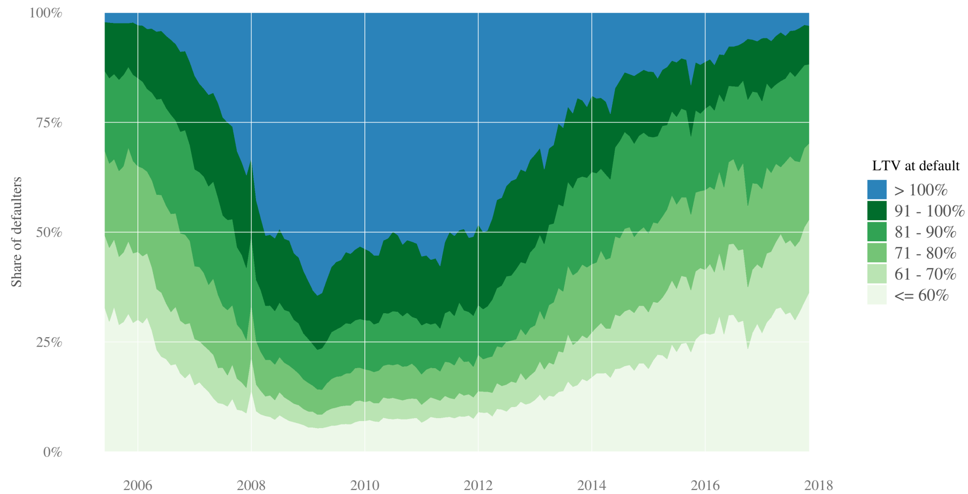
Online Appendix to "Why Do Borrowers Default on Mortgages? A New Method for Causal Attribution"

Peter Ganong and Pascal Noel

Contents

A Figures and tables

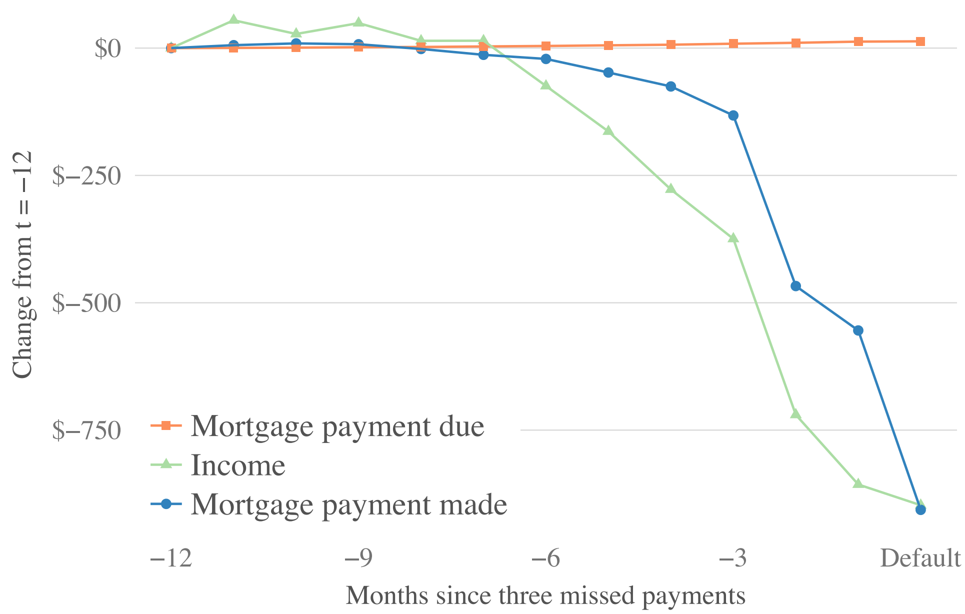
Figure 1: Prevalence of Above Water Mortgage Default



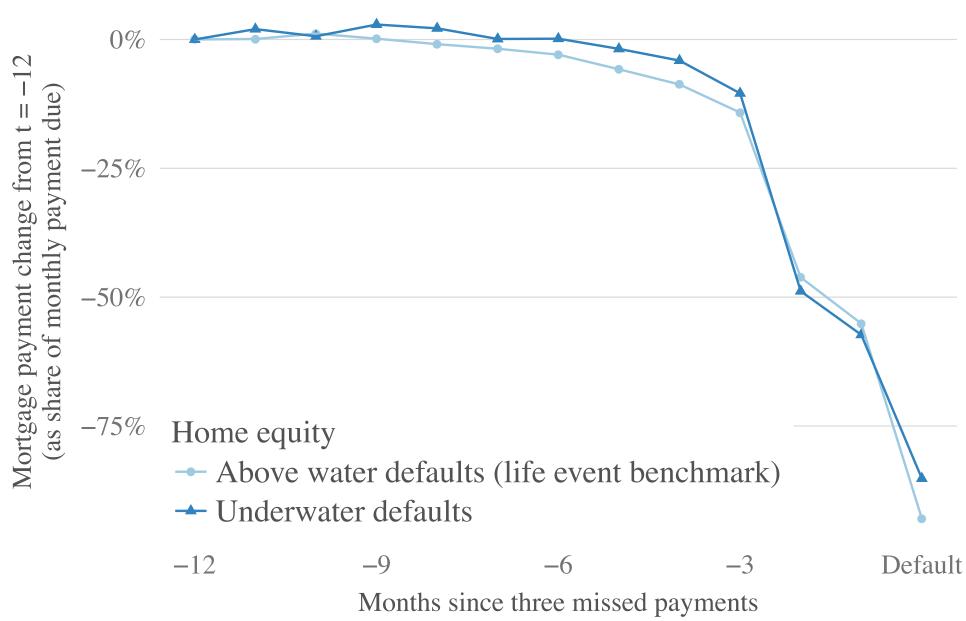
Notes: This figure shows the distribution of the loan-to-value (LTV) ratio at default in the Credit Risk Insight Servicing McDash (CRISM) data. Default is defined as three missed payments.

Figure 2: Payment Due And Payment Made Prior to Default

(a) Payment Due



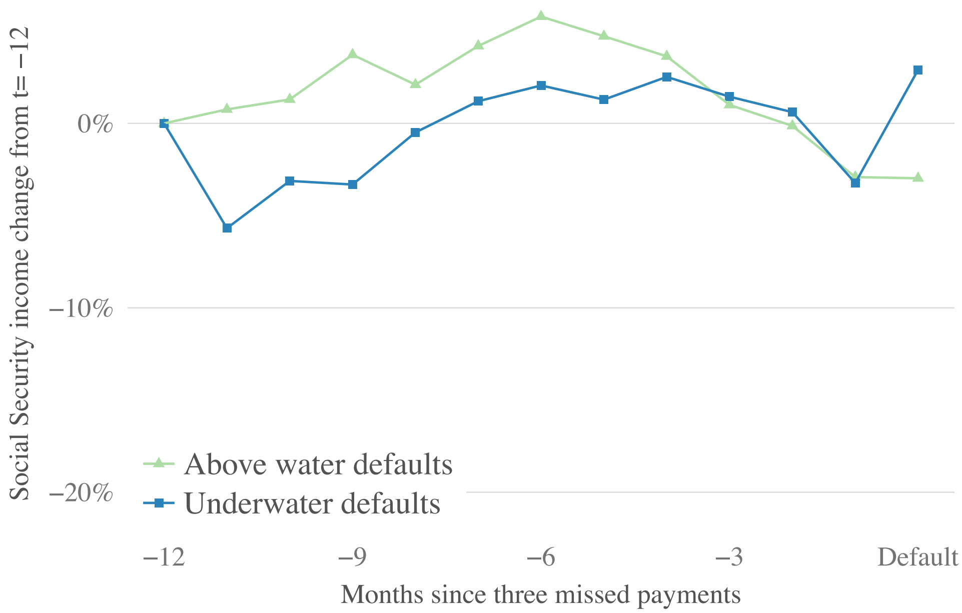
(b) Mortgage Payments by Home Equity



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.

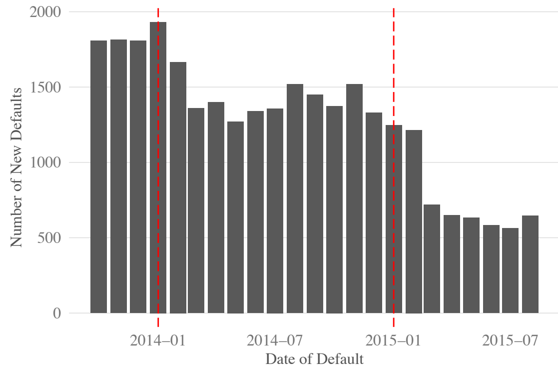
Figure 3: Social Security Income Change by Home Equity



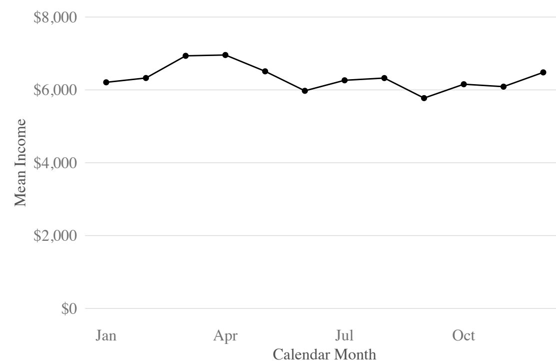
Notes: This figure shows the change in Social Security income in the year prior to default in the JPMCI data.

Figure 4: Seasonality in Mortgage Default

(a) Default Date Distribution



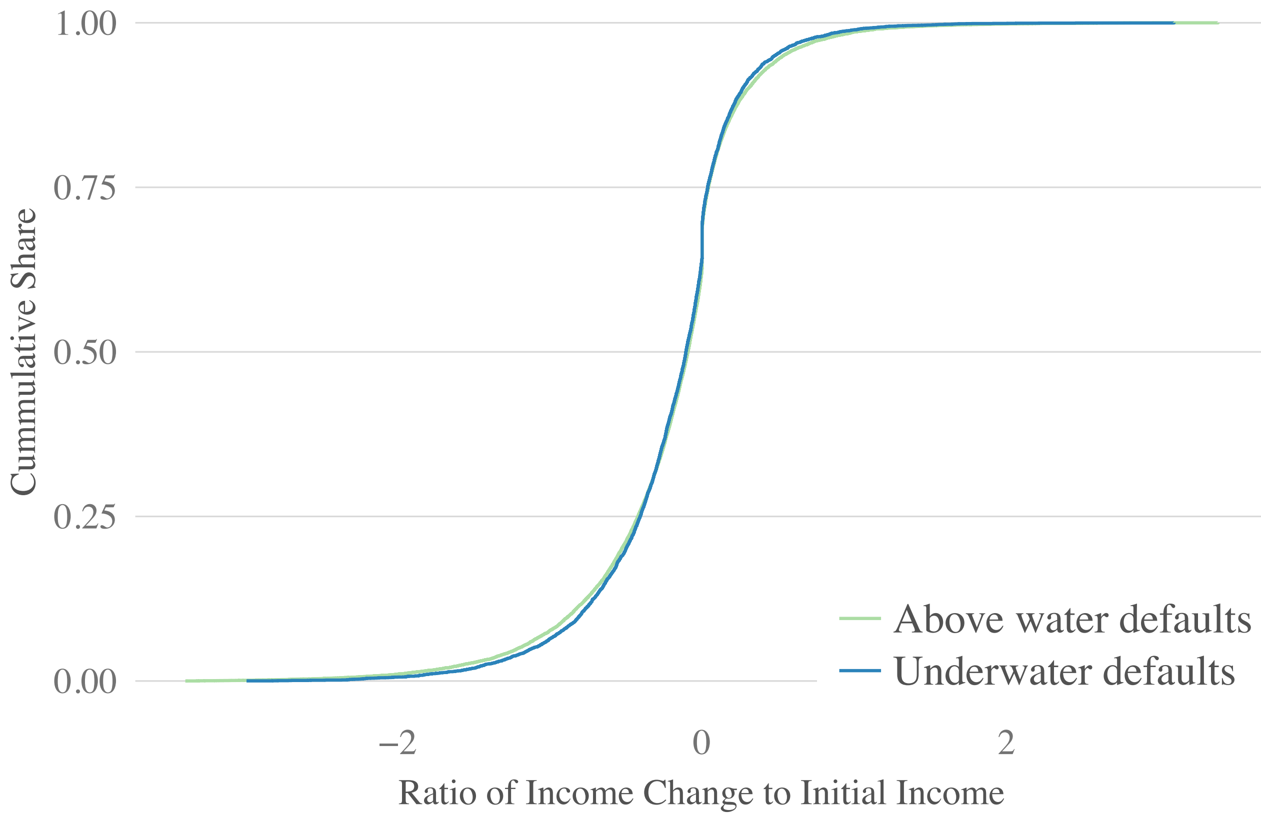
(b) Income by Calendar Month



(c) Effect of Seasonality on "All Underwater Borrowers" Series

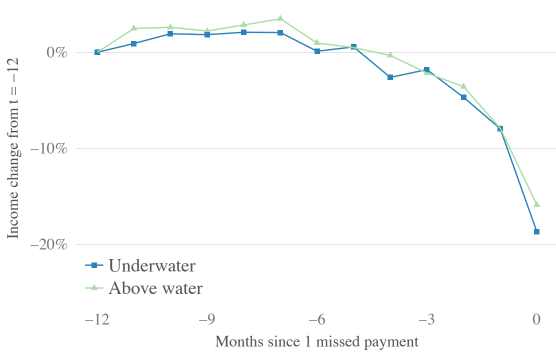
Notes: This figure documents that correlated seasonality in mortgage default and in income generates an S-shape in the "all underwater borrowers" series in Figure 1a. Panel (a) shows the distribution of default dates. In unreported results, we find similar seasonal patterns in broader samples in the JPMCI data and also in the CRISM data. Panel (b) shows mean post-tax income by calendar month. This pattern, too, also holds in broader samples in the JPMCI data. Panel (c) shows the role that correlated seasonality plays in generating the S-shape in the "all underwater borrowers" series. The purple circles replicate the series from Figure 1a. That figure uses equation (4), which indexes time as $t \in \{Oct2013, Nov2013 \ldots Aug2015\}$. The blue squares re-estimate equation (4), where $t \in \{Jan, Feb \ldots Dec\}$. This procedure isolates the effect of seasonal patterns on the "all underwater borrowers" series by capturing seasonality in defaults (shown in panel a) and seasonality in income (shown in panel b). Because defaults are highest at the end of each calendar year (panel a) and income is highest in March and April around tax refund season and December around Christmas (panel b), the "all underwater borrowers" series has peaks in the month of default and roughly eight months prior.

Figure 5: Distribution of Income Change Prior to Mortgage Default

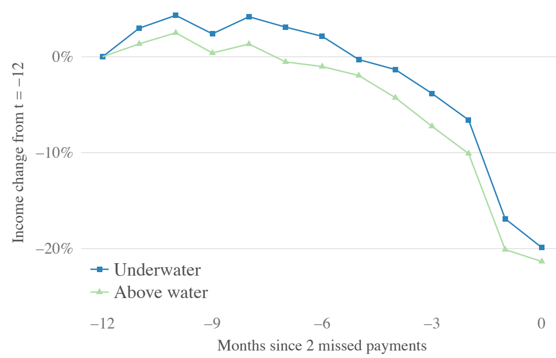


Notes: This figure shows the cumulative distribution function for the change in income, divided by average initial income. Average initial income is computed one year prior to mortgage default and is computed separately for underwater and above water borrowers. The ratio can be less than -1 or more than 1 since some households have income declines or increases that is larger than the average income level. This figure provides an alternative visualization of the histogram in Figure1b.

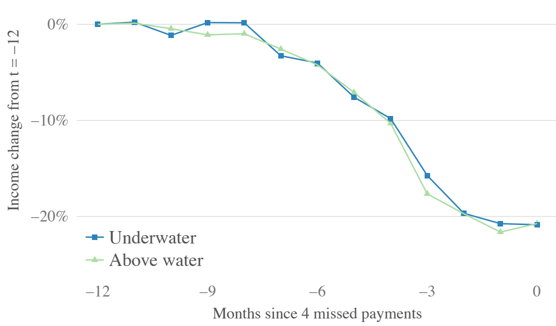
Figure 6: Income by Alternative Missed Payment Thresholds and Home Equity



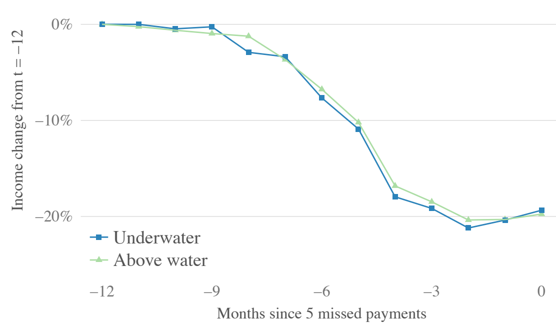
(a) One Month Past Due



(b) Two Months Past Due



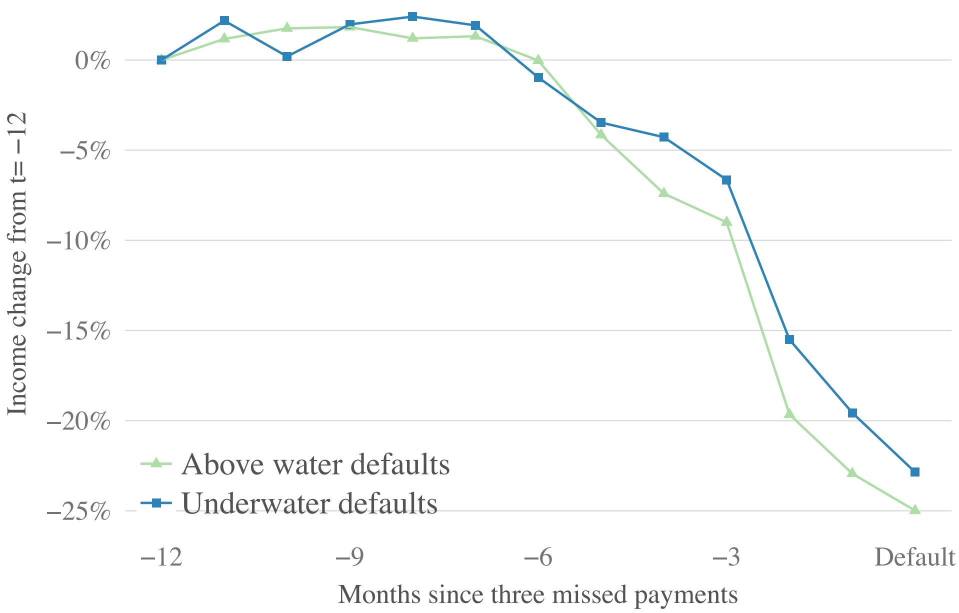
(c) Four Months Past Due



(d) Five Months Past Due

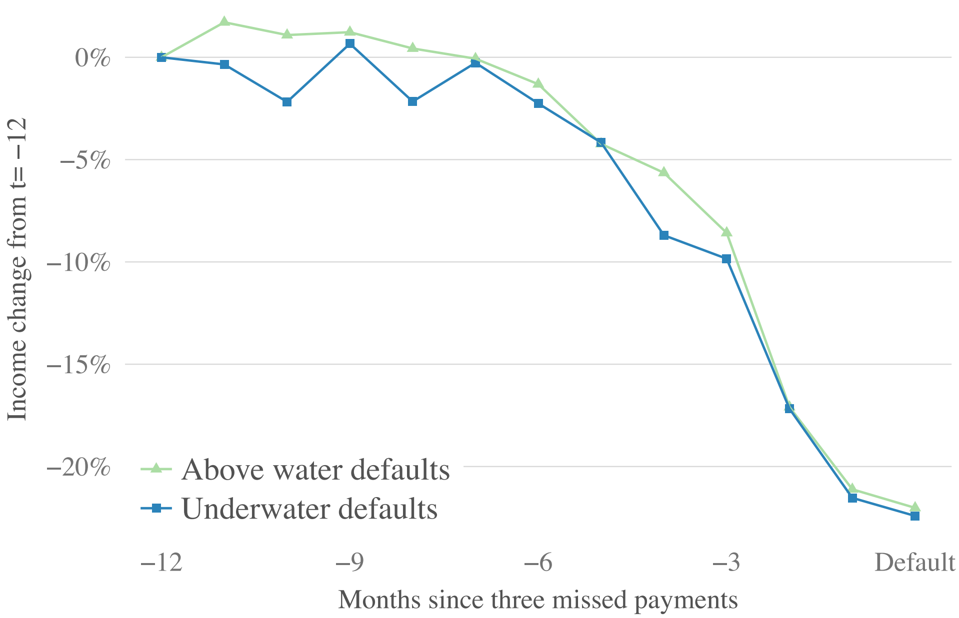
Notes: This figure replicates Figure 1a for alternative months past due thresholds.

Figure 7: Evolution of Income by Home Equity – Foreclosure Sample



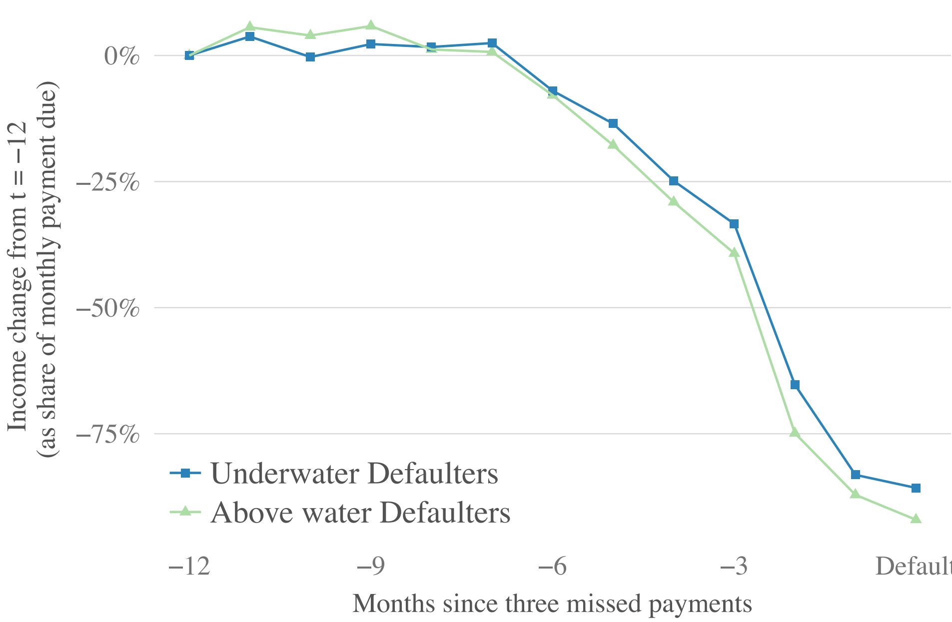
Notes: This figure replicates Figure 1a for the subset of defaulters who ultimately begin the foreclosure process.

Figure 8: Income Prior to Default by State-level Availability of Recourse



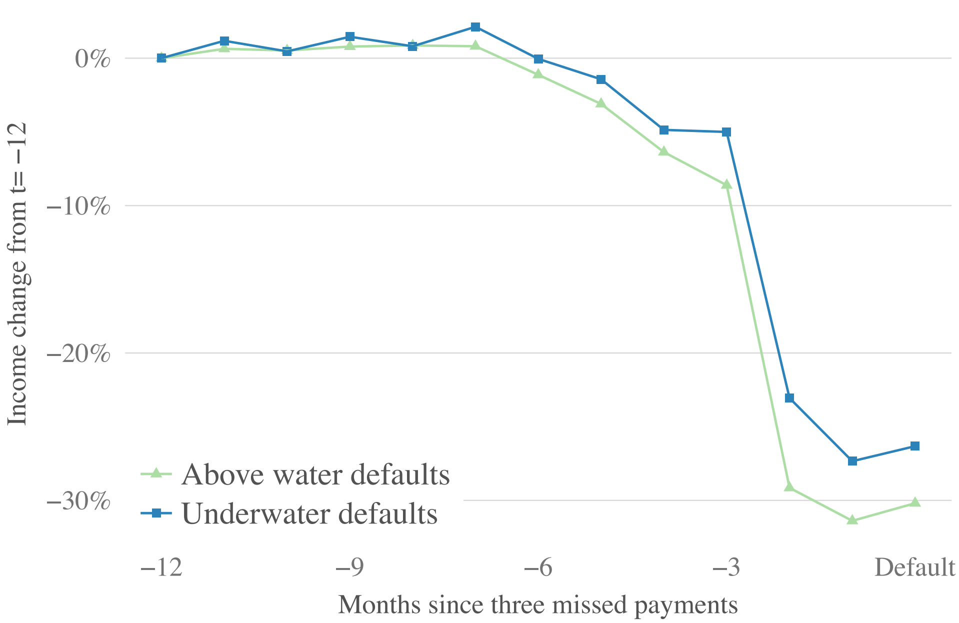
Notes: This figure replicates Figure 1a from the JPMCI data for the subset of states that do not allow mortgage lenders to sue to recover non-mortgage assets. We use the classification of non-recourse states from Ghent and Kudlyak (2011).

Figure 9: Income Change as Share of Payment Due by Home Equity



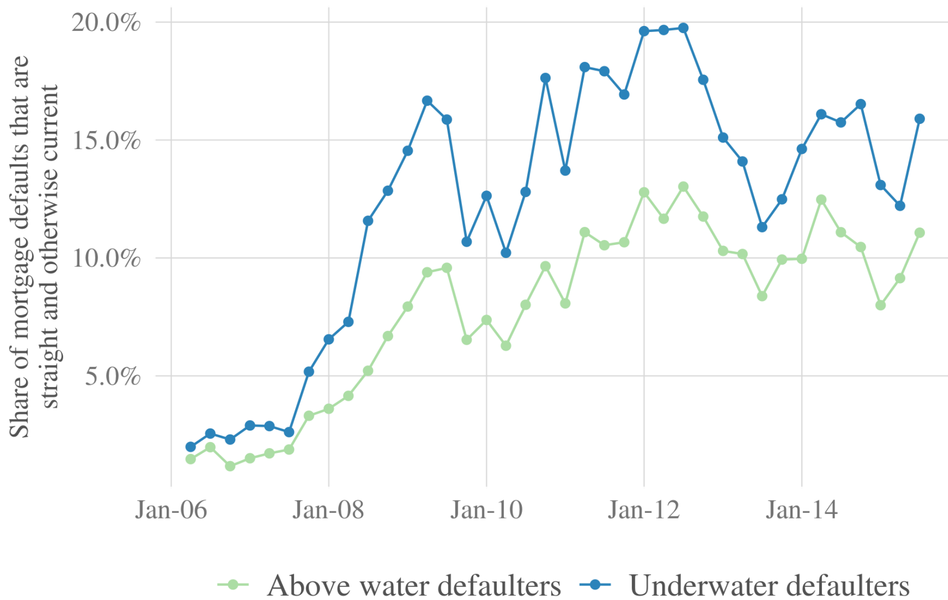
Notes: This figure replicates Figure 1a using a dependent variable of the change in monthly income divided by the average of the monthly mortgage payment due for months -12, -11 and -10 prior to default.

Figure 10: Income Prior to Default by Consecutive Missed Payments



Notes: This figure replicates Figure 1a from the JPMCI data for the subset of borrowers who miss three consecutive payments. Borrowers who miss three consecutive payments are 48 percent of underwater defaults and 39 percent of above water defaults.

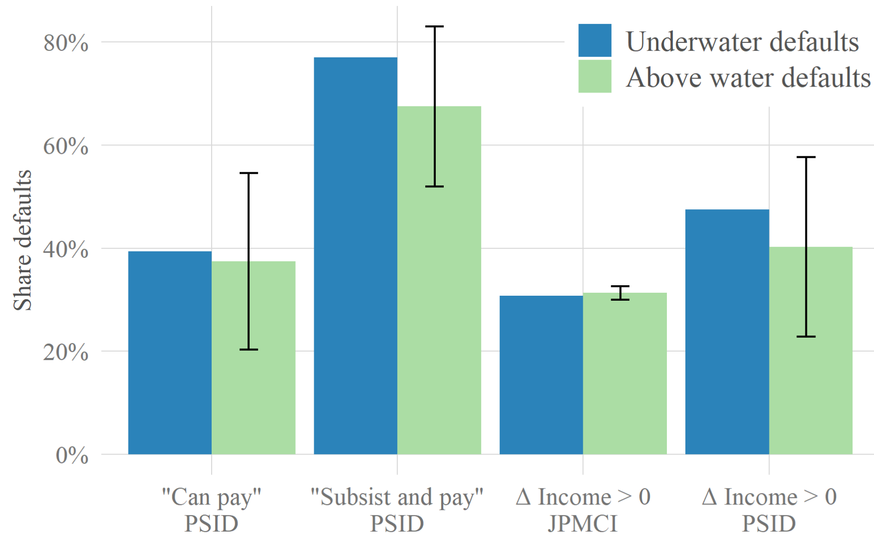
Figure 11: Share of Mortgage Defaults with Consecutive Missed Payments



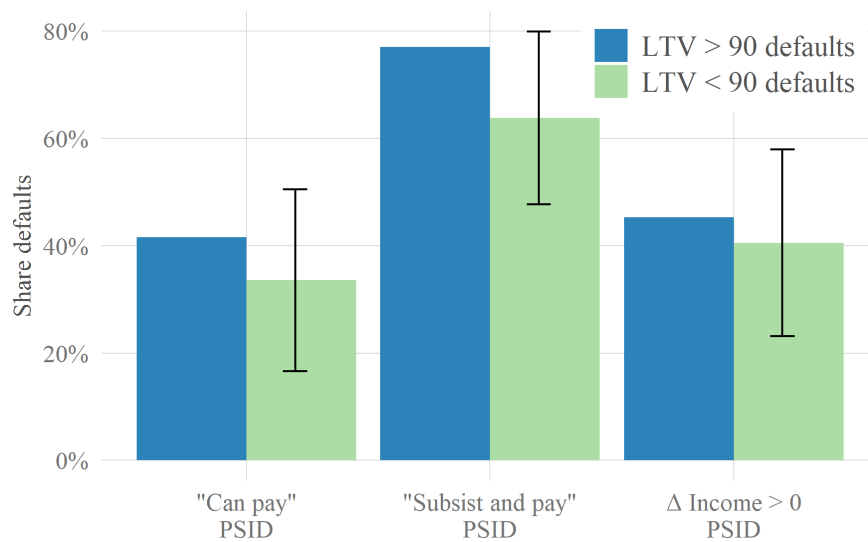
Notes: This figure extends the analysis in Keys et al. (2012) using the CRISM data. That paper measures the share of mortgage defaults that transition straight from 60 days past due to 180 days past due in four months, while remaining otherwise current on all non-HELOC revolving debt. We refer to such defaults as "straight and otherwise current". The average share of defaults that meet this definition is 19.6 percent of defaults for underwater borrowers and 12.3 percent of defaults for above water borrowers. Thus, the excess share of straight and otherwise current defaults for underwater borrowers is 7.3 percent.

Figure 12: Alternative Measures of Strategic Default

(a) Share of Defaults



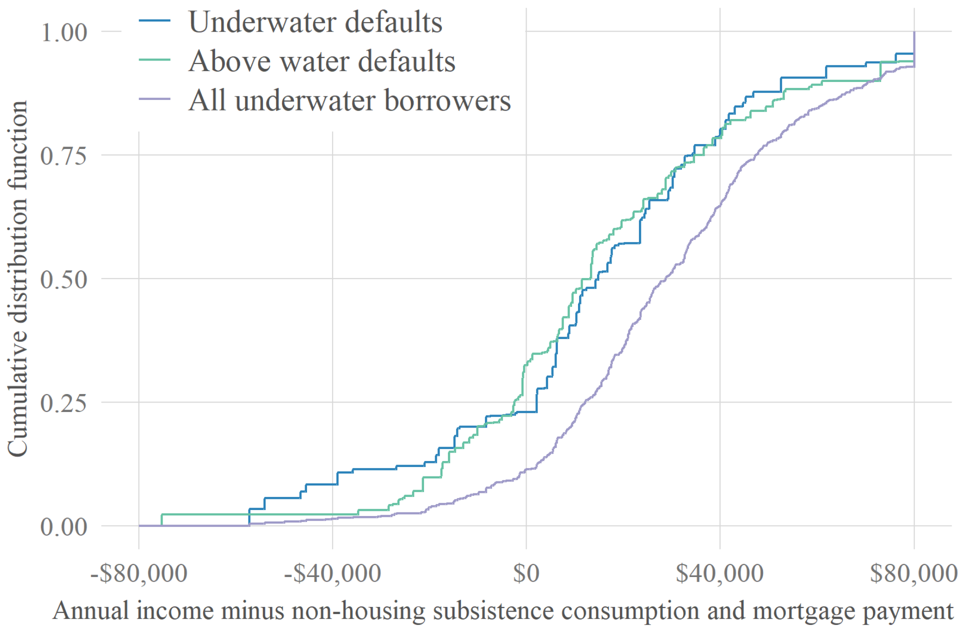
(b) Share of Defaults Using Loan-to-value (LTV) Cutoff of 90



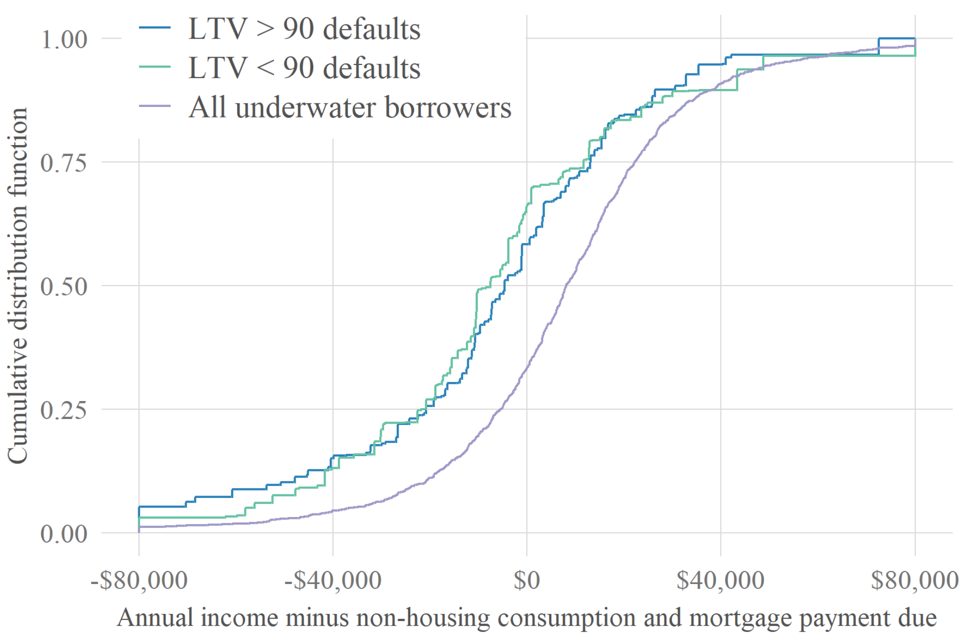
Notes: This figure compares the distribution of income by home equity and default status in the Panel Study of Income Dynamics (PSID) and the bank account data.Gerardi et al. (2018) measure mortgage affordability using income y, mortgage payment m, and non-housing consumption c. That paper classifies a borrower as *can-pay* if she can afford the mortgage without cutting consumption ($y - m - c_{predefault} > 0$) and as *subsist-and-pay* if she can afford a subsistence consumption level and pay her mortgage ($y - m - c_{subsistence} > 0$). See Section 5 for details on these definitions. Panel (a) reports the share of defaults that are classified as strategic using three different empirical criteria: can-pay, subsist-and-pay, and a positive change in income. The vertical lines indicate 95 percent confidence intervals for the difference in shares between above and underwater. 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 13: Alternative Measures of Strategic Default – Distributions

(a) Available Resources Using Subsistence Measure

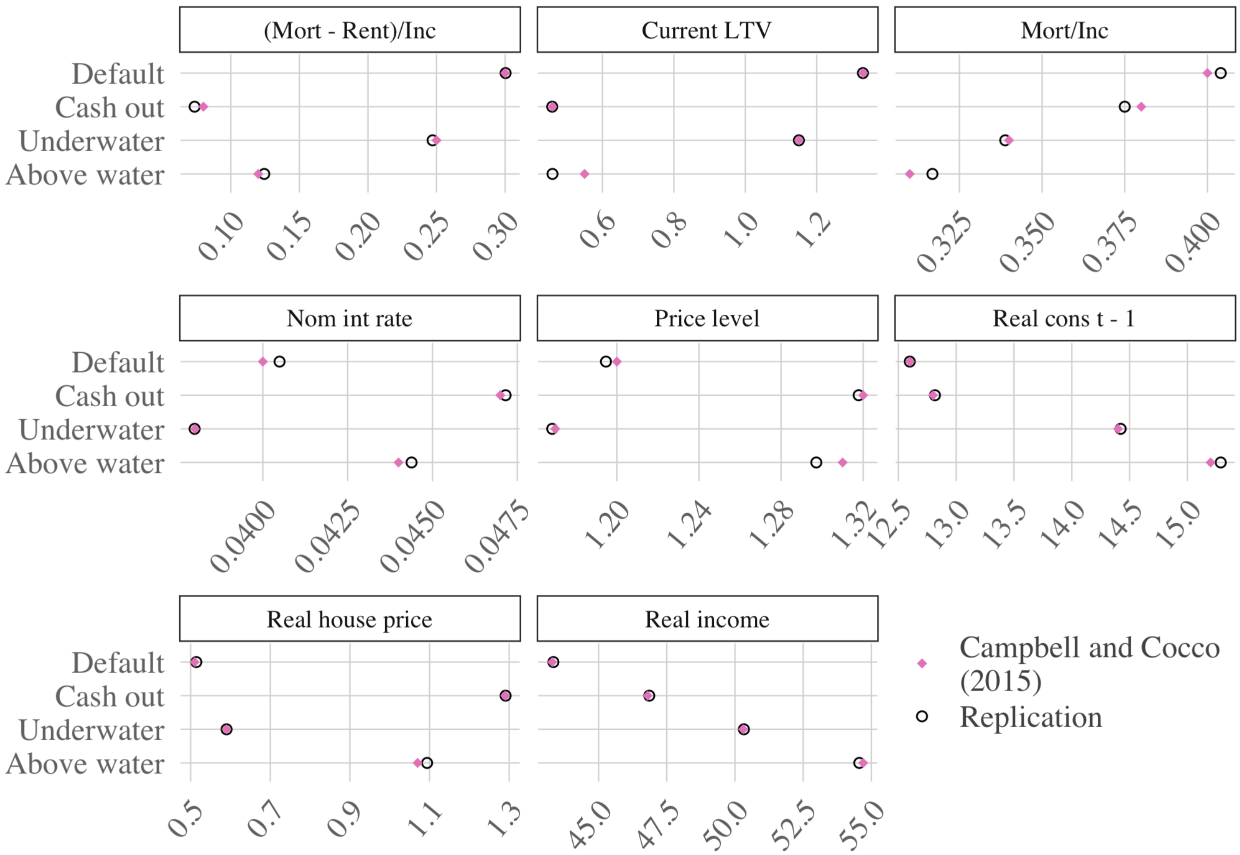


(b) Available Resources Using Loan-to-value (LTV) Cutoff of 90



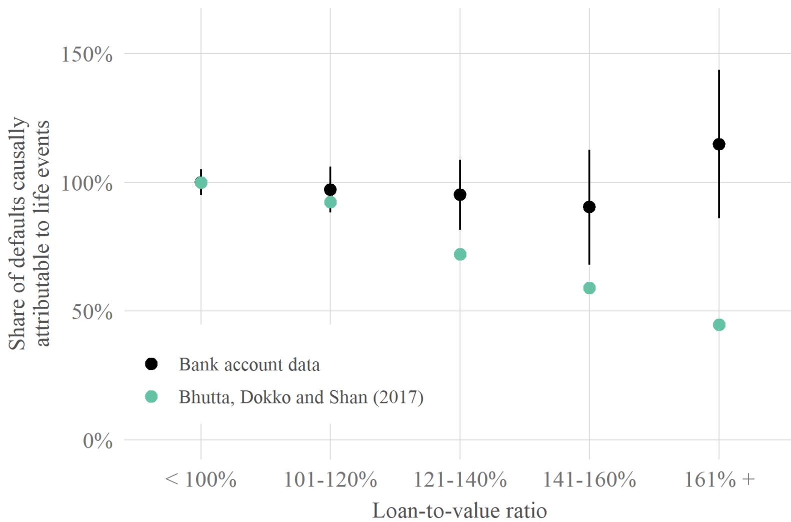
Notes: This figure reports two robustness checks on Figure 5, which uses $y - m - c_{predefault}$ as the x-variable and constructs home equity groups using an LTV cutoff of 100. Panel (a) uses an alternative x-variable $y - m - c_{subsistence}$, where $c_{subsistence}$ is a measure of the expenditure 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 in Gerardi et al. (2018). See Section 5 for details.

Figure 14: 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).

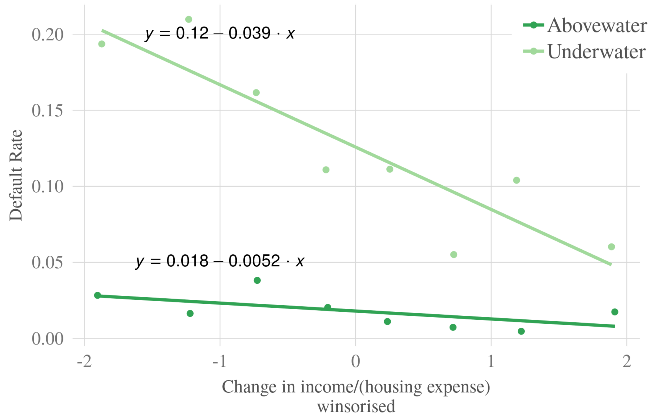
Figure 15: Defaults Causally Attributable to Life Events: Heterogeneity by Loan-to-value Ratio



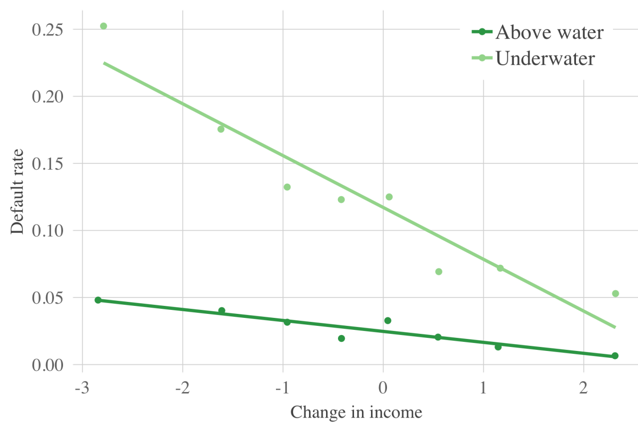
Notes: This figure compares estimates of the share of defaults that are causally attributable to life events in Bhutta, Dokko and Shan (2017) and in the JPMCI bank account data. The bank account estimates use equation (3) with data on the change in income prior to default. The Bhutta, Dokko and Shan (2017) estimates come from Figure 6 of that paper, where the proportion of defaults that are causally attributable to life events is 100 percent minus the share of defaults that are "strategic". This figure replicates Figure 4 using income data (whereas Figure 4 uses balance data).

Figure 16: Double Trigger Is Consistent with Zero Strategic Default

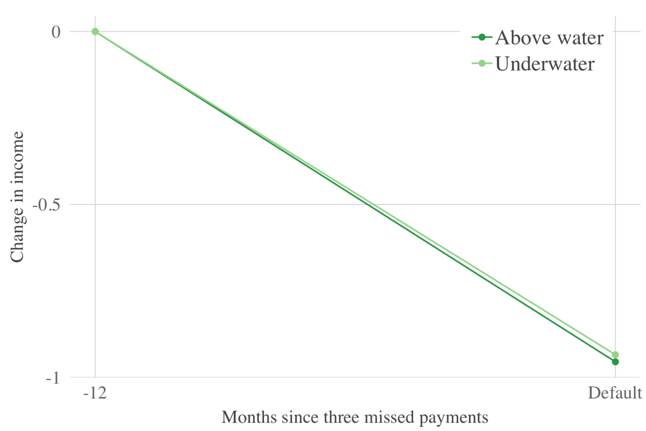
(a) Real Data: Regress Y on T



(b) Simulated Data: Regress Y on T



(c) Simulated Data: $E(T|Y = 1)$



Notes: The top panel computes binned default rates for 8 quantiles of the normalized change in income by home equity using the PSID. The probability of default is higher for underwater borrowers with a decline in income than it is for above water borrowers with a decline in income. We use a simulation to show that this finding is consistent with a finding of no strategic default. The middle and bottom panel show two different ways of visualizing the relationship between income, home equity, and default. Panel (b) replicates the analysis in panel (a), while panel (c) replicates the analysis in Figure 1a. See online Appendix D for details.

Table 1: Summary Statistics: Chase versus CRISM and McDash

Sample	Benchmark	Chase	CRISM	McDash
All mortgages	90 day delinquency rate	3.2%	3.3%	3.8%
All mortgages	Share investor	6.8%	3.9%	5.6%
All mortgages	Share primary occupant	89%	93%	91%
All mortgages	Share underwater	19%	22%	
Defaulters	Share investor	6.4%	4.3%	5.9%
Defaulters	Share primary occupant	90%	94%	92%
Defaulters	Share underwater	50%	58%	
Defaulters	Share of above water defaults with foreclosure within year	40%	55%	
Defaulters	Share of underwater defaults with foreclosure within year	45%	57%	

Notes: This table compares summary statistics regarding the matched mortgage-bank account dataset from Chase to the Credit Risk Insight Servicing McDash (CRISM) dataset in 2011. The CRISM dataset is constructed by linking credit bureau records from Equifax with mortgage servicing records from McDash. We use 2011 as the comparison year because this is the year when US 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 2: Income and Assets of Defaulters by Loan-to-Value

LTV	Drop as share of income	Drop as share of mortgage payment due	Checking Balance
<100	-0.203	-0.878	1,038
100-120	-0.198	-0.795	1,159
120-140	-0.195	-0.785	1,234
140-160	-0.185	-0.707	1,104
160+	-0.235	-0.953	1,080

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 2 in the main text, which describes borrowers one year before default.

Table 3: Distribution of Checking Account Balances of Defaulters

LTV	p10	p25	p50	p75	p90
Above water	-0.6	22.8	277.8	1,118.8	2,610.6
Underwater	0.4	31.4	358.1	1,301.4	2,939.5

Notes: This table shows the distribution of checking account balances 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. A negative balance indicates that an account is in overdraft status. Thus, this table shows that about 10 percent of above water and underwater defaulters have overdrafted their checking accounts. Note that this table describes borrowers at the date of default, which is different from Table 2 in the main text, which describes borrowers one year before default.

B Data appendix

To be included in the analysis sample, we require that the household have an open checking account from one year before default through the date of default.

The income data are available from October 2012 forward. Thus, for inclusion in the income analysis sample, we define the date of default as the first date *after October 2012* when a mortgage was 90 days past due.

The unit of observation is a mortgage. There are 29,034 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 28,589 unique households associated with these 29,034 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.

We apply the same logic to the checking account balance sample. Balance is measured at the beginning of the month. Inclusion in this sample requires having a checking account open in the year prior to the first default, where the date of default is the first default after January 2007.

C Proof of proposition 1

$$\begin{aligned}
\alpha_{G=1} &\equiv \frac{E(Y|G=1) - E(Y_{01}|G=1)}{E(Y|G=1)} \\
&= 1 - \frac{E(Y_{01}|G=1, T^*=0)}{E(Y|G=1)} \\
&= 1 - \frac{P(Y=1|T^*=0, G=1)}{P(Y=1|G=1)} \\
&= 1 - \frac{P(T^*=0|Y=1, G=1)}{P(T^*=0|G=1)}
\end{aligned} \tag{7}$$

where the first step uses assumption 2 (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

$$\begin{aligned}
E(T|Y=1, G=1) = {} & P(T^*=0|Y=1, G=1)E(T(0)|T^*=0, Y=1, G=1). \\
& + (1 - P(T^*=0|, Y=1, G=1))E(T(1)|T^*=1, Y=1, G=1)
\end{aligned}$$

where $T(T^*, G, Y) = T(T^*)$ from assumption 3a. Re-arranging terms gives:

$$\begin{aligned} P(T^* = 0|Y = 1, G = 1) &= \frac{E(T(1)|T^* = 1, Y = 1, G = 1) - E(T|Y = 1, G = 1)}{E(T(1)|T^* = 1, Y = 1, G = 1) - E(T(0)|T^* = 0, Y = 1, G = 1)} \\ &= \frac{E(T(1)) - E(T|Y = 1, G = 1)}{E(T(1)) - E(T(0))} \end{aligned} \tag{8}$$

where the second equality follows from assumption 3a. This object exists because $E(T(1)) - E(T(0)) \neq 0$ by assumption 3b. We can identify $E(T(1))$ because

$$\begin{aligned} E(T(1)) =& E(T|Y = 1, G = 0, T^* = 1)P(T^* = 1|Y = 1, G = 0) \\ =& E(T|Y = 1, G = 0) \end{aligned} \tag{9}$$

where $P(T^* = 1|Y = 1, G = 0) = 1$ by assumption 1.Substitute equation (9) into the numerator of equation (8) to get

$$P(T^* = 0|Y = 1, G = 1) = \frac{E(T|Y = 1, G = 0) - E(T|Y = 1, G = 1)}{E(T(1)) - E(T(0))} \tag{10}$$

This expression captures the numerator of the ratio in equation (7). Applying the same logic to the denominator in the ratio of equation (7) gives

$$\begin{aligned} P(T^* = 0|G = 1) &= \frac{E(T(1)|T^* = 1, G = 1) - E(T|G = 1)}{E(T(1)|T^* = 1, G = 1) - E(T(0)|T^* = 0, G = 1)} \\ &= \frac{E(T(1)) - E(T|G = 1)}{E(T(1)) - E(T(0))} \\ &= \frac{E(T|Y = 1, G = 0) - E(T|G = 1)}{E(T(1)) - E(T(0))} \end{aligned} \tag{11}$$

where $E(T|G = 1)$ includes both underwater defaulters and non-defaulters. We take the ratio of equations (10) and (11). 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 (7) gives

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

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

D Double-trigger simulations

Our empirical finding of no strategic default is consistent with this prior evidence in favor of the "double-trigger" theory of default. We use a simple simulation to illustrate this point. Because the model is highly stylized, our results should be taken more as suggestive illustrations of the economic forces at play rather than as an empirically-realistic model parameterization.

Define $Y(T^*, \eta)$ where Y is a binary variable for default, T^* is a binary random variable that is set to one when the household receives a life event shock and η is a second binary random variable (a preference shock) which determines whether the borrower defaults conditional on receiving a life event shock. We model the default decision as

$$Y(T^*, \eta) \equiv T^* \times \eta$$

so a borrower defaults only if they experience a life event and decide to default. Note that there is no "strategic default" in this model by borrowers who "can pay", to use the language of Gerardi et al. (2018).

We allow the preference shock to vary with household leverage, so the probability of default conditional on a life event can differ between above water and underwater borrowers. We assume that adverse life events T^* are measured with noise. Specifically, we assume the change in observed income T is the sum of T^* and mean-zero noise ε, where ε has a normal distribution with standard deviation σ ($T = -T^* + \varepsilon$, $\varepsilon \sim N(0, \sigma)$). We parameterize the model's three parameters as follows: $P(T^* = 1) = 0.25$, $\sigma^2 = 1.5$, and

$$P(\eta = 1) = \begin{cases} 0.1 & \text{if abovewater} \\ 0.5 & \text{if underwater} \end{cases}.$$

Using actual data, online Appendix Figure 16a shows that among borrowers with a positive income change (who should be able to pay if income is measured without error), the default rate is higher for borrowers who are underwater than for borrowers who are above water.[2]

Using simulated data, online Appendix Figure 16b plots the mean default rate against the eight bins of the change in income using the simulated data. It shows that the intercept is higher and the slope is steeper for underwater borrowers than for above water borrowers.

Finally, we apply the specification used in Section 2 to the simulated data to show that it is consistent with no strategic default and with our findings in the bank data. Online Appendix Figure 16c shows the change in average income among defaulters, separately for underwater and above water. The average change in income is the same for underwater defaults and above water defaults, just as we find in the bank data. Thus, we show that our finding that a life event is a necessary condition for default is consistent with prior double-trigger finds in the literature.

[2] This approximately replicates the Gerardi et al. (2018) Table 4 finding that among "can pay" borrowers, the default rate is higher for underwater borrowers than above water. Gerardi et al. (2018) Table 5 use *residual income*, which is "the difference between household resources and the mortgage payment exactly". We focus here on *income changes* for consistency with our specification in the JPMCI data.