# Essays in Household Finance and Bank Regulation

# Essays in Household Finance and Bank Regulation

A dissertation presented

by

Vijay Tupil Narasiman

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Economics

Harvard University

Cambridge, Massachusetts

July 2016

*Dissertation Advisors:*
**Professor David Scharfstein**
**Professor Jeremy Stein**

*Author:*
**Vijay Tupil Narasiman**

**Essays in Household Finance and Bank Regulation**

# Abstract

My dissertation focuses on topics in household finance and bank regulation. In chapter 1, I estimate the household consumption response to a predictable, quasi-permanent income shock. Credit card spending rises well before the positive shock occurs and then plateaus, suggesting that households are forward-looking and have enough liquidity to increase spending. This type of household behavior is found to be remarkably similar to the simulation of a modified buffer-stock model. The main conclusion is that households appear to be quite sophisticated in their consumption behavior, which has various policy implications.

In chapter 2 (joint with Divya Kirti), we present a model that describes how different types of bank regulation can affect the likelihood of fire sales in a crisis. There are three main results. First, the design of capital requirements affects whether fire sales can occur in the recapitalization process. Second, the interaction between capital and liquidity requirements causes banks to become larger and can also make fire sales more likely. Third, mandatory equity issuance can be a useful policy for limiting fire sales, but only if binding. Collectively, our findings suggest that bank regulation may have a strong effect on the likelihood of fire sales. In addition, time-varying risk weights may more effective than time-varying capital requirements in preventing fire sales.

In chapter 3 (joint with Todd Keister), we investigate whether policy makers should be permitted to bail out financial institutions during a financial crisis. We develop a model that incorporates two competing views about the causes of these crises: self-fulfilling shifts in investors' expectations and deteriorating economic fundamentals. We show that – in both cases – the desirability of allowing intervention depends on a tradeoff between incentives and insurance. If

policy makers can correct incentive distortions through regulation, then allowing intervention is always optimal. If regulation is imperfect and the risk-sharing benefit from intervention is absent, it is optimal to prohibit intervention. Our results show that it is possible to provide meaningful policy analysis without taking a stand on the contentious issue of whether financial crises are driven by expectations or fundamentals.

# Contents

# List of Tables

# List of Figures

# Acknowledgments

I thank my dissertation advisors David Scharfstein, Jeremy Stein, and Samuel Hanson for their invaluable guidance.

Chapter 1 of this dissertation was made possible by a data-use agreement with the JPMorgan Chase Institute (JPMCI), which has created anonymized data assets that are selectively available to be used for academic research. More information about JPMCI anonymized data assets and data privacy protocols are available at https://www.jpmorganchase.com/corporate/institute/data-privacy.htm. All statistics from JPMCI data, including medians, reflect cells with at least 10 observations. The opinions expressed are those of the author alone and do not represent the views of JPMorgan Chase & Co. While working on this paper, I was a paid contractor of JPMCI. Helpful comments on this paper were provided by Pascal Noel, Peter Ganong, Divya Kirti, Filippo Mezzanotti, and several people at JPMorgan Chase.

For Chapter 2, my co-author Divya Kirti is grateful for funding from the Macro Financial Modeling Group. We also thank Adi Sunderam, Paul Tucker, Ben Hebert, Jonathan Rhinesmith, Thomas Powers, Rohan Kekre, and Filippo Mezzanotti for helpful comments.

For Chapter 3, my co-author Todd Keister and I thank Huberto Ennis, Itay Goldstein, two anonymous referees, and seminar participants at CIDE, Cornell, Iowa State, Rutgers, the Society for Economic Dynamics Meetings, the Midwest Macroeconomics Meetings, and the Summer Workshop on Money, Banking, Payments and Finance at the Federal Reserve Bank of Chicago for useful comments.

To my parents Tupil and Ranjana and my wife Preethi

# Chapter 1

# What do adjustable-rate mortgage resets say about household consumption behavior and the income channel of monetary policy?[1]

## 1.1 Introduction

How does household spending respond to predictable income shocks? The answer to this question is related to a number of issues that are fundamental to household finance, such as the level of sophistication of households, their ability to anticipate and act in advance of shocks, and the extent to which they are liquidity constrained. More generally, this question is related to how households make dynamic consumption decisions in the face of uncertainty. Are they relatively unsophisticated, hand-to-mouth consumers? Or do they exhibit buffer-stock behavior?

---

[1]This research was made possible by a data-use agreement with the JPMorgan Chase Institute (JPMCI), which has created anonymized data assets that are selectively available to be used for academic research. More information about JPMCI anonymized data assets and data privacy protocols are available at https://www.jpmorganchase.com/corporate/institute/data-privacy.htm. All statistics from JPMCI data, including medians, reflect cells with at least 10 observations. The opinions expressed are those of the author alone and do not represent the views of JPMorgan Chase & Co. While working on this paper, I was a paid contractor of JPMCI.

Prior work offers somewhat conflicting evidence on this issue, particularly for "poor" households with low income or wealth. For example, Hsieh (2003) and Browning and Collado (2001) find households to be quite sophisticated in anticipating and smoothing non-durable consumption around large, recurring income shocks. Hsieh (2003) finds this to be true even for low-income households. In contrast, Parker *et al.* (2013) find that spending by poor households on large-ticket durable goods is highly responsive to the timing of predictable tax rebates. The same result is found with non-durable goods as well (Souleles (1999), Johnson *et al.* (2006), Parker (2015)), suggesting that households either do not foresee shocks or do not have sufficient liquidity to increase spending in advance of them. In this paper, which is similar to Di Maggio *et al.* (2015) and Keys *et al.* (2014), I attempt to bring further clarity to this issue by investigating how households respond to a "quasi-permanent" income shock that is known to occur at a certain date in the future but is uncertain in magnitude.

Due to the nature of the income shock I study, this paper also speaks to the effectiveness of the income channel of monetary policy, whereby a reduction in market interest rates raises the disposable income of households in a "quasi-permanent" fashion. There are two ways this channel can operate. First, lower market rates allow fixed-rate mortgage holders to refinance and lower their monthly payments. Since these mortgages generally carry 30-year terms, the reduction in payments extends over a fairly long horizon and provides a quasi-permanent increase in household disposable income. Second, lower market rates cause the monthly payments of floating-rate mortgage holders to fall automatically. This also has a quasi-permanent effect on disposable income to the extent that a reduction in current rates lowers the expected interest rates (and therefore the expected mortgage payments) in future months as well.

The general challenge with identifying the effect of an income shock on household spending is finding a suitable control group. For example, with respect to the income channel of monetary policy, a suitable control should not experience a reduction in mortgage payments when interest rates fall but should be otherwise similar to households that do. The problem is that households who experience a reduction in payments tend to be either fixed-rate mortgagees who are able to refinance or floating-rate mortgagees. In contrast, households who do not experience a reduction in payments tend to be fixed-rate mortgagees who are unable to refinance. It certainly seems plausible that the ability to refinance or the choice to have a floating-rate mortgage is correlated with many outcomes of interest, such as household spending. If so, the assumptions required for identification are not met.

Fuster and Willen (2013) were the first to address this issue by studying hybrid adjustable-rate mortgage (ARM) resets. A hybrid ARM is a 30-year mortgage that has an initial term during which monthly payments are fixed. After this fixed term expires, the mortgage "resets", meaning that monthly payments go from being fixed to being linked to the current prevailing interest rate. The length of the initial fixed term is offered in various increments (5 years, 7 years, 10 years, etc) and is chosen at origination.

Hybrid ARMs offer a nice environment for identifying the effects of income shocks. Suppose that when market interest rates are high, two households ($A$ and $B$) originate 5-year fixed term hybrid ARMs in close temporal proximity to each other ($A$ being first). Over the next 5 years, suppose that market interest rates fall dramatically. When household $A$'s mortgage resets, its disposable income rises as its mortgage payment becomes linked to the low prevailing market rate. However, household $B$'s mortgage has not yet reset - it continues paying the same fixed amount for some time. Therefore, the spending of household $B$ is a potential counterfactual for the spending of household $A$ around the latter's reset. The required assumption is that in the absence of household $A$'s reset, $A$ and $B$ would have parallel trends in spending. This certainly seems plausible since both households chose to originate the exact same type of mortgage in the same high interest rate environment, just at different times.

Di Maggio *et al.* (2015) and Keys *et al.* (2014) employ this identification strategy in an event study framework to estimate the effects of income shocks on household consumption. Specifically, they use 5-year hybrid ARMs that are originated in a high interest rate environment (2004-2008) and that reset when market rates are at the zero lower bound (2009-2013). They find strong effects of income shocks on auto spending, their principal measure of household consumption. In the month that a household's mortgage payment falls, there is a sharp increase in auto spending that is sustained for several months. This effect is concentrated in low income households and is weak, if not negligible, in the months leading up to reset. The authors conclude from these results that the income channel of monetary policy is strong, particularly for low-income households, and that these households lack either the foresight or liquidity to increase auto spending in advance of their shocks.

However, this evidence by itself does not provide a complete picture of household behavior or by extension, the income channel. As documented in Di Maggio *et al.* (2015), average monthly household spending on autos is only $300, a small portion of total consumption. In addition, auto purchases are very infrequent: the probability of purchasing a car in a month is less than 1.5%. Therefore, the response of auto spending to income shocks may not represent how broader forms of consumption, namely non-durable goods, respond to the same kinds of shocks.

In this paper, using a methodology similar to Di Maggio *et al.* (2015), I conduct an event study of hybrid ARM borrowers in a 2-year window around reset using anonymized data from JPMorgan Chase (hereafter referred to as "the Bank"). The main innovation compared to prior work is the use of an important and previously unexplored consumption variable: credit card spending.[2] Its importance as a consumption variable is demonstrated by median monthly household credit card spending in my sample being $2400, significantly more than average auto purchases in prior work ($300). In addition, the set of goods purchased with credit card offers a broader perspective on consumption than autos alone.

A second innovation of this paper is the splitting of ARM borrowers into different groups based on how mortgages are amortized before and after reset. While all groups experience a reset, differences in amortization cause some groups to experience much larger reductions in mortgage

[2]Di Maggio *et al.* (2015) and Keys *et al.* (2014) use a measure of credit card spending calculated from outstanding credit card balances and payments and find similar results compared to their main consumption variable of auto spending. While both balances and payments are made available in credit bureau data, banks are not required to report either one to the credit reporting agencies. The Fair Credit Reporting Act only places requirements on banks once they voluntarily choose to provide information (see 15 U.S.C §1681s-2). Therefore, it is possible that credit card spending imputed from credit bureau data is unreliable. The stark differences between the credit card spending results of this paper and those of Di Maggio *et al.* (2015) and Keys *et al.* (2014) provide some evidence in favor of this view.

payments than others. In prior work, just one of these groups is studied in isolation (Di Maggio *et al.* (2015)) or they are all lumped together (Keys *et al.* (2014)). In this paper, I analyze each of these groups separately, allowing for a deeper understanding of household behavior and the income channel.

The first main result of the event study is that the dynamics of credit card spending in response to ARM reset are different from that of auto purchases. For households who experience large payment reductions at reset, credit card spending gradually increases in the months leading up to reset, with no sharp change in the month of reset itself. This suggests that not only do households fully anticipate their resets, they also have enough liquidity to increasing spending before the positive shock occurs. This is quite different from the results on auto spending in Di Maggio *et al.* (2015), where there is a very small increase in spending pre-reset and a sharp increase in the month of reset itself.

The second result is that for the set of ARM borrowers who happen to experience a negligible payment reduction at reset, there is not a statistically significant increase in spending either before or at reset. This is important because it represents a sort of "placebo test" of the previous result. However, based on point estimates for these households, the credit card spending profile actually follows a similar pattern as in the previous result, just with much smaller magnitudes. Spending gradually rises pre-reset, with no sharp change in the month of reset itself. This latter result is puzzling because these households experience only a negligible payment reduction at reset, yet still seem to increase credit card spending leading up to reset.

The third result explores how household wealth affects the response of credit card spending to income shocks. Prior work finds the effect to be concentrated in households with low income and attributes this to tighter liquidity constraints for this group. I find that for households who experience a large payment reduction at reset, there is a similar concentration of the spending response in households with below-median wealth, as measured by liquid assets holdings. However, these spending responses still occur before payment reductions happen at reset. In addition, for households who experience a negligible payment reduction at reset, the spending response is concentrated in households with above-median wealth. These observations seem to be at odds with a pure liquidity constraints story.

In the last portion of the paper, I turn to theory and ask whether there is a model of household consumption that is consistent with the main empirical results: (1) no sharp increase in spending at reset, (2) a gradual increase in spending pre-reset, (3) a similar pattern with much smaller magnitudes for households whose payment reductions at reset are negligible, and (4) heterogeneous responses by household wealth that flip based on the whether the payment reduction at reset is large or negligible.

I start with a standard buffer-stock model of consumption with labor income uncertainty, as in Deaton (1999), and incorporate elements to mimic the environment of the empirical event study. For example, I require agents to make mortgage payments that match the profile of a 5-year hybrid ARM and "start" the model 1 year before mortgage reset. Before reset, payments are fixed at a high value and after reset, payments become a function of the prevailing market rate in the reset month. Note that the current market interest rate will be a key state variable in the pre-reset window because it guides agents as to what the prevailing market rate at reset, and therefore what agents' mortgage payments in the post-reset window, will be.

To assess whether this model comports with the results of the empirical event study, I solve the model numerically and simulate it for a specific path of market interest rates in a 2-year window around the reset month. To emulate the event study, the simulated path of rates is set to reflect what ARM borrowers in my sample actually faced in proximity of their resets. Since the resets in question occur after 2009 when market interest rates were consistently at the zero lower bound (ZLB), the simulated path of rates is set to be low and stable.

The final result of this paper is that the simulated consumption profiles of agents in a modified buffer-stock model are strikingly consistent with the empirically estimated credit card spending profiles described earlier. Thus, the modified buffer-stock model provides a fairly good description of how household credit card spending responds to permanent income shocks. A key property of the model is that agents with high levels of liquid assets act like permanent income consumers while agents with low levels of assets act primarily out of a concern for maintaining an adequate buffer against labor income shocks. It is this dichotomy that causes the simulation to produce the same type of heterogeneous effects by wealth that the empirical event study exhibits.

Two main conclusions are drawn from the collective results of this paper. The first is related to household behavior. Consistent with Hsieh (2003) and Browning and Collado (2001), households appear to display a high degree of sophistication and forward-looking behavior. Not only do they seem to anticipate their resets, their actions in advance of reset match those of a dynamically optimizing buffer-stock agent. This conclusion is most striking for low-asset households, who gradually adjust their spending upward as their resets become more imminent. Such behavior reflects a keen awareness of one's current and future financial situation and demonstrates that even low-asset households have some liquidity to draw on.

How can this result be reconciled with other evidence that households lack sophistication, liquidity, or both? In the case of the evidence on durable goods, it is plausible that poor households, even if they have some liquidity, just do not have enough to purchase large-ticket durables until a shock actually occurs. In the case of the evidence on non-durable goods, Hsieh (1999) and Browning and Collado (2001) claim that since the income shocks in these cases are relatively small, transient, and irregular, households with bounded rationality simply ignore them until they occur, and therefore appear unsophisticated.

However, the results of this paper point to a different explanation. If poor households are indeed buffer-stock agents with a small amount of liquidity, their consumption decisions at any given time will be tightly linked to their perceived exposure to negative labor income shocks going forward. Since a tax rebate is small, transient, and irregular, it cannot have a very lasting impact on a household's ability to weather labor income shocks. Instead of drawing on their liquidity to consume in advance of the shock, poor households would rather retain whatever liquidity they have as a buffer. The decision to hold off on spending until the shock occurs is therefore a rational one. Importantly, the same logic does not apply to shocks that are large, permanent (like ARM resets), or recurring, because these shocks do have a lasting impact on poor households' exposure to negative labor income shocks.

The second main conclusion of the paper is related to monetary policy. The conclusion is that event studies of ARM resets may not be useful for estimating the effects of monetary policy surprises. Given the evidence that households seemed to fully anticipate their resets, the observed spending responses and their relationships to household wealth could be different if the shock is unanticipated. Indeed, in a buffer-stock model, the spending of wealthy households will be more responsive to surprise income shocks since poorer households will use some of the windfall to bolster their asset buffers.

Nevertheless, the analysis of ARM resets in this paper is useful for understanding whether monetary policy can affect current spending by changing household beliefs about future interest rates, holding current rates constant. Both the empirical and theoretical evidence point to wealthy households adjusting their current consumption in response to changes in expected permanent income. If this is true in practice, the ability to guide household beliefs about future interest rates is a powerful monetary policy tool. Specifically, the income channel can operate through an "expectations effect" for wealthy households: the Fed causes expected future rates to fall, households' expected future disposable income rises, and their current consumption rises too.

The rest of the paper is organized as follows. In Section 1.2, I provide an overview of the anonymized data used in the event study and present summary statistics. In Section 1.3, I outline the empirical methodology and provide justification for the identification strategy. In Section 1.4, I present the results of the event study. In Section 1.5, I formally develop the modified buffer stock model and show how it is solved numerically. In section 1.6, I simulate the modified buffer-stock model and compare the results to the event study. Finally, I summarize all of the results and conclude in section 2.5.

## 1.2 Data and summary statistics

The data used in the event study are anonymized and can be split into three categories: mortgages, credit cards, and demographics.

### 1.2.1 Mortgage data

The mortgage data are a monthly panel of customer accounts from April 2009 to present for the universe of mortgages that the Bank services. The anonymized data contains information on the the interest rate, loan amount, term, and home value estimate of every account at origination and as a time series. For hybrid ARMs, the lengths of the initial fixed terms are not given in the data by default and have to be inferred from changes in the interest rate variable.

This paper focuses on 5-year fixed term hybrid ARMs that originate before December 2007 and reset after April 2010. Based on the 5-year fixed term, these dates imply an origination window of April 2005-December 2007 and a reset window of April 2010-December 2012. The rationale for these windows is threefold. First, they ensure that households can experience large payment reductions upon reset. As seen in Figure 1.1, average market interest rates during the origination window (between the first two dotted lines) are approximately 5%, substantially higher than they are during the reset window (between the second two dotted lines). Second, neither the origination nor reset windows overlap with the Great Recession (December 2007 to June 2009), which is desirable from an external validity point of view. Finally, starting the reset window in April 2010 (as opposed to earlier) is necessary for there to be at least one full year of data available before reset, since the data from the Bank begins in April 2009.

A few additional restrictions are also imposed on the mortgage sample. First, mortgage accounts that are modified in any way (term, principal, rate, etc) prior to reset are excluded. In addition, mortgages that are refinanced at any point in their lives are also excluded. These restrictions are imposed because modifications and refinancings can introduce additional shocks and change the nature of the shocks that households experience at reset. They are also endogenous decisions that may be correlated with trend of household spending.

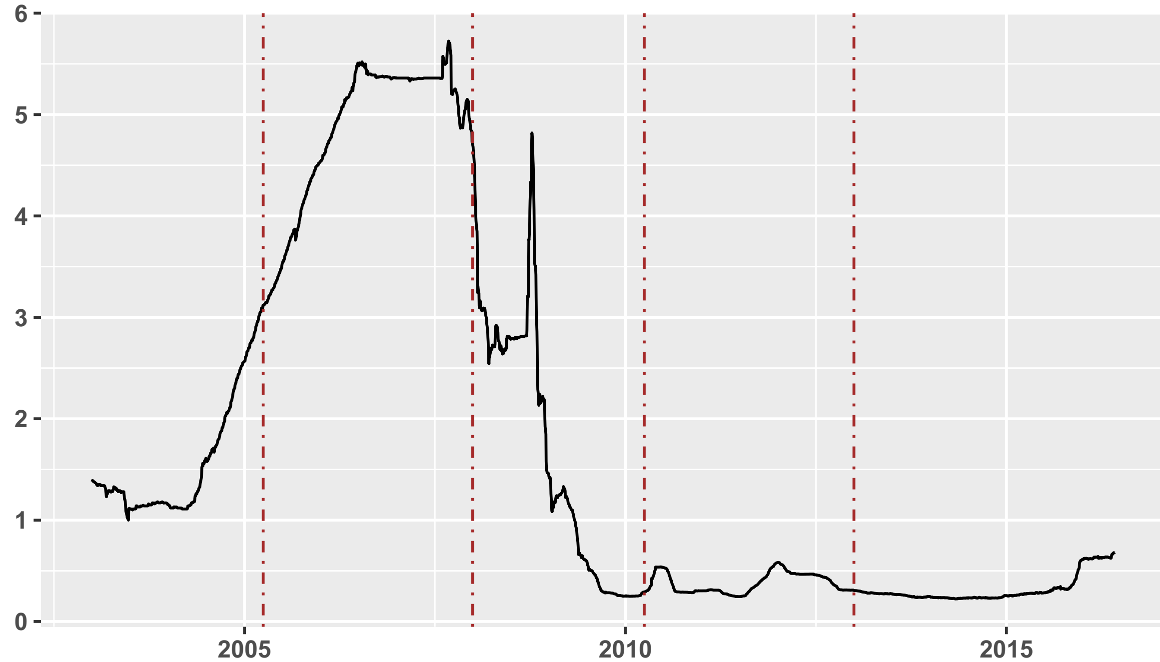



**Figure 1.1:** *USD 3-month LIBOR*

In any given month, the mortgages studied in this paper have either zero or positive amortization, where the former implies that payments are interest-only and the latter implies payments are interest-plus-principal.[3] As indicated in Section 1.1, one of the differences between this paper and prior work is dividing ARM borrowers into three groups based on how amortization changes before and after reset. The first group is labeled "zero-zero" and contains households with zero-amortization (interest-only payments) before and after reset. The second group is "positive-positive" and contains households with positive-amortization (interest-plus-principal payments) before and after reset. The last group is "zero-positive" and contains households with zero-amortization (interest-only payments) before reset and positive-amortization (interest-plus-principal payments) after reset.[4]

[3]Negative amortization mortgages are excluded from the sample because they are relatively few in number.

[4]Each household's amortization group is not specified in the data and must be inferred from how observed mortgage payments change at reset.

While all three amortization groups experience a reset in the interest rates on which their mortgage payments are based, the effect of reset on actual mortgage payments will differ by group. This is the primary motivation behind analyzing these groups separately. Each group effectively experiences a different "treatment", making comparisons across these groups useful for understanding household behavior. An additional reason to analyze these groups separately is that households choose their amortization schedules. If this choice is correlated with the trend of household spending, the groups should be analyzed separately for the purpose of identification.

### 1.2.2 Credit card data

The credit card data are also a monthly panel of customer accounts from April 2009 to present for all households that have ever had a credit card account with the Bank. The anonymized data contains information on total spending, revolving balances, and credit limits. There are three important benefits associated with this data compared to credit bureau records, the primary credit card-related data source in prior work. First, the Bank spending data is built up from actual credit card purchases and does not need to be derived imperfectly from outstanding credit card balances. Second, the Bank spending data can be disaggregated into a broad range of categories. Third, the Bank data on revolving balances, which is the portion of outstanding credit card balances on which interest is paid, is not available in credit bureau records.

A few additional restrictions are imposed on the credit card data as well. First, I require that in the 2-year window around each household's ARM reset month, the median number of monthly transactions on the card exceeds 10. This excludes households whose credit card spending at the Bank is not sufficiently active for it to be representative of their general consumption. Second, ARM households must have held their credit card account at the Bank for at least 2 years prior to their reset dates. This filter is imposed so that the results of the event study are not contaminated by households who just recently opened a credit card account in order to make a large purchase.

### 1.2.3 Demographics data

The Bank has demographic data on the age, zip code, annual income, and liquid asset holdings of all of its customers. Annual income is the Bank's estimate of an account holder's gross annual income. Similarly, liquid asset holdings is the Bank's estimate of an account holder's total liquid assets (cash, money market accounts, etc), not just those held at the Bank. In this paper, I use the liquid asset estimate rather than the income estimate as a proxy for household liquidity, as the former is more related to the concept of cash on-hand. Therefore, when the terms "wealthy" and "poor" are used hereafter, I use them in reference to a household's liquid asset estimate. It should be noted that since these numbers are just estimates, it is better to interpret them as indications of ordinality (relative ordering) rather than cardinality (actual magnitude).

### 1.2.4 Summary statistics

Tables 1.1 and 1.2 show how certain mortgage and demographic-related summary statistics are affected by the various sample restrictions discussed in Sections 1.2.1 and 1.2.2. Table 1.1 contains means and Table 1.2 contains medians.[5]

**Table 1.1:** *Evaluation of sample restrictions (means)*

| sample | n | loan_orig | ltv_orig | rate_orig | age | income | liqassets |
|---|---|---|---|---|---|---|---|
| 30-yr MTG, 2005-2007 orig | 4,772,642 | 214,049 | 0.800 | 0.050 | 44 | 85,873 | 78,396 |
| Not fixed-rate | 912,826 | 315,753 | 0.740 | 0.053 | 48 | 109,867 | 119,248 |
| No mod/refi | 532,223 | 344,535 | 0.710 | 0.049 | 49 | 127,937 | 163,874 |
| 5-yr hybrid ARM | 71,053 | 350,365 | 0.710 | 0.060 | 48 | 124,186 | 135,780 |
| MTG/CC data window | 19,823 | 398,456 | 0.700 | 0.061 | 48 | 138,392 | 150,202 |
| CC activity filter | 5,021 | 473,505 | 0.690 | 0.060 | 48 | 179,339 | 217,783 |

The first column of Tables 1.1 and 1.2 are explained as follows. "30 yr MTG, 2005-2007 orig" refers to the universe of 30-year mortgages (fixed, floating, and hybrid) that were originated between April 2005 and December 2007. "Not fixed-rate" excludes all mortgages that are pure fixed-rate. "No mod/refi" imposes the restriction of no modifications or refinancings during the life of the mortgage. "5-year hybrid ARM" requires the mortgages to be 5-year fixed term hybrid

[5]Throughout this section, medians are reported as the average of all observations lying between the 49th and 51st percentiles. The reason for this alternate definition is that the conventional median corresponds to the actual data of an individual account holder (or the average of two), which the Bank does not make publicly available.

**Table 1.2:** *Evaluation of sample restrictions (medians)*

| sample | n | loan_orig | ltv_orig | rate_orig | age | income | liqassets |
|---|---|---|---|---|---|---|---|
| 30-yr MTG, 2005-2007 orig | 4,772,642 | 161,539 | 0.800 | 0.049 | 43 | 63,048 | 25,178 |
| Not fixed-rate | 912,826 | 226,241 | 0.770 | 0.056 | 47 | 74,980 | 32,983 |
| No mod/refi | 532,223 | 230,144 | 0.750 | 0.048 | 49 | 86,087 | 57,061 |
| 5-yr hybrid ARM | 71,053 | 254,474 | 0.750 | 0.059 | 47 | 87,235 | 46,546 |
| MTG/CC data window | 19,823 | 298,774 | 0.750 | 0.060 | 47 | 98,525 | 53,190 |
| CC activity filter | 5,021 | 359,665 | 0.740 | 0.060 | 49 | 120,440 | 87,948 |

ARMs. "MTG/CC data window" requires there to be continuous, non-missing mortgage and credit card data in a 2-year window around each household's ARM reset. Finally, "CC activity filter" implements the sample restrictions related to credit card usage frequency and account age described in Section 1.2.2.

The remaining columns represent the number of customer accounts within each sample ("n"), the original loan amount ("loan_orig"), the loan-to-value ratio at origination ("ltv_orig"), the interest rate at origination ("rate_orig"), and customer age ("age"), estimated annual income ("income"), and estimated liquid asset balance ("liqassets") as of the earliest possible date in the reset window.

The key insight from Table 1.1 is that as more filters are incorporated, the sample becomes much smaller and more biased toward households with bigger mortgages, higher income and wealth, and older age. For example, there are over 4.7 million customers with 30-year mortgages originated between April 2005 and December 2007. After limiting the sample to 5-year hybrid ARM borrowers that meet the mortgage and credit card data sufficiency requirements in the last two rows of Tables 1.1 and 1.2, the remaining sample size is only 5,021. In addition, the average mortgage balances, income, and liquid asset holdings of the final sample are more than twice as large as the same figures for the initial sample. Table 1.2 tells the same story but also conveys that the sample is quite skewed - means are considerably greater than medians.

Clearly, the final sample of households is not perfectly representative of the typical household with a mortgage. Therefore, any results from analyzing this sample are only applicable to households that are generally older and wealthier, rather than the universe at large. This is an important caveat, considering that one of the main conclusions of this paper is related to how sophisticated households are. In spite of this bias, the final sample does not just contain households that are very high-earning and wealthy. For example, the 25th percentile of income and liquid asset holdings for the final sample are $79,000 and $36,000, respectively.

The sample of households in the final row of Tables 1.1 and 1.2 can further be broken down by amortization group: zero-zero, positive-positive, and zero-positive (see Section 1.2.1). The households in these three groups will form the core samples for the empirical analysis of this paper. Table 1.3 presents detailed mortgage summary statistics for these three groups, at both origination and reset. Each group has two rows in the table, one for means and one for medians. The column "pmt_orig" refers to the initial 5-year fixed payment, "pmt_reset" refers to the first post-reset payment, and "fico" refers to the customer's FICO score as of 12 months before reset.

**Table 1.3:** *Mortgage characteristics by amortization group*

| summary_stat | amort_final | n | loan_orig | ltv_orig | rate_orig | pmt_orig | ltv_reset | rate_reset | pmt_reset | fico |
|---|---|---|---|---|---|---|---|---|---|---|
| mean | zero_zero | 912 | 389,294 | 0.720 | 0.065 | 2,046 | 1.070 | 0.032 | 988 | 781 |
| mean | positive_positive | 1,182 | 327,825 | 0.710 | 0.060 | 1,945 | 0.870 | 0.032 | 1,439 | 783 |
| mean | zero_positive | 2,564 | 581,440 | 0.670 | 0.059 | 2,770 | 0.980 | 0.031 | 2,696 | 782 |
| median | zero_zero | 912 | 288,704 | 0.770 | 0.064 | 1,521 | 1.040 | 0.031 | 748 | 791 |
| median | positive_positive | 1,182 | 207,659 | 0.750 | 0.059 | 1,251 | 0.850 | 0.031 | 916 | 795 |
| median | zero_positive | 2,564 | 481,439 | 0.720 | 0.058 | 2,257 | 0.980 | 0.031 | 2,230 | 793 |

Across the three amortization groups, samples size ranges from 912 to 2,564. In terms of mortgage characteristics, all groups have relatively high credit quality as measured by FICO score. The zero-zero and positive-positive groups appear quite similar at origination based on loan size, loan-to-value (LTV) ratios, rates, and payments. In addition, both groups experience relatively large reductions in mortgage payments at reset, though the reduction is larger for the zero-zero group. The substantial difference in LTV ratios at reset is because the positive-positive group pays down principal while the zero-zero group does not in the 5 year gap between origination and reset. The zero-positive group is quite different from the others according to Table 1.3, with higher loan balances and payments at origination. Most importantly, this group experiences a very small change in its payment at reset compared to the other two groups.

Figure 1.2 presents a frequency distribution of the change in households' monthly mortgage payments at reset by amortization group, where negative values on the x-axis represent reductions in payments. Consistent with Table 1.3, households in the zero-zero group have the largest payment reductions at reset, followed by the positive-positive group. The reasons for this relative ordering are given precisely in Appendix A.1. The intuition is that only the interest portion of the positive-positive group's payments experience a decline at reset due to a lower rate. The principal portion effectively does not. In contrast, the full interest-only payment of the zero-zero group experiences a decline.

Figure 1.2 also highlights that among the three groups, the zero-positive group experiences the smallest payment reduction, which is close to zero on average. There are two two offsetting forces: the decline in the interest rate at reset makes payments smaller but post-reset payments now include a principal portion, making payments larger. The net effect depends on the specific interest rate reduction that a given household experiences at reset (see Appendix A.1).

One piece of external validation for the anonymized data used in this paper is that the characteristics of the the zero-zero group documented in Table 1.3 and Figure 1.2 are fairly consistent with those of the households studied by Di Maggio *et al.* (2015) (not shown here). The reason that this specific comparison represents a validation is because Di Maggio *et al.* (2015) focus exclusively on 5-year fixed term hybrid ARM borrowers that have interest-only payments for the first 10 years after origination, a restriction that is conceptually similar to the definition of the zero-zero group in Section 1.2.1. It is therefore comforting that despite being drawn from different data sources, these two samples exhibit similar mortgage characteristics.

Table 1.4 presents demographics data by amortization group as of April 2009. Consistent with Table 1.3, the zero-zero and positive-positive groups are broadly similar along the dimensions of age, income, and wealth. Also consistent with Table 1.3, the zero-positive group is older and has higher income and wealth. While statistics on the geographic concentration of households in eac group are not reported in Table 1.3, there are some sharp differences. Such heterogeneity would

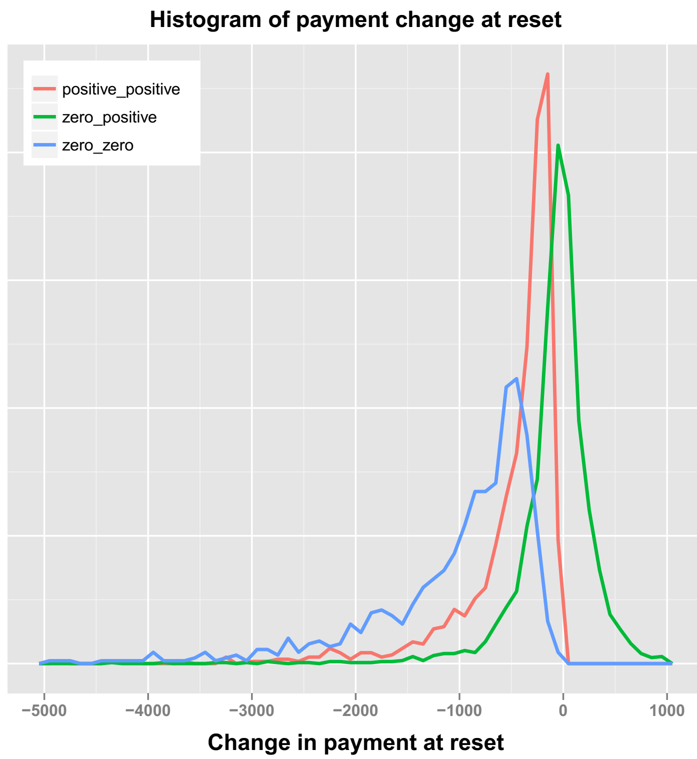



**Figure 1.2:** *Frequency distribution of payment change at reset*

**Table 1.4:** *Demographic characteristics by amortization group*

| summary_stat | amort_final | n | age | income | liqassets |
|---|---|---|---|---|---|
| mean | zero_zero | 912 | 47 | 167,043 | 178,661 |
| mean | positive_positive | 1,182 | 47 | 180,854 | 163,725 |
| mean | zero_positive | 2,564 | 49 | 185,778 | 254,969 |
| median | zero_zero | 912 | 47 | 127,065 | 77,953 |
| median | positive_positive | 1,182 | 47 | 115,097 | 73,884 |
| median | zero_positive | 2,564 | 49 | 126,423 | 110,578 |

be worrisome if identification relied on comparisons across these three groups. However, as will be described in Section 1.3, the empirical analysis will be based on within-group comparisons.

Finally, Table 1.5 presents summary statistics on credit card usage for each amortization group during the 2 year window around households' reset months. The columns show credit card revolving balance ("cc_revbal"), credit card limit ("cc_creditlimit"), number of credit card transactions ("cc_txns"), and creedit card spend ("cc_spend"). The main takeaway from this table is that credit card spending is a large component of consumption for these households, especially compared to autos. Median monthly credit card spending ranges from $2,334 to $2,508, compared to average monthly spending on autos of $300 over a similar time period (Di Maggio *et al.* (2015)).

**Table 1.5:** *Credit card usage characteristics by amortization group*

| summary_stat | amort_final | n | cc_revbal | cc_creditlimit | cc_txns | cc_spend |
|---|---|---|---|---|---|---|
| mean | zero_zero | 912 | 4,939 | 30,997 | 38 | 3,818 |
| mean | positive_positive | 1,182 | 4,845 | 31,559 | 38 | 4,105 |
| mean | zero_positive | 2,564 | 6,010 | 36,650 | 37 | 4,037 |
| median | zero_zero | 912 | 0 | 26,272 | 29 | 2,415 |
| median | positive_positive | 1,182 | 0 | 25,804 | 29 | 2,401 |
| median | zero_positive | 2,564 | 0 | 31,737 | 29 | 2,553 |

Before reviewing the empirical methodology in the next section it is worth summarizing the two key points from this section. First, compared to the 30-year mortgage universe, the sample used in this paper is biased toward households with higher wealth and income, and probably higher consumption as well. However, the sample still includes many households with moderate levels of wealth and income. Second, there are many observable differences between the three amortization groups, which justifies analyzing each of them separately.

## 1.3 Empirical methodology

The research design of this paper is similar to Di Maggio *et al.* (2015). The objective is to estimate the causal effect of mortgage reset on a number of outcomes, with a focus on credit card spending (hereafter, credit card spending will be used to refer to the outcome). The analysis is done separately for each amortization group (zero-zero, positive-positive, and zero-positive) to account for observable (see Tables 1.3-1.5) and potentially unobservable differences across groups that could be correlated with the trend of credit card spending.

For each amortization group, I run an event study in a 2-year window around households' mortgage reset months. The window length of 2 years was chosen to balance three considerations. First, the window should be wide enough to be able to identify any effects of mortgage reset that occur before (i.e. anticipation effects) or after (i.e. delayed reactions) reset actually takes place. Second, since every household needs to have continuous data in the full window around its reset month, a window that is too wide would result in a small sample size. Finally, we will see in Section 1.3.3 that as the window gets wider, the ability to precisely identify causal effects becomes more limited. A window length of 2 years was chosen taking all of these issues into consideration.

### 1.3.1 Identification assumption

The identification assumption that underlies the event study is as follows: households in the same amortization group but with different origination dates would share common trends in credit card spending in the 2-year window around each household's reset month if reset were not to occur. The exact implications of this assumption are made more clear in Figure 1.3, which shows a timeline of mortgage origination and reset for two hypothetical households, $A$ and $B$.

As noted in Section 1.2.1, this paper focuses on households with 5-year fixed term hybrid ARMs originated between April 2005 and December 2007 and therefore resetting between April 2010 and December 2012. Suppose that households $A$ and $B$ are of the same amortization type and originate their mortgages on the two ends of the origination window, as demonstrated in Figure 1.3. This means that household $A$'s "event window", defined as the 2-year window around its reset month, is April 2009 to April 2011. Similarly, household $B$'s event window is December 2011 to December 2013.

With this framework, the identification assumption can be rearticulated in the context of households $A$ and $B$. The assumption states that in a hypothetical world in which $A$'s and $B$'s hybrid ARMs do not reset, their credit card spending would share common trends within each of their event windows. In other words, household $B$ is a good counterfactual for household $A$ from April 2009 to April 2011 and vice versa from December 2011 to December 2013.

Why might this assumption be valid? First, households $A$ and $B$ made fairly similar choices with respect to their mortgages. They both chose to originate 5-year hybrid ARMs with the same

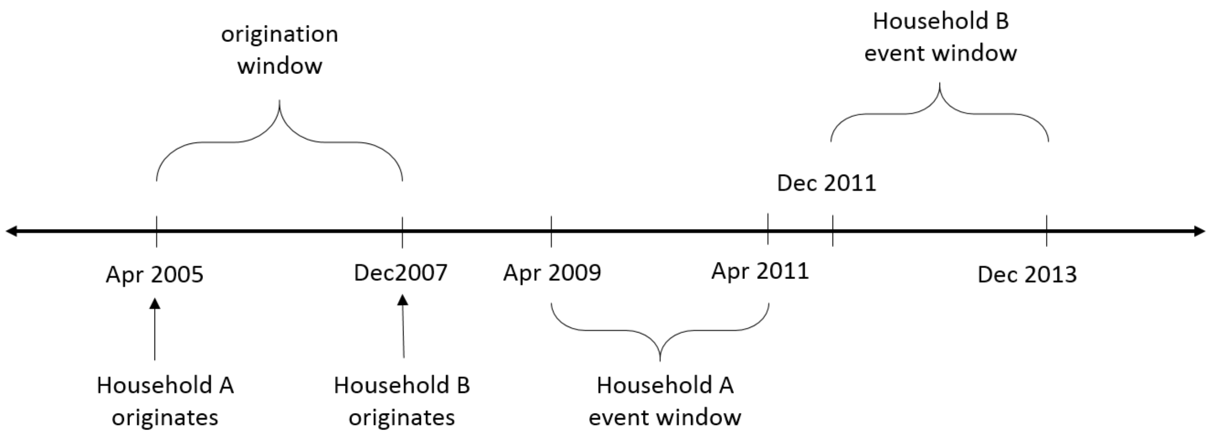



**Figure 1.3:** *Timeline of origination and reset*

pre/post-reset amortization structure. They also chose to originate in a relatively high interest rate period, before the beginning of the Great Recession. Therefore, even if household credit card spending is correlated with mortgage-related decisions, households $A$ and $B$ can still be good counterfactuals for each other.

Second, it is not essential that households $A$ and $B$ share common trends in credit card spending in the full time series. In fact, it is reasonable for households to change their spending patterns around salient events like a home purchase. For example, households may curtail spending as a consequence of feeling the pinch from making a large down payment. On the other hand, households may also increase spending on a host of goods that are complementary to homes (furnishings, decorative items, etc). However, these changes in spending patterns are likely to be limited to a narrow window around mortgage origination. For the event study of this paper, the common trend assumption for identification only needs to apply within each household's event window, which is 4-6 years after origination.

Therefore, even if one thinks that the trend of credit card spending for household $B$ would be different from that of household $A$ in the immediate vicinity of the former's origination in December 2007, this does not invalidate identification unless this difference persists through household $A$'s event window, which is 16-40 months later. Similarly, during household $B$'s event window, household $A$ is more than 6 years removed from its origination date and so is unlikely to still be affected by the origination.

### 1.3.2 Formal specification

The formal regression specification for the event study is given below in (1.1). Note that this specification is applied separately to each of the three amortization groups.

$$Y_{i,t} = \sum_{k=-12}^{12} \beta_k \mathbf{1}\{\tau_{i,t} = k\} + \beta_L \mathbf{1}\{\tau_{i,t} < -12\} + \beta_R \mathbf{1}\{\tau_{i,t} > 12\} + \lambda_i + \eta_t + \varepsilon_{i,t} \tag{1.1}$$

The dependent variable $Y_{i,t}$ is the value of an outcome variable (i.e. credit card spending) for household $i$ in month $t$. The variable $\tau_{i,t}$ is "event time": the difference in months between $t$ and the month in which household $i$ makes its last pre-reset payment. For example, suppose that a household's first post-reset payment occurs in July 2011. Then $\tau_{i,t} = 1$ for this household when $t =$ July 2011. For $t =$ June 2011 (the last month in which a pre-reset payment is made), $\tau_{i,t} = 0$ and for $t =$ August 2011, $\tau_{i,t} = +2$.

The coefficients of interest in (1.1) are the $\beta_k$'s, the coefficients on monthly event time dummies between -12 (one year before reset) and +12 (one year after reset), i.e. the event window. There are also dummies for being to the left and right of the event window, with coefficients $\beta_L$ and $\beta_R$, respectively. Finally, the specification includes both household ($\lambda_i$) and monthly ($\eta_t$) fixed effects. The household fixed effect controls for any time-invariant household attributes that may be correlated with the level of credit card spending. The time fixed effect controls for any temporal effects on credit card spending that apply to all households in the sample.

The interpretation of $\beta_k$ is the causal impact of mortgage reset on credit card spending $k$ months after reset occurs. However, due to the inclusion of fixed effects, $\beta_k$ is only identified up to a constant. I choose this constant to be $\beta_{-12}$ to highlight that the event study is effectively a difference-in-differences framework. The interpretation of the normalized coefficient $\beta_k - \beta_{-12}$ is the causal impact of mortgage reset on credit card spending $k$ months after reset relative to 1 year before reset. Put another way, $\beta_k - \beta_{-12}$ represents the difference between treatment and control households when the former is in event month $k$ relative to the same difference when the former is in event month $-12$.

The normalization of coefficients is formally presented below. First, define $t^*$ implicitly with $\tau_{i,t^*} = -12$. This definition implies that $t^*$ is the month in which household $i$ is 12 months before reset. Using (1.1), we have

$$\begin{aligned} Y_{i,t^*} &= \beta_{-12} + \lambda_i + \eta_{t^*} + \varepsilon_{i,t^*} \\ &= \sum_{k=-12}^{12} \beta_{-12}\mathbf{1}\{\tau_{i,t} = k\} + \beta_{-12}\mathbf{1}\{\tau_{i,t} < -12\} + \beta_{-12}\mathbf{1}\{\tau_{i,t} > 12\} + \lambda_i + \eta_{t^*} + \varepsilon_{i,t^*} \end{aligned} \tag{1.2}$$

Then, subtract $Y_{i,t^*}$ in (1.2) from $Y_{i,t}$ in (1.1).

$$\begin{aligned} Y_{i,t} - Y_{i,t^*} = \sum_{k=-12}^{12} (\beta_k - \beta_{-12})\mathbf{1}\{\tau_{i,t} = k\} + (\beta_L - \beta_{-12})\mathbf{1}\{\tau_{i,t} < -12\} \\ + (\beta_R - \beta_{-12})\mathbf{1}\{\tau_{i,t} > 12\} + (\eta_t - \eta_{t^*}) + (\varepsilon_{i,t} - \varepsilon_{i,t^*}) \end{aligned} \tag{1.3}$$

The specification in (1.3) is the regression that is run for each amortization group. It is clear from (1.3) that the coefficients on the event time dummies will reflect the desired normalization relative to $\beta_{-12}$. In addition, the coefficient on the event time dummy for $k = -12$ will always be zero by construction.

### 1.3.3 Identification revisited

The formal specification for the event study given in (1.1) permits a more nuanced discussion of how causal effects are identified in this paper. The identification assumption stated in Section 1.3.1 is that households in the same amortization group but with different origination dates would share common trends in credit card spending in the 2-year window around each household's reset month if reset were not to occur. Using the notation of (1.1), this assumption is equivalent to

$$E(\varepsilon_{i,t} \mid \lambda_i, \eta_t, \tau_{i,t}) = 0 \tag{1.4}$$

That is, conditional on household, time, and event window effects, there is no expected difference between actual credit card spending and spending predicted by (1.1).

A similar identification assumption is stated in Di Maggio *et al.* (2015) and is used to justify that for a given household, all other households with different reset dates are appropriate counterfactuals, no matter when their resets actually occur:

> In other words, we estimate the consumption response of the households who experienced a reduction in the interest payment, relative to that of households holding the same mortgage, but with a different reset date... the assumption is that households whose mortgage is reset in May 2010 are basically comparable to households that experience their reset, say in December.

However, based on the specification in (1.1), households that are appropriate counterfactuals must have reset dates that are sufficiently far apart in time from the household in question. We will show this with the following example.

Suppose household $A$ experiences its first month after reset in June 2011, which is given by $t = t'$. This means that $\tau_{A,t'} = 1$. Let household $B$ experience its second month after reset in June 2011, or $\tau_{B,t'} = 2$. This implies that households $A$ and $B$ have reset dates that are one month apart. Suppose that we try to estimate $\beta_1 - \beta_{-12}$ using a difference-in-differences (DD) estimate $\hat{\beta}$, where household $A$ is the treatment group and household $B$ is the control. Since $t' =$ June 2011 is the first month after reset for the treatment household ($A$), this implies that May 2010 $= t' - 13$ is 12 months before reset for household $A$. Therefore, the DD estimate takes the following form.

$$\hat{\beta} = (Y_{A,t'} - Y_{B,t'}) - (Y_{A,t'-13} - Y_{B,t'-13}) \tag{1.5}$$

$$= ((\beta_1 + \lambda_A + \eta_{t'} + \varepsilon_{A,t'}) - (\beta_2 + \lambda_B + \eta_{t'} + \varepsilon_{B,t'})) \tag{1.6}$$

$$- ((\beta_{-12} + \lambda_A + \eta_{t'-13} + \varepsilon_{A,t'-13}) - (\beta_{-11} + \lambda_B + \eta_{t'-13} + \varepsilon_{B,t'-13}))$$

$$\implies E(\hat{\beta}) = (\beta_1 - \beta_{-12}) - (\beta_2 - \beta_{-11}) \tag{1.7}$$

Expression (1.5) is the usual form of the DD estimator: the difference in the outcome $Y$ between the treatment ($A$) and control ($B$) when the treatment group is 1 month after reset minus the same difference when the treatment group is 12 months before reset. Expression (1.6) comes from substituting (1.1) into (1.5). Expression (1.7) comes from canceling terms, taking expectations of both sides, and employing the identification assumption in (1.4).

Notice that the expression in (1.7) does not properly identify $\beta_1 - \beta_{-12}$ as intended. Household $B$ is not an appropriate counterfactual because it is also being "treated" during the time period from $t' - 13$ to $t'$. While household $A$ experiences the treatment of being 1 month after relative to 12 months before reset, household $B$ experiences the treatment of being 2 months after relative to 11 months before reset. The effect of household $B$'s treatment is given by $\beta_2 - \beta_{-11}$, which is precisely the bias displayed in (1.7).

This problem can be addressed by comparing household $A$ to a household whose treatment status does not change over the same time period. For example, suppose the control group is household $C$, who in June 2011 is at least one year before reset ($\tau_{C,t'} < -12$). In this case, the DD estimate takes the following form.

$$\begin{aligned}
\hat{\beta} &= (Y_{A,t'} - Y_{C,t'}) - (Y_{A,t'-13} - Y_{C,t'-13}) \\
&= ((\beta_1 + \lambda_A + \eta_{t'} + \varepsilon_{A,t'}) - (\beta_L + \lambda_C + \eta_{t'} + \varepsilon_{C,t'})) \\
&\quad - ((\beta_{-12} + \lambda_A + \eta_{t'-13} + \varepsilon_{A,t'-13}) - (\beta_L + \lambda_C + \eta_{t'-13} + \varepsilon_{C,t'-13})) \\
\implies E(\hat{\beta}) &= \beta_1 - \beta_{-12}
\end{aligned}$$

The DD estimate is unbiased because from $t' - 13$ to $t'$, household $C$'s treatment status does not change. It is always to the left of its event window.

The general rule is that for household $A$ who experiences event month $k$ in month $t$, valid counterfactuals are households who are outside of their event windows when household $A$ is experiencing event time between $[-12, k]$, or $[t - (k + 12), t]$. This means that for the set of valid counterfactual households, represented by the set $\Omega(k, t)$, $t$ must be more than 12 months before reset or $t - (k + 12)$ must be more than 12 months after reset. The latter condition is equivalent to $t$ being more than $24 + k$ months after reset. Therefore, we have $\Omega(k, t) = \{j : \tau_{j,t} < -12 \text{ or } \tau_{j,t} > 24 + k\}$, where $\Omega(k, t)$ is the set of valid counterfactual households for a treatment household that experiences event month $k$ in month $t$. Clearly, not all households who just have different origination dates are valid counterfactuals.

I close this section by noting two important properties of the set $\Omega(k, t)$. First, as the event window gets wider than 12 months on each side of reset, the set of valid counterfactual households becomes smaller. It is easy to show that with an event window length $L$ on each side of reset, the set of valid counterfactual households is $\Omega(k, t, L) = \{j : \tau_{j,t} < -L \text{ or } \tau_{j,t} > 2L + k\}$. Therefore, as $L$ gets larger, a household has to reset even farther from $t$ to be a valid counterfactual. As mentioned at the beginning of this section, this is another reason why a very wide event window is untenable. Parameter estimates would be very imprecise with a small set of valid counterfactual households.

Second, as $k$ gets larger, the set $\Omega(k, t)$ becomes smaller because it becomes less likely for a candidate counterfactual household to be more than $24 + k$ months after reset in month $t$. This means that holding other things equal, the normalized coefficients $\beta_k - \beta_{-12}$ should be estimated with less and less precision as $k$ goes from -12 to 12. We shall see this in the event study results, which we turn to next.

## 1.4 Event study results

In this section, I present the results of the event study specified in (1.3). The results are displayed in graphical form, with the x-axis representing the event window from 12 months before to 12 months after reset and the y-axis representing the magnitude of the normalized event month coefficients. Recall from Section 1.3 that the interpretation of the coefficient for event month $k$ is the effect of mortgage reset on the outcome in month $k$ relative to 12 months before reset, relative to control.

In addition to credit card spending, results are shown for a number of other outcome variables. First, to highlight the kinds of changes that occur at reset, graphs are presented for the outcome variables "mortgage rate" and "monthly mortgage payment". Each graph will have three lines, one for each amortization group (zero-zero, positive-positive, and zero-positive).

Next, I focus on the main credit card spending outcome variable, with each amortization group studied separately. In addition to reporting the baseline results, I also report the results for two subsets of each group: above-median and below-median liquid assets, representing wealthy and poor households, respectively. This split will establish whether there is any heterogeneity in the effect of mortgage reset by wealth. I also report for each group a decomposition of the credit card spending outcome into 9 different categories of spending (i.e. general retail, home improvement, leisure, etc). The purpose of this decomposition is to assess whether the effect is concentrated in one or two particular categories or broadly spread across many of them. The last outcome variable studied is revolving credit card balances, i.e. credit card debt.

In all of the graphs, 95% confidence intervals for estimated coefficients are provided, with standard errors clustered at the household level.[6] In addition, to remove the effects of outliers, the data for each regression are winsorized at the 1% and 99% levels. For certain outcome variables, results are shown for the natural log of the variable to estimate relative effects. For these log event studies, a small positive number is added to any zero-valued outcome variables in the data.

### 1.4.1 Mortgage rates and payments

The results of the mortgage rate event studies for each amortization group are given in Figure 1.4. The first result to note is that before the first post-reset month (x-axis $< 1$), all three groups are at zero. This is expected because 5-year hybrid ARMs have a fixed mortgage rate until reset occurs.

Upon reset, the mortgage rate of all three groups falls dramatically, with the decline ranging from 276 to 332 basis points. The reason that the zero-zero group has a larger decline than the others is because these mortgages tended to be originated closer to the end of the origination window (December 2007) than the other two groups. As seen in Figure 1.1, the end of the origination window is when market interest rates peaked, meaning that the zero-zero group would have originated at higher rates than the other groups. Moreover, market rates do not exhibit much variation during the reset window - they are consistently low. Therefore, households in the zero-zero group exhibit the largest decline in their mortgage interest rates.

[6] Clustering was also done at the monthly level for robustness. The resulting standard errors were generally smaller than when clustering by household.

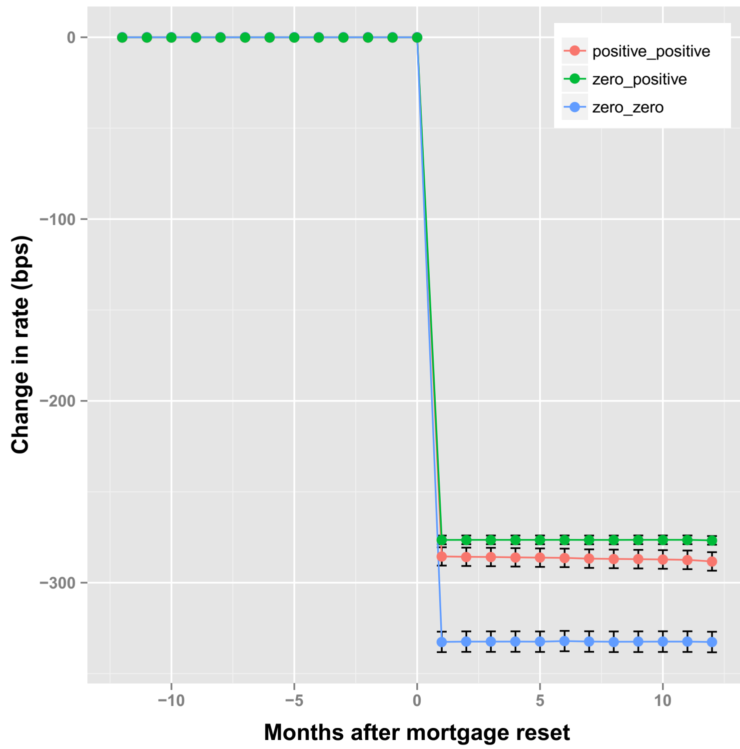



**Figure 1.4:** *Event study of mortgages rates*

The event study results for monthly mortgage payments are given in Figure 1.5. Consistent with Figure 1.2 and the analysis of Appendix A.1, there is considerable heterogeneity in payment reductions across amortization groups. Despite experiencing very similar interest rate changes at reset, households in the zero-zero, positive-positive, and zero-positive groups experience average payment reductions of approximately $1000, $500, and $60, respectively.

How large are these payment reductions relative to pre-reset payments? This question can be answered by using the natural log of mortgage payments as the outcome variable, the results of which are shown in Figure 1.6. It is clear that the relative payment reductions are quite large for the zero-zero and positive-positive groups: approximately 75 and 30 log points, respectively.[7] Meanwhile, the zero-positive group experiences only a 2 log point decline in payments.

[7] Log point declines of 75 and 30 correspond to percentage declines of 53% and 25%, respectively.

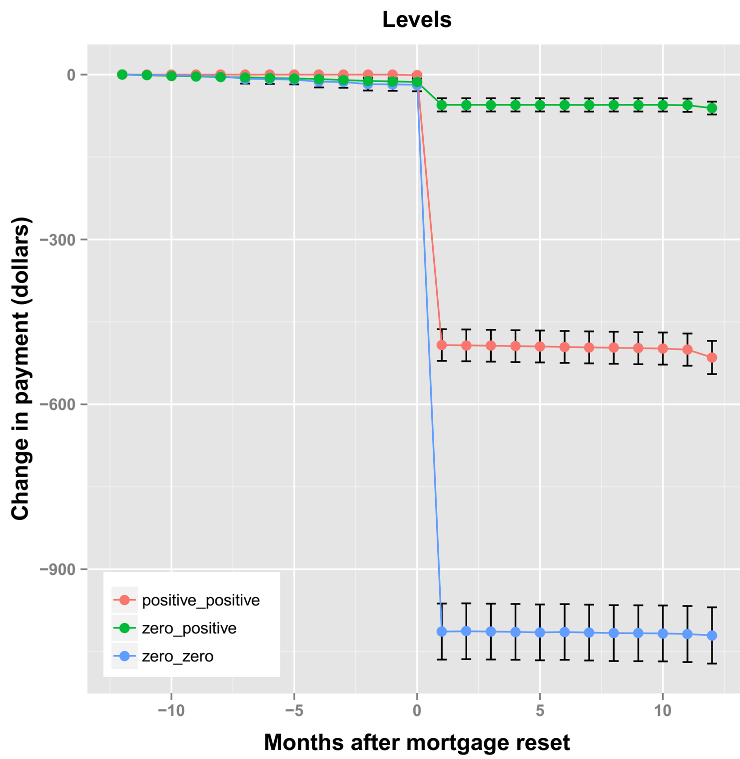



**Figure 1.5:** *Event study of mortgage payments*

The preceding results indicate that while the zero-zero and positive-positive groups experience large absolute and relative declines in mortgage payments at reset, the zero-positive group's payments change very little. As we continue to the main outcome of credit card spending in Section 1.4.2, it should be kept in mind that this disparity between the zero-positive group and the others provides an informal check on the the research design of this paper. Specifically, suppose that large effects on credit card spending are estimated for the zero-zero and positive-positive groups (who experience large payment reductions). These effects will only have credibility as causal impacts of payment reductions if the same effects are **not** observed in the zero-positive group (who experience very small payment reductions). Otherwise, estimated effects could be caused by something else that occurs at reset or a faulty identification strategy.

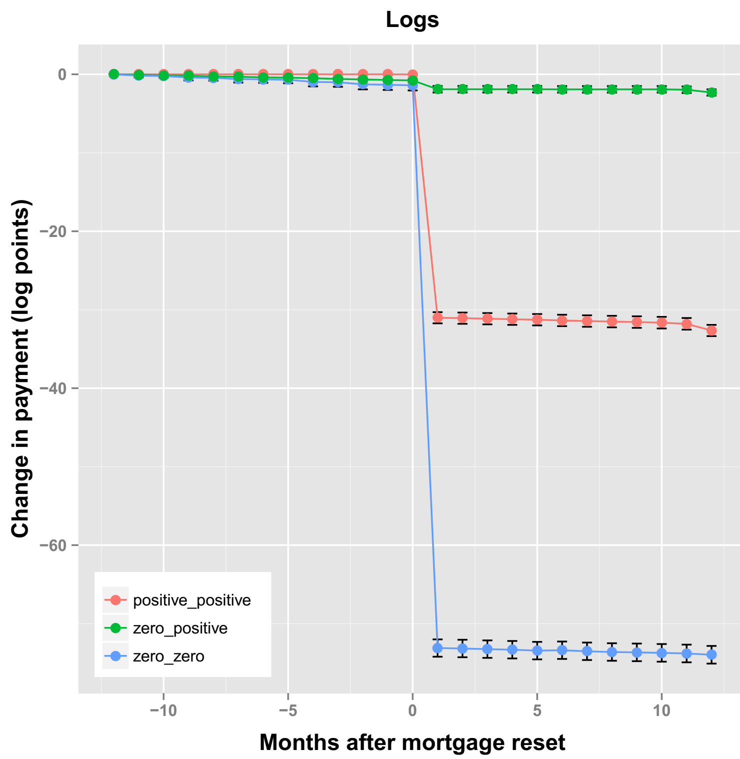



**Figure 1.6:** *Event study of mortgage payments*

### 1.4.2 Credit card spending

The event study results for credit card spending are presented separately for each amortization group: zero-zero, positive-positive, and zero-positive. The coefficients in the graphs represent the amount by which credit card spending in each event month in the event window exceeds spending from 1 year before reset, relative to control households. By construction, the coefficients for event month -12 will equal zero. In addition to the baseline results, results are also shown for sub-samples of each amortization group based on household wealth.

#### Zero-zero group

The credit card spending results for the zero-zero group are reported in Figure 1.7. There are three salient features. First, credit card spending rises dramatically over the course of the entire event window. One year after reset, spending is a statistically significant $500 higher than it is 1 year before reset, which represents about 50% of the average payment reduction of $1000 for this group (see Figure 1.5). Second, the increase in spending occurs gradually over several months, reaches statistical significance before reset, and is limited to the pre-reset window. When households actually experience the decline in payments in event month = 1, there is no sharp jump in spending and it is relatively flat thereafter. This is very different from the results on auto spending in Di Maggio *et al.* (2015). Third, as expected based on the discussion at the end of Section 1.3.3, coefficients are estimated with less precision as event time goes from -12 to 12.

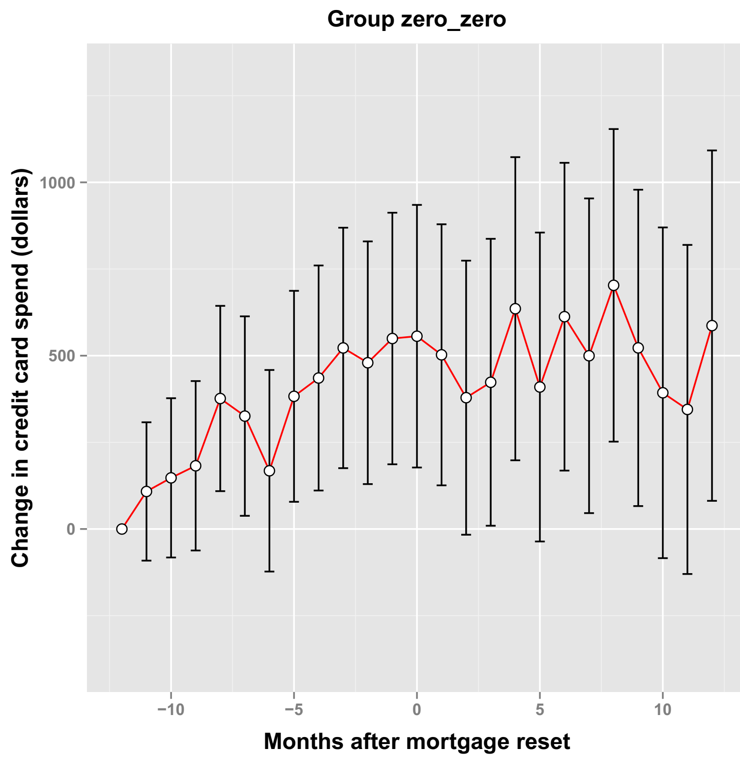



**Figure 1.7:** *Event study of credit card spending for zero-zero group*

Results are reported for the log of credit card spending in Figure 1.8. The coefficients exhibit the same pattern as that of the levels results in Figure 1.7. Credit card spending rises by a statistically significant 40 log points over the course of the pre-reset window and if anything, declines slightly in the post-reset window. It should be noted that 40 log points is more than the 20% that results from putting the levels result in relative terms.[8] The reason for this disparity is

[8]The figure of 20% is obtained by dividing the $500 increase in spending (Figure 1.7) by median monthly spending

the presence of households that start with a very small amount of spending 12 months before reset and increase spending modestly during the pre-reset window. For these households, the relative change in spending will be extremely large and positive.

However, this does not mean the log results are spurious. If there were households with modest amounts of spending 12 months before reset who reduced spending to very small amounts during the pre-reset window, their relative changes in spending would be extremely large and negative, offsetting the earlier effect. The result that the relative effect remains large and positive therefore means that there are more households of the former type (large relative spending increases during the pre-reset window) than the latter (large relative spending reductions during the pre-reset window).

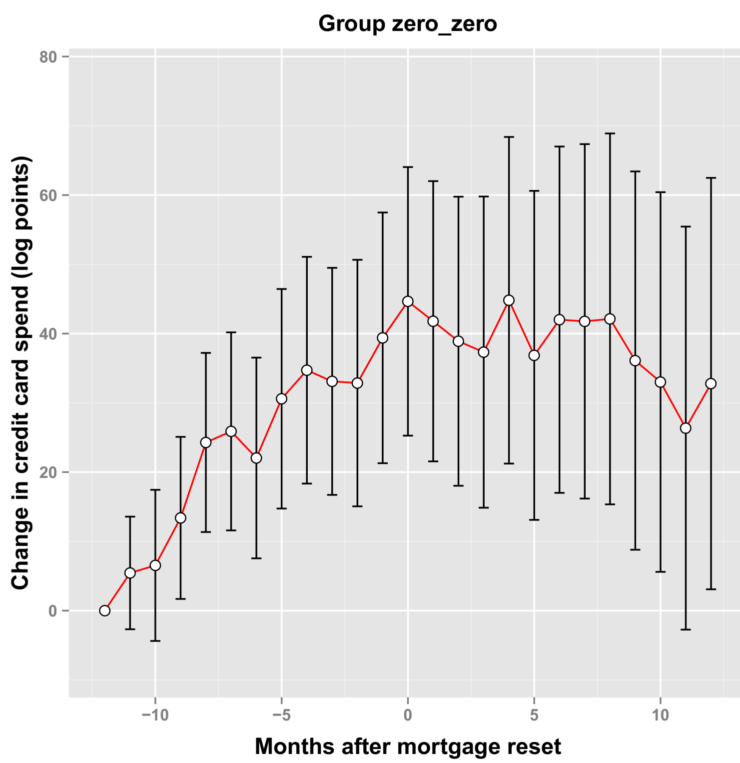



**Figure 1.8:** *Event study of log credit card spending for zero-zero group*

---

for the zero-zero group of $2400 (Table 1.5).

Next, I explore to what extent the effects documented in Figures 1.7 and 1.8 are exhibited by zero-zero households with different levels of wealth. Figure 1.9 addresses this question by reporting event study results for two separate sub-samples: above-median (left plot) and below-median (right plot) liquid assets. The plots (which have the same y-axis scale) clearly indicate that the effect is concentrated in low-wealth households. The coefficients for the high-wealth sample are never statistically different from zero while the coefficients for the low-wealth sample match the pattern of Figures 1.7 and 1.8, with larger magnitudes.

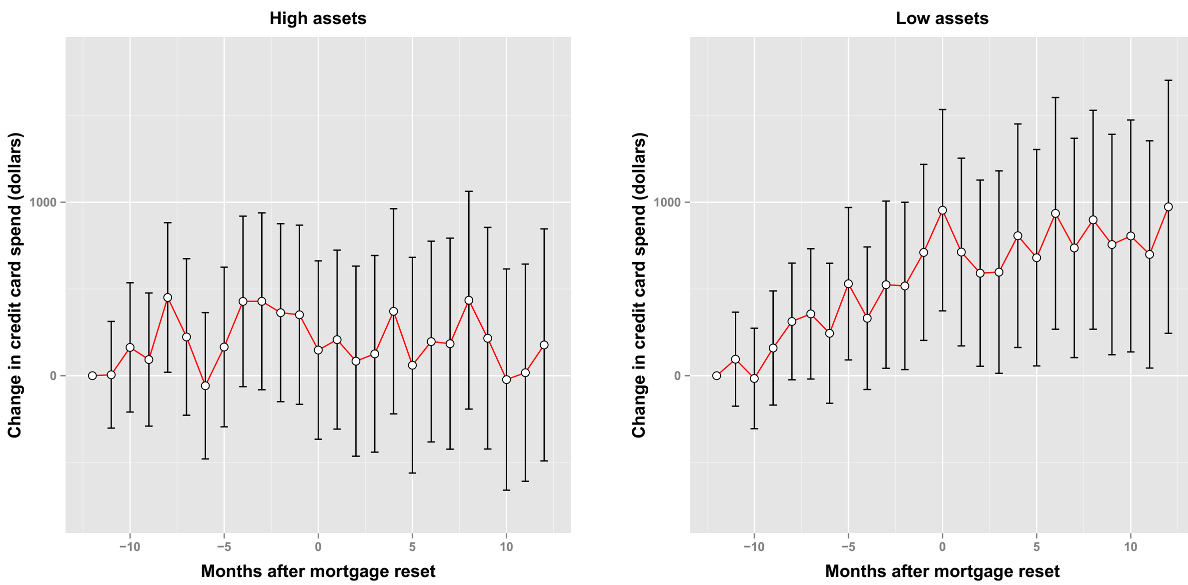



**Figure 1.9:** *Event studies of credit card spending for zero-zero group by household wealth*

In summary, households in the zero-zero group exhibit large increases in credit card spending in both absolute and relative terms due to the decline in mortgage payments at reset. The increase is gradual and comes entirely before reset occurs, with no sharp increase in spending at or after reset. In addition, the effect is concentrated in households with low levels of wealth.

### Positive-positive group

The results for the positive-positive group are very similar to those of the zero-zero group. In Figure 1.10, the results for the positive-positive group are shown for both the level (left) and log (right) of credit card spending. As with the zero-zero group, there is an economically and statistically significant increase in spending during the pre-reset window ($700 or 40 log points), with no sharp change at reset and a flat profile afterward.

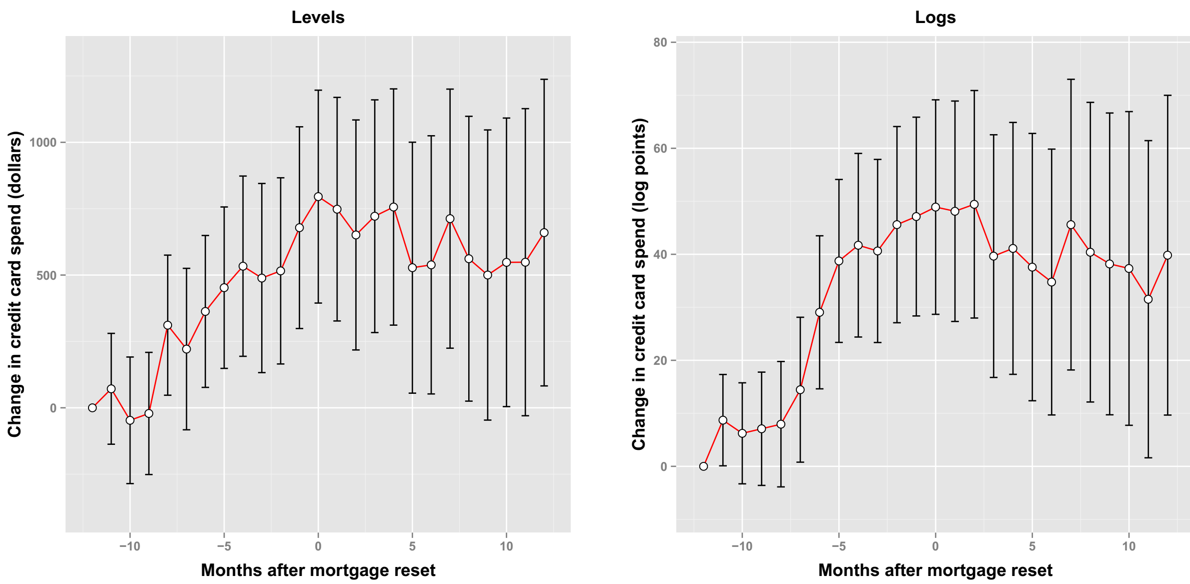



**Figure 1.10:** *Event studies (levels and logs) of credit card spending for positive-positive group*

Figure 1.11 shows that the positive-positive group also displays the same kind of heterogeneity in wealth as the zero-zero group. The pre-reset increase in spending is clearly exhibited in households with lower levels of liquid assets and does not appear strongly among wealthier households.

### Zero-positive group

In Figure 1.12, the event study results on credit card spending are shown for the zero-positive group. Recall from Figure 1.5 that this group experiences a very small reduction in payments at reset on average. Figure 1.12 shows that while the shape of the profile of coefficients resembles that of the zero-zero and positive-positive groups (Figures 1.7 and 1.10), the magnitudes are much smaller and there is no statistical significance. Specifically, while there is an increase in credit card spending of about $150 during the pre-reset window, this point estimate is not significant and is much smaller than the spending increases exhibited by the zero-zero and positive-positive groups.

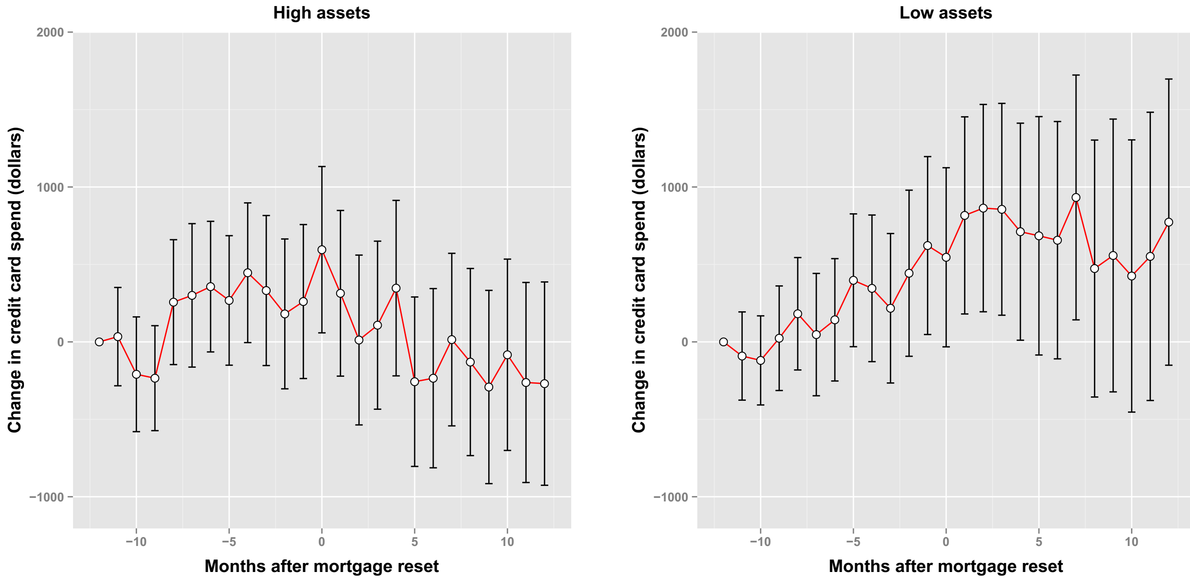



**Figure 1.11:** *Event studies of credit card spending for positive-positive group by household wealth*

After reset, the zero-positive group exhibits no sharp increase in spending - the coefficients remain insignificant and flat around $150.

This result for the zero-positive group lends credibility to the claim that the estimated effects for the zero-zero and positive-positive groups in Sections 1.4.2 and 1.4.2 are indeed caused by payment reductions associated with mortgage reset. Recall that the zero-positive group experiences an interest rate reduction but only a tiny reduction in mortgage payments at reset on average. Therefore, the result that this group also exhibits a small, insignificant spending response in Figure 1.12 means that the large responses exhibited by the zero-zero and positive-positive groups are probably not spurious (i.e. driven by something else correlated with mortgage reset or a faulty identification strategy).

The log plot for the zero-positive group in Figure 1.13 tells a different story: it closely matches the log results of the zero-zero and zero-positive groups (Figures 1.8 and 1.10) in terms of economic and statistical significance. Since the zero-positive levels result is clearly not significant in Figure 1.12, this is a bit puzzling. The logs result must be driven by households that start with a tiny amount of spending 12 months before reset and increase spending modestly during the pre-reset window. For these households, the relative change in spending is quite large and positive but the

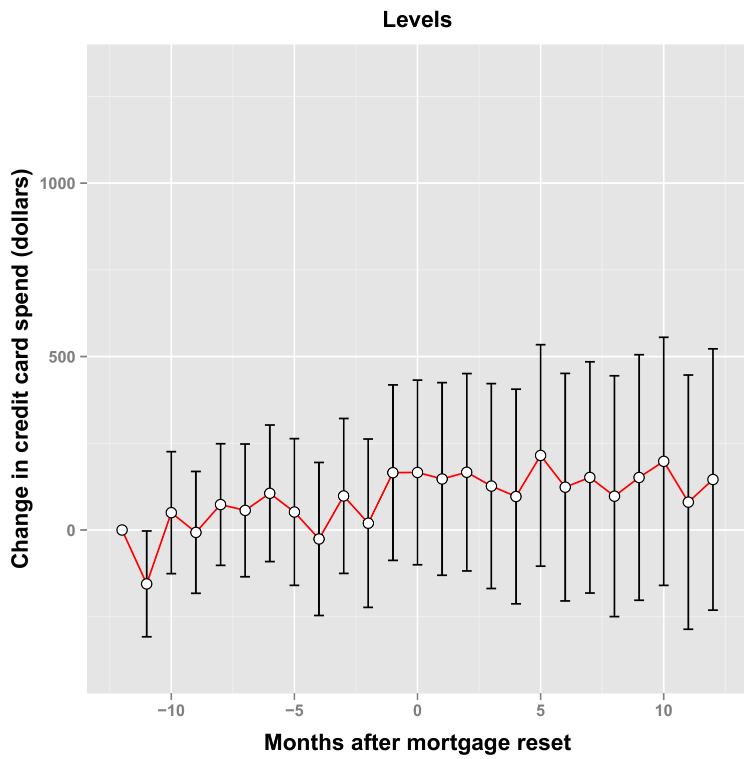



**Figure 1.12:** *Event study of credit card spending for zero-positive group*

absolute change is small.

For this explanation to be robust, there must not be many households that have a modest amount of spending 12 months before reset and reduce spending to a tiny amount during the pre-reset window, as this would result in large, negative effects that would offset the previous effect. Therefore, the level and log results collectively imply that while zero-positive households generally do not increase spending by a large dollar amount, the relative effect is strong due to the "net" presence of households increasing spending modestly but from a small base.

Figure 1.14 shows the zero-positive event study results separately for households with above-median (left) and below-median (right) wealth. Interestingly, to the extent that an effect exists, it is concentrated among wealthier households. Meanwhile, the coefficients for low-wealth zero-positive households on the right side of Figure 1.14 are stable around zero.

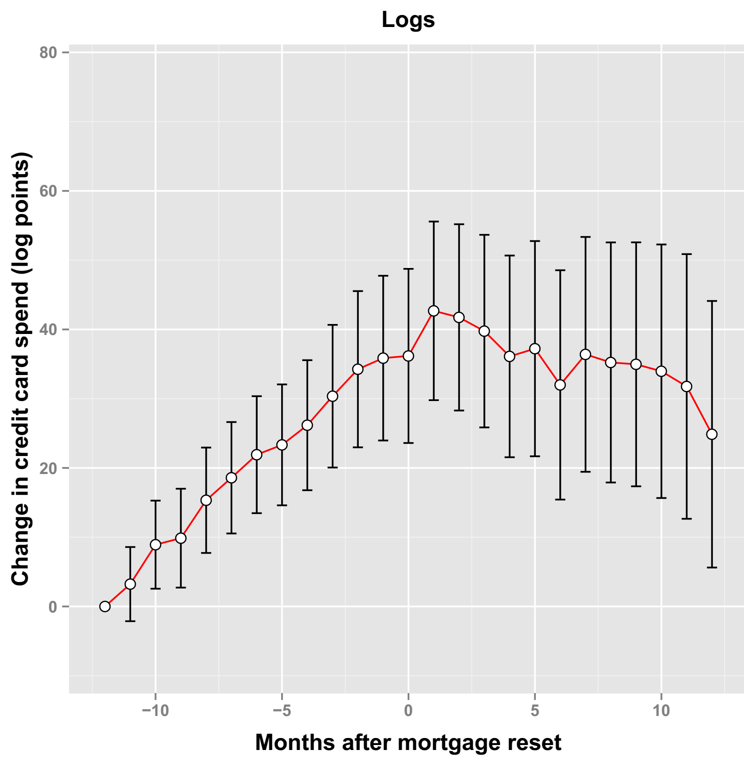



**Figure 1.13:** *Event study of log credit card spending for zero-positive group*

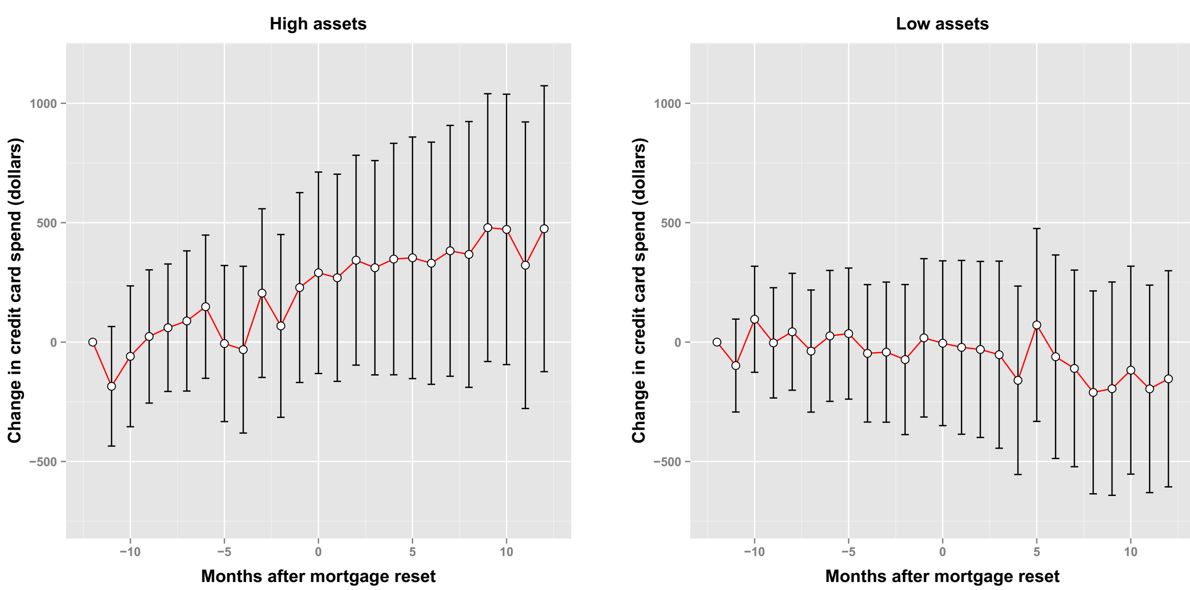



**Figure 1.14:** *Event studies of credit card spending for zero-positive group by household wealth*

### Decomposition of credit card spending

In sections 1.4.2-1.4.2, it was shown that the general pattern of credit card spending around mortgage reset is as follows: a gradual pre-reset increase, no sharp increase at reset, and stability thereafter. In other words, spending increases are concentrated almost entirely in the pre-reset window. In this section, I investigate which categories of spending are driving this result. Specifically, I report event study results using 9 different categories of credit card spending as outcome variables for each of the three amortization groups. The categories are auto parts/maintenance, health care, home improvement, leisure, general retail, services, staples, transportation, and miscellaneous.

Instead of presenting the event study results in the form of line graphs as in the previous sections, for simplicity I present just one statistic for each spending category/amortization group. The statistic is the average of the normalized $\beta_k$'s from event months 1 to 12. The interpretation of this statistic is the average amount by which post-reset spending in a particular category exceeds spending 12 months before reset, relative to control. A more intuitive interpretation is the amount by which spending in a particular category increases during the pre-reset window, relative to control.[9]

Figure 1.15 presents the decomposition statistics for all three amortization groups. The cumulative size of the bars above each amortization group represents the effect of mortgage reset on total credit card spending. This aggregate effect is in turn decomposed into 9 different categories.[10] For example, in the positive-positive group, leisure spending increases by $162 pre-reset whereas for the zero-zero group, home improvement spending increases by roughly $179. Across all three amortization groups, the pre-reset increase is spending, while concentrated in certain discretionary categories like leisure, home improvement, and general retail, is spread fairly broadly. If the increase was confined to just one or two categories, this might indicate that something else may be happening around mortgage reset that is causing spending to increase. However, this does not

[9]This interpretation assumes that the profiles of category-level spending are similar to those of total spending, i.e. relative stability post reset. While results confirming this are not shown here, this is indeed the case.

[10]The disaggregated results do not sum exactly to the results for total spending because the data is winsorized separately for each regression.

appear to be the case.

Figure 1.16 presents the same decomposition as Figure 1.15 using the logs of credit card spending in each category as outcome variables. As such, Figure 1.16 accounts for the possibility that different amortization groups might have different baseline levels of spending in each category. These results provide further evidence of the same point - the pre-reset increase in spending caused by mortgage reset is spread broadly across several categories.

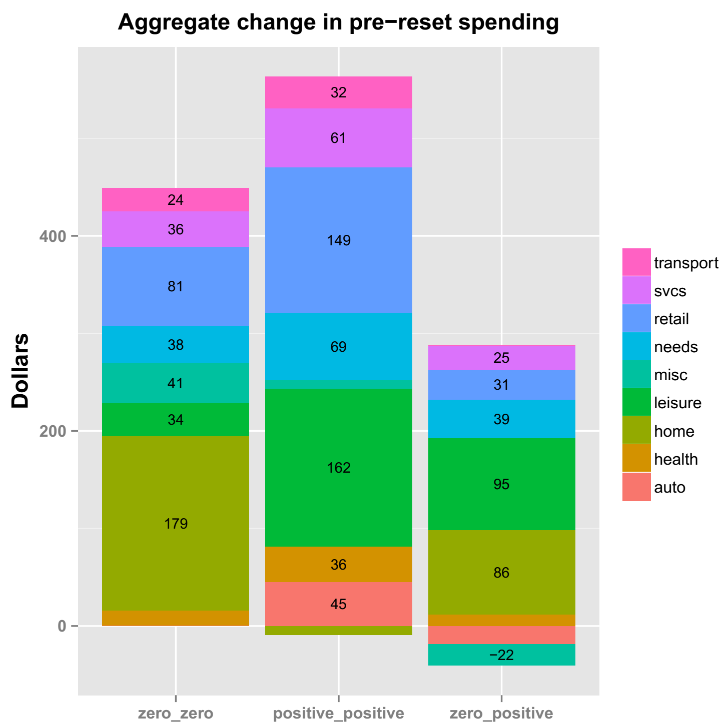



**Figure 1.15:** *Decomposition of credit card spending effects (dollars)*

### 1.4.3 Credit card revolving balances

The last outcome variable that is used in the event study is credit card revolving balances. A household's revolving balance is the portion of its outstanding credit card balance on which it pays interest. The remaining portion of the outstanding balance is the "float" on which interest is not paid. Therefore, carrying a revolving balance is synonymous with borrowing on one's credit

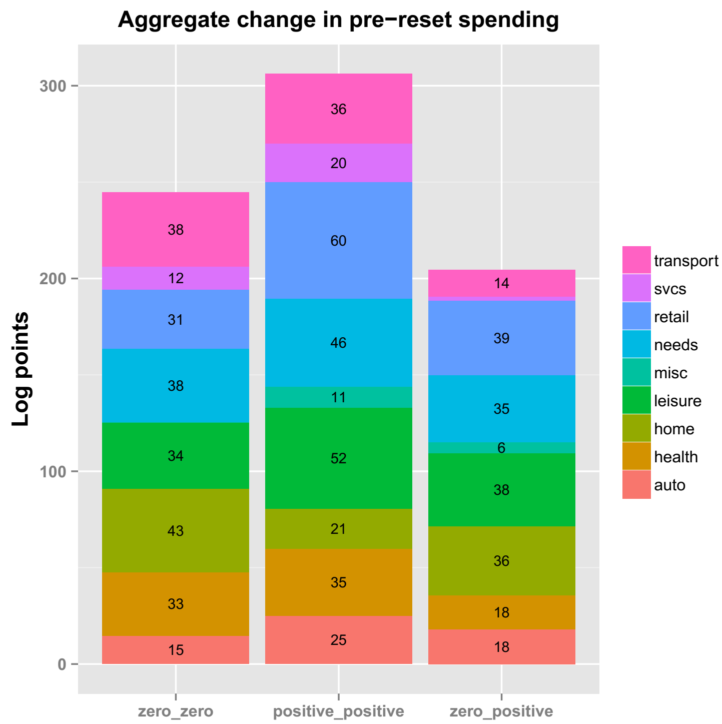



**Figure 1.16:** *Decomposition of credit card spending effects (log points)*

card. In this section, I discuss how credit card revolving balances are affected by mortgage reset.

Figure 1.17 reports the event study results on credit card revolving balances for the zero-zero (left) and positive-positive (right) groups. Although the coefficients are not highly significant, the results seem to exhibit the familiar pattern of a pre-reset increase, no sharp increase at reset, and stability afterward. Recall that for these amortization groups, the spending result was found to be concentrated in households with low levels of liquid assets (Figures 1.9 and 1.11). Therefore, Figure 1.17 suggests that the pre-reset spending increases exhibited by poor households in the zero-zero and positive-positive groups are at least partially financed by credit card debt.

In contrast, Figure 1.18 shows that zero-positive households do not exhibit the same pattern in revolving balances, which remain relatively stable throughout the event window. This result lends credibility to the claim that the estimated effects on revolving balances for the zero-zero and

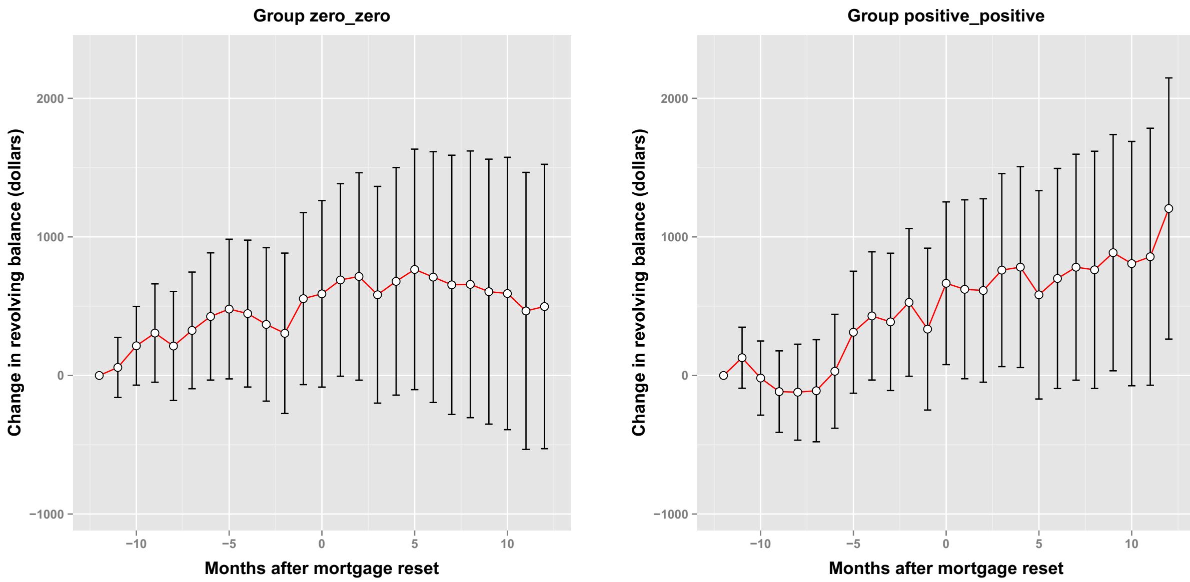



**Figure 1.17:** *Event studies of credit card revolving balances for zero-zero and positive-positive groups*

positive-positive groups in Figure 1.17 are indeed due to payment reductions at reset. Moreover, recall that to the extent there is any spending response among zero-positive households during the pre-reset window, it is concentrated in wealthier households (Figure 1.14). Together with Figure 1.18, this means that if wealthy, zero-positive households did indeed increase spending, it must have been financed by other borrowing or savings (not credit card debt from this Bank).

### 1.4.4 Summary of results and implications

In the preceding sections, I presented a number of event study results for various permutations of outcome variables, amortization groups, and wealth-based subsamples. In this section, I provide a brief summary of these results and discuss their implications on household behavior.

The first result is that households in the zero-zero and positive-positive groups (those that experience large payment reductions at reset) exhibit large, significant increases in credit card spending during the event window. However, unlike the pattern of auto spending documented in Di Maggio *et al.* (2015), the increases in credit card spending are gradual and occur entirely pre-reset. The identification strategy underlying these results is bolstered by the finding that the

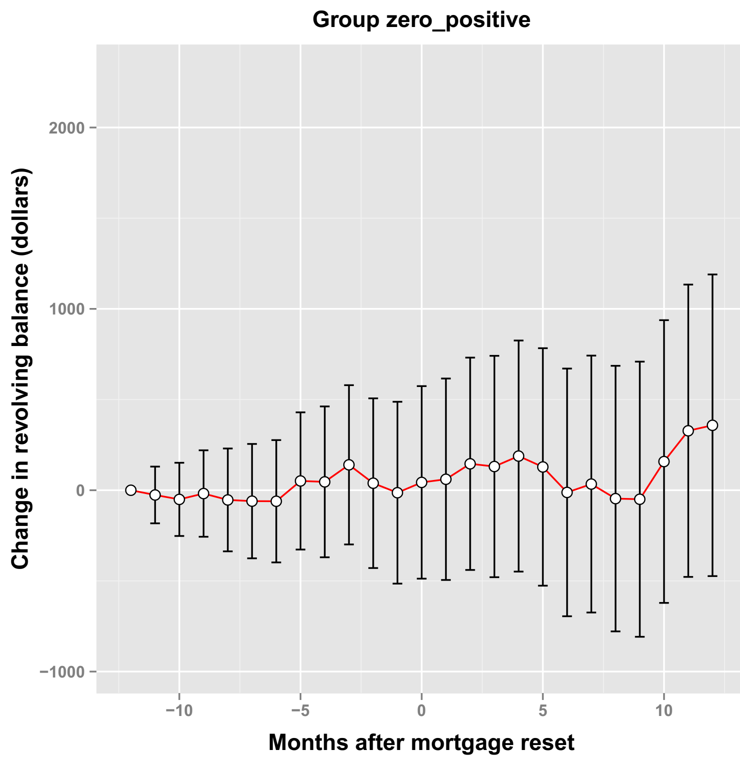



**Figure 1.18:** *Event study of credit card revolving balances for zero-positive group*

zero-positive group, which experiences a similar reset in mortgage rates but only a tiny reduction in payments, exhibits much smaller, insignificant increases in spending during the pre-reset window.

Second, for the zero-zero and positive-positive groups, the spending effects described above are limited to households with low levels of liquid assets and appear to be financed at least in part by credit card debt. In contrast, to the extent that any spending effect is present among zero-positive households, it appears to be concentrated among households with high levels of liquid assets and is not financed with credit card debt. Finally, all of the aforementioned spending increases are fairly widely spread across multiple categories of spending, with a slight concentration in discretionary goods.

There are several interesting implications of these results on household behavior. First, the results indicate that households fully anticipate and act in advance of their ARM resets. Instead of increasing spending sharply at reset, the increases comes gradually beforehand. This kind of household behavior is certainly forward-looking and has the feel of dynamic optimization. It also implies that households, even those with below-median wealth, are not fully liquidity constrained because they are able to increase spending well in advance of the positive shock.

Second, wealth appears to affect the way that household spending responds to ARM resets, but in a manner that actually depends on the exact nature of the reset. For example, when reset is synonymous with a quasi-permanent income shock as it is for the zero-zero and positive-positive groups, low-wealth households appear much more responsive. When reset simply reflects an adjustment to the mortgage rate with no significant change to permanent income (the zero-positive group), high-wealth households appear more responsive.

Is there a model of consumption that produces this kind of household behavior in the presence of ARM resets? Clearly, models in which agents are very myopic or unsophisticated would not be consistent with the empirical results. Neither would models in which all agents are awash with liquidity and consume according to the permanent income hypothesis. In the next two section, I incorporate ARM resets into a model that fits somewhere in between, i.e. a buffer-stock model, and assess via simulation whether its predictions are consistent with the empirical evidence.

## 1.5 Modified buffer-stock model

The standard buffer-stock model of Zeldes (1989), Deaton (1991), and Carroll (1992) describes how agents make consumption decisions in the face of labor income uncertainty and borrowing constraints. In this section, I develop a modified version of the model in which agents also have to make mortgage payments that are similar to the payments on a hybrid ARM. That is, mortgage payments are fixed until some known date in the future, at which point they reset based on the prevailing market interest rate. Agents in the modified model therefore face two sources of uncertainty: labor income (as before) and post-reset mortgage payments.

In a standard model where labor income is iid each period, the only state variable for an agent is her "cash on-hand", the amount of funds available for consumption through her asset holdings from the previous period and labor income from the current period. However, the modified model with ARM reset has two additional state variables: the current market interest rate and the number of periods until reset occurs. While these variables do not directly impact current cash on-hand, they do affect the distribution of future cash on-hand, which makes them relevant for buffer-stock agents' current consumption.

In accordance with the empirical event study, the model will have three types of agents: zero-zero, zero-positive, and control. The positive-positive group is unmodeled because as shown in Section 1.4, its behavior is quite similar to that of the zero-zero group. In the following subsections, I review the environment and assumptions of the modified buffer-stock model, describe the agent's problem, assign parameter values, and demonstrate how the model is solved numerically.

### 1.5.1 Model environment and assumptions

In this section, I describe the basic environment of the model. The discussion centers around the part of the model that departs from the standard version: mortgage payments. Then, I motivate the principal assumptions of the model and evaluate them in the context of the empirical event study.

#### Environment

Time is indexed in months relative to reset, which occurs at $t = 1$. Each period, agents maximize the sum of an infinite stream of discounted CRRA utility, where $\delta$ is the monthly exponential discount factor and $\rho$ is relative risk aversion. The model begins at -12 (1 year before reset) and is analyzed through +12 (1 year after reset) to match the empirical event study. Monthly labor income $y_t \in \{y_L, y_H\}$ is iid where $\text{Prob}(y_t = y_L) = p_y$. Agents can lend and borrow at a constant risk-free rate $r_f$ but face an exogenous borrowing constraint $b$.

The main departure from the standard buffer-stock model is that in each period, agents have to make a mortgage payment $z_t$. The time profile of payments is similar to that of a 5-year fixed term hybrid ARM that resets at $t = 1$. There are three types of agents: control (c), zero-zero (zz), and zero-positive (zp). The latter two agent types correspond to the classes of ARM borrowers in the empirical portion of the paper (see Section 1.2.1).[11] Before reset ($t < 1$), all three types have interest-only payments equal to $\frac{r_{orig}}{12}B$, where $r_{orig}$ is the initial fixed rate of the hybrid ARM and $B$ is the original outstanding balance. Post-reset payments ($t \geq 1$) for each agent type are given below.

$$z_c = \frac{r_{orig}}{12}B \tag{1.8}$$

$$z_{zz}(r_1) = \frac{r_1 + m}{12}B \tag{1.9}$$

$$z_{zp}(r_1) = \frac{\frac{r_1+m}{12}}{1 - (1 + \frac{r_1+m}{12})^{-300}}B \tag{1.10}$$

where $r_1$ is the value of $r_t$, the market interest rate that ARMs are indexed to (i.e. Libor), that prevails when reset occurs at $t = 1$ and $m$ is the margin that is applied on top of this rate for credit risk. For control agents in (1.8), payments do not change at reset. For zero-zero agents in (1.9), payments remain interest-only but are based on $r_1 + m$ rather than $r_{orig}$. For zero-positive agents in (1.10), payments switch to interest-plus-principal and equal the 300 month (25-year) annuitized value of $B$ with an annualized interest rate of $r_1 + m$.[12] In all cases, post-reset payments do not change again after $t = 1$, which explains why there is no $t$ subscript in (1.8)-(1.10).

The market interest rate $r_t$ evolves as follows. In every period before reset ($t < 1$), $r_t$ either stays the same or goes up by a small amount $\Delta$ with probability $p_r$:

$$\begin{aligned} r_{t+1} &\in \{r_t, r_t + \Delta\} \\ p_r &= \text{Prob}(r_{t+1} = r_t + \Delta) \end{aligned} \tag{1.11}$$

In addition to being simple, this interest rate structure is chosen to roughly correspond to the environment that households in the empirical event study faced in proximity of their ARM resets.

[11] Positive-positive households are not simulated because in Section 1.4.2, their actual behavior was very similar to zero-zero households. The simulated behavior of these two types of households would be very similar too.

[12] The 300 month or 25 year term is because after a 5-year fixed term hybrid ARM resets, the mortgage has a remaining term of 25 years.

Recall from Figure 1.3 that the reset window in the event study is from April 2010 to December 2012. Based on Figure 1.1, it is clear that starting well before the beginning of the reset window, market interest rates were effectively at the zero lower bound (ZLB). In fact, the Federal Reserve set the target Fedfunds rate to zero in December of 2008. Therefore, in the months leading up to their mortgage resets, households were indeed facing an environment in which rates could either stay at the ZLB or go up each period. Notably, rates ultimately stayed at the ZLB throughout the entire reset window (see Figure 1.1), which will inform how the model is simulated in Section 1.6.

**Motivation and evaluation of assumptions**

It is important to motivate the key assumptions of this model and evaluate their consistency with the environment of the empirical event study. First, the way that hybrid ARMs are modeled here is different from how they work in practice. Here, it is assumed that there is one reset date, after which there is no uncertainty in payments. In practice, hybrid ARMs continue to reset periodically after the first occurrence, usually at a frequency of 1 year. Rather than modeling interest rates over a much longer horizon, we simply assume that there is only one reset date. In addition, the model assumes that mortgage payments are made in perpetuity, not just for the typical 30 year term. This assumption is made so that after reset, the agent's problem is stationary over an infinite horizon, and therefore more easily solvable.

Second, the model assumes that agents lend and borrow at the same, constant rate $r_f$. One issue with this assumption is that the lending/borrowing rate is decoupled from $r_t$, the market interest rate that ARM resets are tied to. While this is probably not true in reality, the assumption is needed to avoid modeling the evolution of $r_t$ over the infinite horizon of the model. A second issue with this assumption is that lending and borrowing rates for households are typically very different in practice, particularly if a household's primary means of borrowing is a credit card. The problem with setting the borrowing rate higher is that in order for the buffer-stock model to be numerically solvable, the discount rate must exceed the lending/borrowing rate.[13] Satisfying this condition would require a discount rate that is unrealistically high.

[13]This condition is sufficient for the Bellman functional to be a contraction, which allows for the contraction-mapping theorem to be invoked.

Finally, the model assumes that agents discount utility exponentially, which is at odds with a large body of evidence on individuals having time inconsistent preferences (citations needed). Exponential discounting is chosen for its simplicity.

### 1.5.2 Agent's problem

The dynamic optimization problem for agents of type $i \in \{c, zz, zp\}$ is as follows:

$$\begin{aligned} &\max_{c_{i,t}} E_{t_0}\left[\sum_{t=t_0}^{\infty} \delta^{t-t_0}\left(\frac{c_{i,t}^{1-\rho}-1}{1-\rho}\right)\right] \\ &\text{s.t. } a_{i,t} = (1+r_f)a_{i,t-1} + y_{i,t} - z_{i,t} - c_{i,t} \\ &\quad a_{i,t} \geq b \end{aligned} \tag{1.12}$$

where $t_0$ is each month in the event window ($t_0 \in \{-12, -11, ..., 12\}$) and $c_t$ is consumption in period $t$. Assets held at end of period $t$ are given by $a_{i,t}$ and evolve according to (1.12): assets from the previous period are grossed up by $r_f$, current period income $y_{i,t}$ is added, and the mortgage payment $z_{i,t}$ and consumption $c_{i,t}$ for the current period are subtracted. Assets each period must exceed $b$, the borrowing constraint. The mortgage payment $z_{i,t}$ equals $\frac{r_{orig}}{12}B$ for $t < 1$ and is given by (1.8)-(1.10) for each agent type for $t \geq 1$. The initial conditions for the agent's problem are $a_{i,-13}$ (initial asset holdings) and $r_{-12}$ (the market interest rate 12 months before reset).

### 1.5.3 Parameter selection

To the extent possible, parameters are selected to roughly match the characteristics of households in the empirical event study (see Tables 1.3 and 1.4). For example, agents' monthly labor income process is parameterized by $\{y_L, y_H\} = \{\frac{40{,}000}{12}, \frac{80{,}000}{12}\}$ and $p_y = 0.1$. This means that in the good state of the world with probability 0.9, annual income is \$80,000 and in the bad state with probability 0.1, annual income is \$40,000. In addition, agents' original mortgage balance $B$ is set to \$350,000.

Other parameters are set based on a combination of existing literature, summary statistics from the Bank's data, and intuition. In accordance with (citation needed), agents' monthly discount factor $\delta$ is set to 0.99 and relative risk aversion $\rho$ is set to 2. Based on data from the Bank, the credit-risk margin $m$ that post-reset mortgage payments are based on is set to 3%. The monthly risk-free rate $r_f$ is arbitrarily set to a low value of $\frac{1\%}{12}$. Finally, the borrowing constraint $b$ equals 0, meaning that agents are unable to borrow.

The interest rate at which mortgages are originated ($r_{orig}$) is defined implicitly so that zero-positive agents experience no change in their mortgage payment if the prevailing market interest rate at reset ($r_1$) equals 0. The motivation for this condition is that in the empirical event study, mortgage payments for the zero-positive group change very little at reset (see Figure 1.5), and reset occurs when market interest rates are at the ZLB. The definition of $r_{orig}$ is given in (1.13), where the left-hand side is the pre-reset mortgage payment and the right-hand side is the post-reset mortgage payment from (1.10) with $r_1 = 0$.

$$\frac{r_{orig}}{12} B = \frac{\frac{m}{12} B}{1 - (1 + \frac{m}{12})^{-300}} \tag{1.13}$$

$$\implies r_{orig} = 5.69\%$$

For the evolution of the market interest rate $r_t$, the amount by which rates can go up each period ($\Delta$) is set to 0.25%. The probability that the market rate rises each period ($p_r$) is given three possible values: 0.0, 0.1, and 0.25. The model is solved for multiple values of $p_r$ because this parameter plays an important role in setting agents' pre-reset expectations for how mortgage payments will change at reset. In Section 1.6, I compare simulations of the model for different values of $p_r$, which builds intuition about household behavior under the model.

### 1.5.4 Solving the agent's problem

The agent's problem from Section 1.5.2 is solved in two steps. First, since the problem in every post-reset period ($t \geq 1$) is stationary, it can be solved using standard Bellman iteration techniques. Second, to find the pre-reset value functions, the value function from the first step is used to solve the agent's problem in $t = 0$, which is used to solve the problem in $t = -1$, and so on until $t = -12$.

**Solving the stationary post-reset problem**

In each period after reset, agents of type $i \in \{c, zz, zp\}$ face the same problem. Let $V_{i,1}(a_{t-1}, y_{i,t}, r_1)$ be the post-reset value function for agents of type $i$. $V_{i,1}$ can be defined iteratively as follows:

$$V_{i,1}(a_{i,t-1}, y_{i,t}, r_1) = \max_{c_{i,t}} \{u(c_{i,t}) + \delta E_t[V_{i,1}(a_{i,t}, y_{i,t+1}, r_1)]\} \tag{1.14}$$

$$\text{s.t. } a_{i,t} = (1 + r_f)a_{i,t-1} + y_{i,t} - z_i(r_1) - c_t$$

$$a_{i,t} \geq b$$

$V_{i,1}$ is a function of asset holdings at the end of the previous period ($a_{i,t-1}$), current period labor income ($y_{i,t}$), and the prevailing market interest rate at $t = 1$ ($r_1$). The sole state variable in the problem is asset holdings $a_{t-1}$, whose evolution is determined by $y_{i,t}$ and $r_1$. The effect of $r_1$ is through post-reset mortgage payments $z_i(r_1)$ defined in (1.8)-(1.10).

Since $\delta = 0.990 < 0.999 = \frac{1}{1+r_f}$, the functional defined by (1.14) is a contraction and by the contraction mapping theorem, the value function and policy function $c^*_{i,t}(a_{i,t-1}, y_{i,t}, r_1)$ can be found by numerically iterating the functional until it converges.

**Solving for pre-reset value functions**

In each period before reset ($-12 \leq t \leq 0$), the agent's problem is no longer stationary. There is also an additional state variable: the current interest rate $r_t$. In addition, the pre-reset value function will depend on the value of the parameter $p_r$, the probability that market rates rise each period. Letting $p^j$ index the possible values of $p_r$, the pre-reset value function takes the form $V^j_{i,t}(a_{i,t-1}, y_{i,t}, r_t)$ and is defined below.

$$V^j_{i,t}(a_{i,t-1}, y_{i,t}, r_t) = \max_{c^j_{i,t}} \left\{u(c^j_{i,t}) + \delta E_t\left[V^j_{i,t+1}(a^j_{i,t}, y_{i,t+1}, r_{t+1}) \;\middle|\; p_r = p^j\right]\right\} \tag{1.15}$$

$$\text{s.t. } a^j_{i,t} = (1 + r_t)a_{i,t-1} + y_t - \frac{r_{orig}}{12}B - c^j_{i,t}$$

$$a^j_{i,t} \geq b$$

$$t \in \{0, -1, ..., -12\}$$

There are four key differences between (1.15) and the post-reset value function in (1.14). First, the continuation value in (1.15) is based on next period's value function, highlighting the non-stationarity of the pre-reset problem. This means that in pre-reset periods, two agents that are the same in every respect except for how far away they are from reset will generally make different consumption decisions.

Second, the pre-reset value function in (1.15) is an explicit function of $r_t$. This dependence arises from the fact that the post-reset value function in (1.14) depends on post-reset mortgage payments, which depend on $r_1$. Prior to reset, the distribution of $r_1$ depends on the current value of $r_t$. Therefore, $r_t$ is a state variable because as it changes, it affects agents' perceptions about what their post-reset mortgage payments will be.

Third, the pre-reset value function in (1.15) depends on the parameter $p_r$. This dependence is due to the fact that $p_r$ determines agents' expectations of future interest rates, which affects the nature of the expectations operator on the right side of (1.15). This is why the operator is conditioned on $p_r^j$. Finally, in (1.15) the mortgage payment $z_i$ equals $\frac{r_{orig}}{12}B$ for all agent types because as discussed earlier, all three agents types make interest-only payments pre-reset.

The pre-reset value functions and policy functions ${c_{i,t}^{j}}^{*}(a_{t-1}, y_t, r_t)$ are found by first inserting the post-reset value function $V_{i,1}$ into the right hand side of (1.15) and numerically solving for $V_{i,0}^j$ (noting that $V_{i,1}$ will not depend on $j$). Then, $V_{i,0}^j$ is inserted on the right hand side of (1.15) and $V_{i,-1}^j$ is numerically solved for. This process continues until $V_{i,-12}^j$ is obtained. It should be noted the control agents' problem is stationary even in pre-reset periods since nothing occurs at reset for them. Therefore, the control agents' value and policy functions in all periods are obtained by solving (1.14).

In the next section, we use the policy functions developed in this section to simulate the model and compare it to the results of the empirical event study.

## 1.6 Simulating the model

In Section (1.5), I developed and solved a modified buffer-stock model that describes household consumption behavior in the face of labor income and hybrid ARM reset uncertainty. In this section, I simulate household consumption for a specific realization of market interest rates that attempts to mimic the environment of the empirical event study of Section 1.4. Recall that households in the event study have hybrid ARMs that reset between April 2010 and December 2012, a period before which and during which market interest rates remained consistently at the ZLB. Therefore, the modified buffer-stock model will be simulated under a realization of interest rates of the form $r_{-12} = 0 = r_t$ for all $t \leq 1$. That is, at one year before reset ($t = -12$), market rates equal zero and they remain at zero until reset occurs at $t = 1$. Importantly, agents in the simulation do not know the realized path of interest rates beforehand. They simply know that each period, rates can either remain the same or go up by $\Delta$ with probability $p_r$, as specified by (1.11).

In the remainder of this section, I first go into detail on how the modified buffer-stock model is simulated. I then report the simulation results for zero-zero and zero-positive agents and show that they are very similar to the results of the empirical event study in Section 1.4. Finally, I explain the intuition for why the model produces these results and discuss their implications.

### 1.6.1 Simulation details

The output of Section 1.5 is two sets of policy functions for consumption for agents of type $i \in \{c, zz, zp\}$. First, there are the pre-reset policy functions ${c_{i,t}^{j}}^{*}(a_{i,t-1}, y_{i,t}, r_t)$ for $t \in \{-12, -11, ..., 0\}$, where $j$ indexes values of $p_r$. Second, there are the post-reset policy functions $c_{i,t}^{*}(a_{i,t-1}, y_{i,t}, r_1)$ for $t \in \{1, 2, ..., 12\}$.

The goal is to simulate paths of consumption for zero-zero and zero-positive agents from $t = -12$ to $t = 12$ for different levels of initial asset holdings ($a_{-13} \in \{0, 20000, 40000, 60000, 80000\}$) and different probabilities of interest rate increases ($p_r \in \{0, 0.1, 0.25\}$), relative to control types. The first step is generating 10,000 paths of labor income. Then, for each of these income paths, each of the three values of $p_r$, each of the four values of $a_{-13}$, and the realized path of interest rates ($r_t = 0$ for all $t \leq 1$), I use the policy functions from Section 1.5 to produce simulated paths of consumption from $t = -12$ to $t = 12$ for the zero-zero and zero-positive types.

Each of these consumption paths has an associated "control" path, derived from the control policy functions, which is what a control type consumes in each period with the same level of asset holdings and the same income realization. The control paths are subtracted from their "treatment" paths to produce consumption paths that represent the causal effects of ARM reset. Finally, for each triplet of (agent type, $a_{-13}$, $p_r$), the average consumption path is computed across the 10,000 simulated income paths.

The final outputs of the simulation are therefore average causal effects of ARM reset on consumption in each period of the "event window" from $t = -12$ to $t = 12$ as functions of agent type, initial asset holdings, and the probability of interest rate increases. As was the intention, the interpretation of the simulation output is very similar to the interpretation of the coefficients estimated in the empirical event study in Section 1.4, with one exception. As noted in Section 1.3, the empirical event study coefficients are difference-in-differences (DD) estimates - the coefficients represent the causal effects of ARM reset on spending in a particular event month, relative to 12 months before reset. In contrast, the simulation output does not have the second differencing. While it can be put in DD form, doing so would discard potentially valuable information. Therefore, the simulation output will be viewed in both single and double-differenced forms.

### 1.6.2 Simulation results

The simulated causal effects of ARM reset on consumption for zero-zero and zero-positive agents are reported in Figures 1.19-1.22. The results for each type of agent are discussed in detail below.

**Zero-zero agents**

Figure 1.19 reports the single-differenced simulation results for zero-zero agents while Figure 1.20 reports the double-differenced results. The three plots in each figure represent different values of $p_r$ and within each plot, each line represents different values of initial wealth.

The main conclusion from these results is that for a given value of $p_r$, consumption rises gradually from $t = -12$ to $t = 1$ (the pre-reset periods), exhibits no sharp increase at reset, and is stable thereafter. Moreover, this effect gets stronger as initial agents' wealth gets smaller (comparing the green line to the blue line). These features of the double-differenced simulation in Figure 1.20 are strikingly consistent with the results of the empirical event study for zero-zero households in Figures 1.7 and 1.9.

Another thing to note is that in the single-differenced simulation results of Figure 1.19, all of the lines for wealthy households are well above zero at 12 months before reset ($t = -12$), meaning that consumption is significantly higher than control. This is not the case in the double-differenced results of Figure 1.20 though, where the lines for wealthy households are exactly at zero at $t = -12$ for mechanical reasons. This highlights one limitation of using a difference-in-differences (DD) approach, as was done in the empirical event study of this paper. For example, if the single-differenced consumption of wealthy agents mirrors the green line in Figure 1.19 with $p_r = 0$, a DD framework would show a very small effect for these households, as in Figure 1.20. But in reality, the consumption of these agents is uniformly higher in the pre-reset window.

One interesting feature of the double-differenced results in Figure 1.20 is that as the probability of interest rate increases goes up (i.e. moving from left to right in the plots), the amount by which consumption increases during the pre-reset periods becomes larger for wealthy agents only. For example, poor agents (the blue lines) build up to a double difference of \$825 in pre-reset periods for all values of $p_r$. In contrast, the double difference that wealthy agents (the green lines) build up to is higher when $p_r$ is higher. This same feature can be observed in the single-differenced results in Figure 1.19. Here, wealthy agents (the green lines) build up consumption to roughly \$900 at

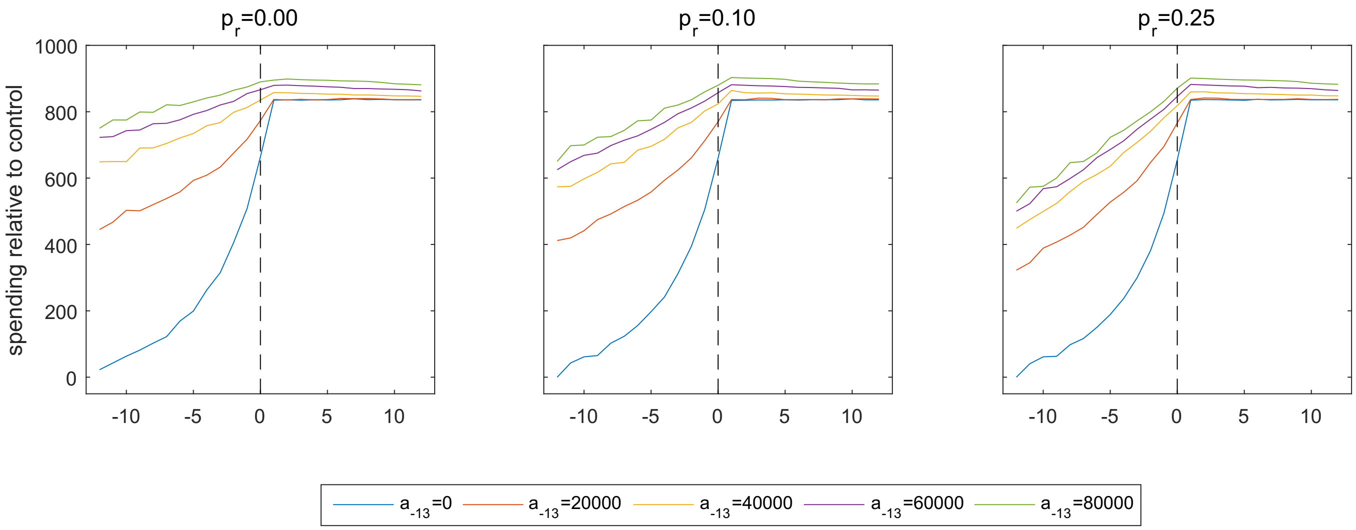



**Figure 1.19:** *Simulation output for zero-zero agents (single difference)*

$t = 1$ for all values of $p_r$ but start at a lower level when $p_r$ is higher. While this aspect of the simulation results is not directly translatable to the empirical event study, we shall see in Section 1.6.3 that it is related to wealthy buffer-stock agents acting like permanent income consumers.

### Zero-positive agents

The simulated causal effects of ARM reset on consumption for zero-positive agents are given in Figures 1.21 and 1.22. The first thing to note is that when $p_r = 0$, there is no difference in consumption relative to control in any time period. This is because according to (1.13), the mortgage payments of zero-positive agents do not change at all if interest rates equal zero at reset. This is indeed the case for the realized path of interest rates in the simulation ($r_t = 0$ for all $t \leq 1$). Furthermore, when $p_r = 0$ there is no chance of interest rates rising during the pre-reset periods, meaning that agents are guaranteed a post-reset rate of zero. Therefore, when $p_r = 0$ zero-positive agents know for sure in every pre-reset period that their payments will not change at reset, so there is no reason for there to be any difference in consumption relative to control.

Second, the single-differenced simulation results of Figure 1.21 are always weakly below zero before reset, meaning that simulated consumption for zero-positive agents is never higher than

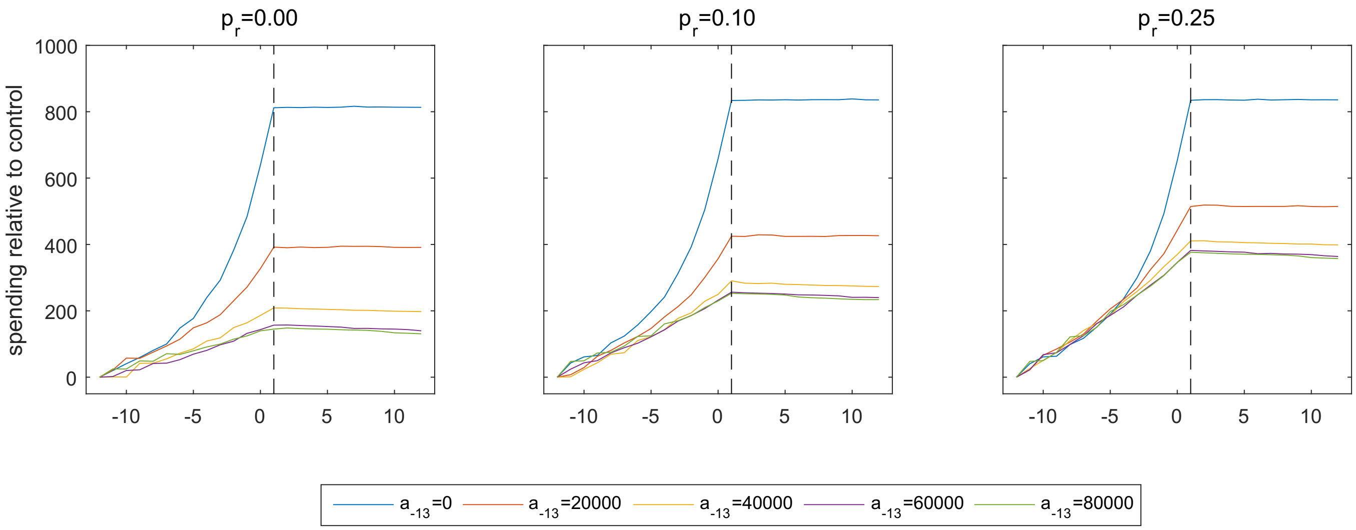



**Figure 1.20:** *Simulation output for zero-zero agents (double difference)*

control. This result, which cannot be inferred from the double-differenced results of Figure 1.22, is because the best case scenario for zero-positive agents is that interest rates at reset equal zero and mortgage payments remain the same. If rates get any higher, mortgage payments go up. As a result, zero-positive agents will never consume more than control agents, whose payments do not change at reset. However, as reset approaches and the prospect of mortgage payments remaining the same becomes more certain, zero-positive consumption converges to the control.

The main conclusion from Figures 1.21 and 1.22 is that when $p_r > 0$, the effect of reset on consumption is quite small relative to that of zero-zero agents and to the extent that there is an effect, it appears to be concentrated among wealthier households . For example, in the double-differenced results in Figure 1.22 with $p_r = 0.1$, the maximum consumption increase between $t = -12$ and $t = 1$ is roughly \$75, compared to a maximum increase of roughly \$900 for zero-zero agents in Figure 1.20. In addition, the effect on zero-positive agents in Figure 1.22 is stronger for agents with higher levels of initial wealth (the green lines vs the blue lines), particularly when $p_r$ is larger. This feature of the simulation (a small effect that is concentrated among wealthy agents) is also remarkably consistent with the results of the empirical event study for zero-positive households in Figures 1.12 and 1.14.

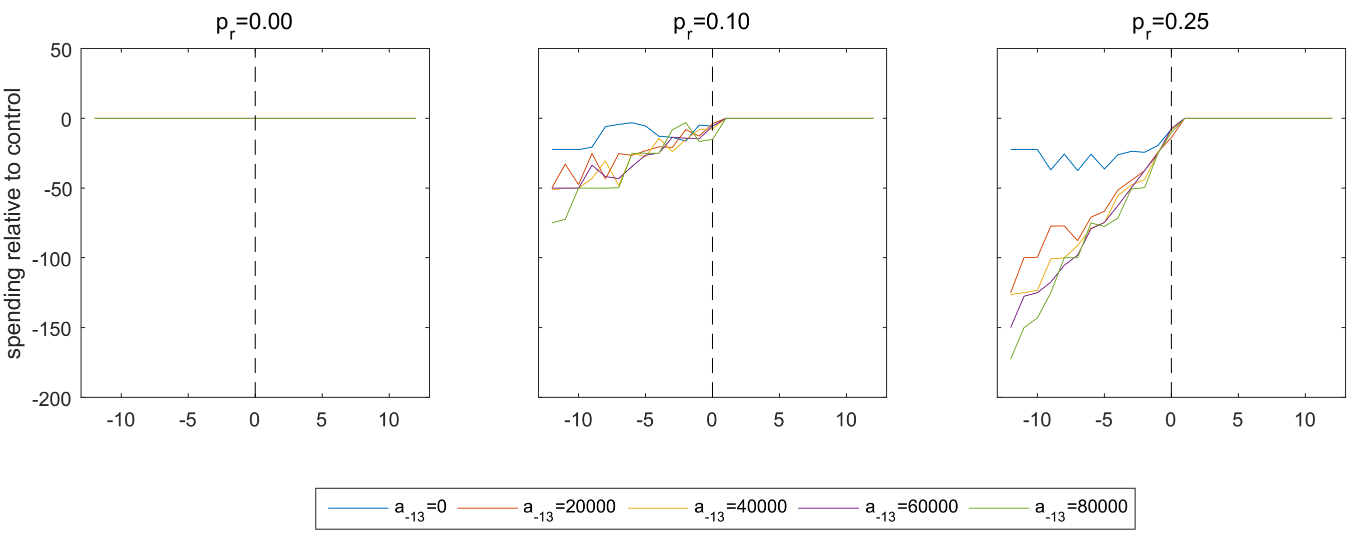



**Figure 1.21:** *Simulation output for zero-positive agents (single difference)*

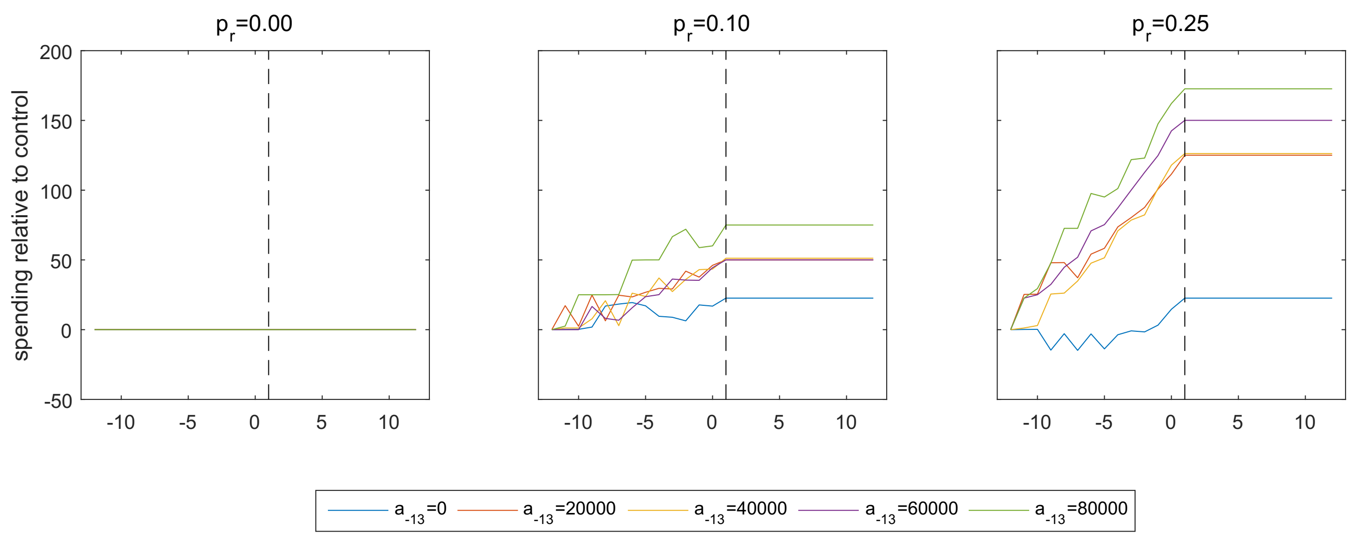



**Figure 1.22:** *Simulation output for zero-positive agents (double difference)*

### 1.6.3 Explanation and discussion of results

In the previous section, we observed that the simulated causal effects of ARM reset on consumption for zero-zero and zero-positive agents are very similar to the empirical event study results of Section 1.4. It appears that the modified buffer stock model of Section 1.5 describes the behavior of households fairly well in the context of ARM resets. In this section, I explain the mechanisms of the model that produce the results in Figures 1.19-1.22 and discuss their broader implications.

First, it is useful to define the random variable $\tilde{r}_{1,t}$, the interest rate that prevails at reset ($t = 1$) as a function of the current interest rate $r_t$. Note that (1.16) is derived from the interest rate process detailed in (1.11).

$$\tilde{r}_{1,t} = r_t + \Delta \sum_{k=t}^{0} \mathbf{1}\{r_{k+1} = r_k + \Delta\} \tag{1.16}$$

Using (1.8)-(1.10) and (1.16), we can define another random variable $\tilde{X}_{i,t}$, the amount that mortgage payments fall at reset for agents of type $i \in \{c, zz, zp\}$.

$$\tilde{X}_{i,t} = -\left(z_i(\tilde{r}_{1,t}) - \frac{r_{orig}}{12}B\right) \tag{1.17}$$

The value of (1.17) is shown for each agent type below. Note that $\tilde{X}_{c,t} = 0$ by construction because control agents experience no change in payments at reset.

$$\tilde{X}_{zz,t} = \frac{B}{12}(r_{orig} - (\tilde{r}_{1,t} + m)) \tag{1.18}$$

$$\tilde{X}_{zp,t} = \frac{B}{12}\left[r_{orig} - \frac{\tilde{r}_{1,t} + m}{1 - (1 + \frac{\tilde{r}_{1,t}+m}{12})^{-300}}\right] \tag{1.19}$$

With this framework, we can precisely identify what agents expect to happen at reset when they are in the pre-reset window. Recall that in the simulation, all agents face a market interest rate realization of zero in every period. Therefore, agents' expected payment reductions as of each pre-reset period can be determined by taking expectations of (1.18) and (1.19) conditional on $r_t = 0$.

We can then describe how these expectations affect dynamic consumption decisions in a modified buffer-stock model and how these decisions are influenced by wealth. We will see that the consumption of poor agents will be highly sensitive to the perceived risk of their asset buffers being depleted by labor income shocks. In contrast, wealthy agents have such large buffers that they do not need to worry about asset depletion - they act like they are permanent income consumers. This analysis is carried out for each agent type below.

### Zero-zero agents

The expected payment reduction at reset conditional on $r_t = 0$ for zero-zero agents is given in (1.20), using the definition of $\tilde{X}_{zz,t}$ in (1.18) and $\tilde{r}_{1,t}$ in (1.16).

$$E(\tilde{X}_{zz,t} \mid r_t = 0) = \frac{B}{12}(r_{orig} - (m + (1-t)p_r\Delta)) \tag{1.20}$$

Using the parameter values for $B$, $r_{orig}$, and $m$ and the assumption that $0 \leq p_r \leq 0.25$ from Section 1.5.3, we have

$$\begin{aligned} & E(\tilde{X}_{zz,t} \mid r_t = 0) \approx 785 - 73(1-t)p_r \\ \Longrightarrow \;& E(\tilde{X}_{zz,t} \mid r_t = 0) \in [785 - 18(1-t), 785] \end{aligned} \tag{1.21}$$

Expression (1.21) shows that in pre-reset periods, zero-zero agents expect a relatively large mortgage payment reduction at reset, even with the probability of interest rate increases ($p_r$) at its maximum value of 0.25. This is true going back 12 months before reset ($t = -12$) as well, when the minimum expected payment reduction according to (1.21) is $785 - 18(13) = \$548$. This is unsurprising: prior to reset, agents that make interest-only payments before and after reset will expect large reductions in payments at reset when current interest rates are zero.

The fact that $E(\tilde{X}_{zz,t} \mid r_t = 0)$ is large for each $t < 1$ is central to why there is a gradual increase in pre-reset consumption for zero-zero agents in Figures 1.19 and 1.20. The large expected payment reduction means that prior to reset, zero-zero agents expect that after reset, their asset buffers will have less of a chance of being depleted by negative labor income shocks. This is because when mortgage payments are much smaller, it is easier to maintain consumption in the face of a negative labor income shock without drawing too heavily on assets. Negative labor income shocks still pose a serious threat before reset occurs, but this threat naturally subsides with the passage of time. As each period passes, there is less of a chance that a negative labor income shock occurs before reset, as there is simply less time until reset.

Therefore, as reset approaches and agents get closer to being "in the clear", they require less of a precautionary buffer and can steadily increase consumption. This effect should be most prominent for agents whose precautionary buffer is of first-order importance, i.e. poor agents. This is precisely what is observed in the empirical (Figures 1.7 and 1.9) and double-differenced simulated (Figure 1.20) results for the zero-zero group: a gradual increase in consumption leading up to reset, with the effect being stronger among agents with lower initial wealth. Despite being relatively poor, households still have enough liquidity to increase consumption relative to control in the pre-reset periods, before their mortgage payments are actually reduced.

What about agents with high levels of wealth? The effect described above does not apply to them because they have a comfortable asset buffer. In other words, wealthy agents are already "in the clear" before reset occurs. This is related to the point that in standard buffer-stock models, wealthy agents effectively consume their expected permanent incomes each period. In the modified model, since mortgage payment reductions constitute a permanent shock, wealthy agents will immediately (at $t = -12$) draw down on their assets and increase consumption by roughly the amount that they expect mortgage payments to decline at reset. As time passes, wealthy agents will continue adjusting consumption by however much their expected payment reductions at reset change.

This intuitive description of the behavior of wealthy agents can be related to the single-differenced simulation results in Figure 1.19 by differentiating the expected payment reduction in (1.20) with respect to $p_r$ and $t$.

$$\frac{\partial E(\tilde{X}_{zz,t} \mid r_t = 0)}{\partial p_r} = -\frac{B}{12}(1-t)\Delta < 0 \tag{1.22}$$

$$\frac{\partial E(\tilde{X}_{zz,t} \mid r_t = 0)}{\partial t} = \frac{B}{12}\left[p_r\Delta - \frac{\partial r_t}{\partial t}\right]$$

$$\implies \frac{\partial^2 E(\tilde{X}_{zz,t} \mid r_t = 0, \frac{\partial r_t}{\partial t} = 0)}{\partial p_r \partial t} = \frac{B}{12}\Delta > 0 \tag{1.23}$$

Expression (1.22) shows that as $p_r$ gets lower, the expected payment reduction at reset becomes higher. This makes sense, as a lower $p_r$ means that it is more likely that the prevailing interest rate and mortgage payments at reset are low. The implication is that as $p_r$ falls, agents' expected permanent income rises. This is why going from right to left in Figure 1.19, wealthy agents (the green lines) consume more relative to control in any given event month when $p_r$ is lower.

Expression (1.23) shows that as time passes and interest rates remain the same ($\frac{\partial r_t}{\partial t} = 0$, as is the case in the simulation), the expected payment reduction grows by more when $p_r$ is larger. This also makes sense - if the perceived probability of an interest rate increase is extremely low, an occurrence of rates staying the same is very much consistent with agents' beliefs and so expected payment reductions remain relatively unchanged. On the other hand, if the perceived probability of an interest rate increase is higher, an occurrence of rates staying the same is more surprising and expected payment reductions must be revised upward. This is why going from left to right in Figure 1.19, wealthy agents increase consumption by more between $t = -12$ and $t = 1$ as $p_r$ gets higher.

#### Zero-positive agents

For zero-positive agents, the main finding from the double-differenced simulation results in Figure 1.22 is that when $p_r > 0$, the effect of reset on consumption is quite small relative to that of zero-zero agents and to the extent that there is an effect, it appears to be concentrated among households with higher levels of initial wealth (the green lines as opposed to the blue lines). To explain this result, it is useful to first formally show that for zero-positive agents, the expected payment reduction at reset conditional on current rates being zero is weakly negative, i.e. $E(\tilde{X}_{zp,t} \mid r_t = 0) \leq 0$.

$$0 = \frac{B}{12}\left(r_{orig} - \frac{m}{1-(1+\frac{m}{12})^{-300}}\right) \tag{1.24}$$

$$\geq \frac{B}{12}\left(r_{orig} - E\left[\frac{\tilde{r}_{1,t}+m}{1-(1+\frac{\tilde{r}_{1,t}+m}{12})^{-300}} \,\middle|\, r_t = 0\right]\right) \text{ if } \tilde{r}_{1,t} \geq 0 \tag{1.25}$$

$$= E(\tilde{X}_{zp,t} \mid r_t = 0) \tag{1.26}$$

where (1.24) uses the definition of $r_{orig}$ in (1.13), (1.25) uses

$$\frac{\partial}{\partial \alpha}\left[\frac{\alpha+m}{1-(1+\frac{\alpha+m}{12})^{-300}}\right]>0$$

and (1.26) uses the definition of $\tilde{X}_{zp,t}$ in (1.19). The result that expected payment reduction is weakly negative is the reason why the lines in the single-differenced simulation results in Figure 1.21 are always below zero. If agents only expect payments to rise, there is no reason for them to consume more than control agents.

But why is there not a large pre-reset increase in the double-differenced simulation results in Figure 1.22, as there is in Figure 1.20 for zero-zero agents? Recall that poor zero-zero agents increase consumption because their precautionary motives get lower as they get closer to being "in the clear" at reset, when mortgage payments are expected to fall by a large amount. However, zero-positive agents face a fundamentally different situation. According to (1.26), payments for zero-positive agents are always expected to rise at reset. In fact, the best-case scenario is for interest rates to remain at zero, in which case mortgage payments remain the same. Zero-positive agents do not become "in the clear" at reset and so prior to reset, they have the same precautionary motives as the control group.

Therefore, if any effect exists for zero-positive agents between $t=-12$ and $t=1$, it must come from agents for whom the precautionary motive is not a major factor in determining consumption, i.e. wealthy agents. As noted earlier, the consumption of wealthy agents in the modified buffer-stock model only responds to changes in expected payment reductions at reset. While $E(\tilde{X}_{zp,t} \mid r_t=0)$ cannot be computed analytically from (1.19) for zero-positive agents as it could for zero-zero agents, it can be verified by Monte Carlo simulation that $E(\tilde{X}_{zp,t} \mid r_t=0)$ has similar properties as the expected payment reduction for zero-zero agents, namely the derivatives in (1.22) and (1.23).[14]

This means that wealthy zero-positive agents should behave similarly to wealthy zero-zero

[14]These are verified by Monte Carlo simulation for $0 \leq p_r \leq 0.25$ and $-12 \leq t \leq 1$.

agents between from $t = -12$ to $t = 1$. Specifically, if $p_r$ is higher, the expected payment reduction at reset will be more negative for zero-positive agents. But as time passes and interest rates remain the same, the expected payment reduction approaches its upper bound of zero more quickly when $p_r$ is larger. This explains the behavior of wealthy agents in the single-differenced simulation results in Figure 1.21. Consumption is below control by an amount that increases with $p_r$ but as time passes, consumption gradually converges to that of the control group.

Note that a DD framework can also lead to mistaken conclusions for the zero-positive group. For example, the double-differenced simulation results in Figure 1.22 and the DD event study results in Figure 1.14 could be interpreted as wealthy agents spending more than control agents in the pre-reset periods, with the gap widening with time. However, these double-differenced results could actually arise from single-differenced results that mirror Figure 1.21, which shows wealthy agents actually consuming less than control, with the gap narrowing with time.

## 1.7 Summary of results and conclusion

In this paper, I explore how household credit card spending responds to a predictable, uncertain, quasi-permanent income shock in the form of a hybrid ARM reset. The main empirical finding is that spending rises gradually in advance of reset but is relatively smooth in the month of reset itself and stable thereafter. The implication is that households not only anticipate their resets but also have sufficient liquidity to increase spending in advance of it. These results are substantially different from prior work, namely Di Maggio *et al.* (2015), who find that auto spending increases sharply in the month of reset and very little beforehand.

I also show that the results of the empirical event study have a theoretical basis. In a modified buffer-stock model that incorporates hybrid ARM payments, the simulated behavior of agents in the model is remarkably consistent with the results of the empirical event study. In the simulation, agents with high levels of wealth behave like permanent income consumers. Knowing that a large expected payment reduction is coming at reset, wealthy agents draw on their assets to finance higher consumption from the very beginning of the event window. In addition, they gradually increase their spending by a small amount during the pre-reset window, as their expected mortgage payment reductions at reset get slightly larger when interest rates remain flat over time.

In contrast, agents with low wealth do not have enough liquidity to consume their expected permanent incomes one year before reset occurs. But they do have enough liquidity to increase spending gradually (and in aggregate, dramatically) over the entire pre-reset window. As each month passes, there is less and less of a concern that their limited asset buffers will be wiped out by a labor income shock before reset. Moreover, after reset, they can be less worried about these kinds of shocks in general because mortgage payments will be substantially and permanently lower. This reasoning allows them to do increasingly less precautionary saving (increasingly more consumption) in the pre-reset window.

The first main point to take away from this paper is that households, even those with below-median wealth, appear to be quite sophisticated in their spending responses to income shocks. Even though the sample studied in this paper is slightly biased toward households that may be more sophisticated to begin with, this conclusion is still at odds with the large body of evidence that households seem to sharply increase spending precisely when they are hit by other kinds of income shocks. However, these shocks are generally temporary, small, and non-recurring. If poor households follow a buffer-stock model, it may actually be optimal for them to ignore these kinds of shocks until they actually occur. Unlike an ARM reset, these shocks will not have a very lasting impact on poor households' ability to weather labor income shocks. Therefore, it is possible for households to be rational in both ignoring these shocks until they occur while using their liquidity to spend in advance of larger, more permanent shocks.

The second key point is related to the use of ARM resets to assess the income channel of monetary policy. This paper offers evidence that ARM resets are anticipated and acted upon in advance by households. If this is the case, then event studies of ARM resets (as in this paper) will not necessarily give reliable estimates of the effects of monetary policy *surprises*. This is because households may respond differently to anticipated and unanticipated shocks. Indeed, in a buffer-stock model, the spending of wealthy households may be *more* responsive to surprise income shocks than that of poorer households as the latter would use some of their windfall to bolster their asset buffers.

Nevertheless, the event study results of this paper are still useful for evaluating monetary policy because they suggest that wealthy households with mortgages respond to changes in their expected future payments. Therefore, this paper suggests that the income channel can operate through an "expectations" effect. Specifically, if wealthy households learn that it is more likely for interest rates to fall in the future, they will have higher expectations of permanent income and will increase their current spending.

# Chapter 2

# How is the likelihood of fire sales in a crisis affected by the interaction of various bank regulations?[1]

## 2.1 Introduction

It is commonly believed that the financial crisis of 2008 was made significantly worse by banks engaging in fire sales of risky assets (Brunnermeier, 2009). These fire sales are generally attributed to two sources. First, short-term creditors of banks refused to roll over their loans, forcing banks to sell assets in order to repay these loans (Shleifer and Vishny, 2011). Second, after an initial shock led banks to suffer losses on their holdings, banks sold assets in order to recapitalize (Hanson *et al.*, 2011). Though these two explanations for fire sales are well-accepted, there is little theoretical work on the optimal course of action for banks facing creditor withdrawals or with insufficient capital. One could argue that in the former situation, expediency might take precedence over optimality. Following a shock, however, banks generally have a longer horizon over which to recapitalize, making this situation suitable for a theoretical model.

In this paper, we model the recapitalization process in a setting where banks face a (potentially

[1]Co-authored with Divya Kirti

risk weighted) regulatory capital requirement. After experiencing an initial shock that causes them to fall short of the requirement, banks choose the optimal combination of asset sales and equity issuance that restores their capital ratio.[2] We then analyze how bank behavior is impacted when additional regulations, such as liquidity requirements and mandatory equity issuance, are in place. It is important to think about how regulations put in place to solve different problems interact with each other.[3] Can high risk weights be counterproductive? Can liquidity requirements interact with capital requirements in a harmful manner?

We assume that banks are risk-neutral, act in the interests of their existing shareholders, and that all assets are priced in a risk-neutral manner. Under these assumptions, we show that the optimal bank choice is shaped solely by risk-shifting motives. Intuitively, as banks act on behalf of shareholders with limited liability, actions that allow them to retain more risk, while still satisfying regulatory requirements, are desirable because value can be transferred from creditors to shareholders. Importantly, we assume that in the absence of shocks, risk-shifting motives play a very limited role in bank decision-making. In particular, the composition of the asset side of bank balance sheets is taken to be exogenous in our model before the shock takes place.

While these assumptions are admittedly strong, the purpose of our model is simply to show how risk-shifting can influence banks' recapitalization decisions in the midst of a crisis. We argue that such analysis is useful to the extent that shareholder value maximization, which is the root of risk-shifting, is an important consideration for banks. Consistent with the risk shifting view, weaker banks in the Euro area concentrated their balance sheets into domestic sovereign debt following the Euro area crisis (Acharya and Steffen, 2015; Crosignani, 2015).[4] Prior work on the crisis, including Hanson *et al.* (2011) and French *et al.* (2010), uses debt overhang as the overarching

[2]We assume that the calculation of a bank's capital ratio involves marking-to-market the value of its assets. According to the new Basel III regulations, this is the case for assets designated as "trading" and "available for sale" but not "held to maturity". The first two buckets make up over 50% of total bank assets (source).

[3]The post crisis regulatory agenda includes many separate regulatory thrusts. Andy Haldane listed ten separate areas at a speech given at a 2015 conference (slides here). These different areas have largely been treated as independent or complementary of each other.

[4]While risk shifting is one explanation for this pattern, high expected returns following fire sales (Shleifer and Vishny, 2011; Diamond and Rajan, 2011) and financial repression (Becker and Ivashina, 2014) might also be relevant.

framework, which implicitly makes the same assumption. There is also a literature on risk shifting by banks in more 'normal' circumstances (Stiglitz and Weiss, 1981; Hellmann *et al.*, 2000; Acharya and Viswanathan, 2011; Dell'Ariccia *et al.*, 2016).[5] In addition, it is precisely in the midst of crises, when the probability of insolvency is nontrivial, that shareholders have the greatest ability to shift risk on to creditors. Similarly, in the absence of a crisis, shareholders are almost fully exposed to any risks that they take because the probability of insolvency is so low. Therefore, it seems appropriate for banks to incorporate risk-shifting into their post-shock recapitalization plans while focusing on other factors before the shock occurs.

Having established that the bank recapitalization process is shaped by risk-shifting motives, we then solve the model. When the banks experience a shock and face risk-weighted capital requirements, banks choose to recapitalize by concentrating their holdings into one particular asset and shedding the others. The "desired" asset is selected based on two criteria: the underlying risk of the asset's return and the asset's risk weight. This is a manifestation of risk-shifting and resembles regulatory arbitrage: banks choose to concentrate their portfolios in the asset that provides the greatest risk relative to the amount of capital that must be held against it.

This result leads to the first main result of our paper: when banks recapitalize after a shock, the choice of risk weights in the capital requirement strongly affects whether fire sales[6] can occur in equilibrium. If all risk weights are identical, as in a simple leverage ratio requirement, then the optimal choice for banks is to sell relatively safe assets. However, if risky assets are given sufficiently high risk weights, banks will find it optimal to sell these assets in a fire sale. This result is driven by banks' risk-shifting motives. If a risky asset has a high risk weight, retaining this asset requires banks to hold relatively more capital for a given value of total assets, which represents a transfer of value from shareholders to creditors. For a high enough risk weight,

[5] Bahaj and Malherbe (2016) provide empirical evidence on banks' responses to changes in capital requirements consistent with a model based on maximizing value for existing shareholders.

[6] The term fire sale is generally used for assets that have substantial illiquidity such that if a large quantity was sold, the price would drop substantially. An example would be subprime MBS. On the other hand, if a large quantity of a relatively safe, liquid asset like GSE debt was sold, there would likely be little price impact. Therefore, we do not think of the latter situation as a fire sale.

banks would rather sell the risky asset and retain a lower risk-weight asset, even though it is safer, because it does not require as much capital to be held.

In contrast, when risk weights are identical, all assets have the same capital charge regardless of riskiness. In this case, risk-shifting motives lead to a simple choice: retain the risky assets and sell the relatively safer ones. These results are pertinent in light of the recently adopted Basel III capital regulations. One component of the regulations is a simple leverage ratio requirement, which lowers the risk of fire sales according to our model. However, the regulations also modify the existing regime by assigning higher risk weights to a variety of risky assets, which arguably raises the risk of fire sales.

Our second result is that if, in addition to a capital requirement, banks face a liquidity requirement that requires them to retain a minimum amount of "safe" assets, banks have an incentive to become large in scale and could be pushed toward a fire sale of risky assets in response to a shock. The rationale for the scale is that liquidity requirements force banks to hold assets they would otherwise want to sell. Banks make up for this by building up holdings of desired assets to the maximum extent possible, to dilute the holdings of the undesired assets.

Suppose that in the absence of the liquidity requirement, banks prefer to sell the relatively safe assets and retain risky assets. The liquidity requirement can be interpreted as diminishing the appeal of this action. Banks want to hold a concentrated portfolio of risky assets, but the liquidity requirement forces them to hold a diversified portfolio of risky and safe assets, which is costly when there is a risk-shifting motive. Alternatively, the bank can recapitalize by selling risky assets, which are likely not subject to liquidity requirements. Though this action involves holding a concentrated portfolio of safer assets, this portfolio could still be more appealing than a diversified portfolio of the risky and safe assets, especially if the liquidity requirement is strict (more diversification required) and risky assets have high risk weights (making them less desirable). This result is particularly important given that Basel III introduces enhanced liquidity requirements to be implemented alongside stricter risk-weighted capital requirements.[7]

Our first two results suggest that undercapitalized banks may respond to a shock by engaging in fire sales of risky assets. One common view is that these fire sales can be averted by forcing banks to issue equity (Hanson *et al.*, 2011), the idea being that equity issuance recapitalizes banks while rendering it unnecessary to sell assets. Our third result is that there is theoretical justification for mandatory equity issuance. In the absence of the mandate, banks wish to concentrate their portfolios in one asset and dispose of the others to the maximum extent possible given liquidity requirements. Mandatory equity issuance forces banks to hold assets they would otherwise sell, which could potentially prevent fire sales of illiquid assets. However, if the constraint is not binding for banks, fire sales may still occur, suggesting that mandatory equity issuance amounts should be aggressive.

Collectively, our results suggest that the assignment of risk-weights can be an important determinant of how banks choose to recapitalize in a crisis. While assigning high risk weights to risky, illiquid assets may have favorable ex-ante incentives, doing so might generate unintended ex-post incentives for banks to engage in fire sales, particular if liquidity requirements are in place.[8] Moreover, cyclical risk weights (CRWs) could be a powerful tool in the midst of a crisis, potentially more powerful than cyclical capital requirements (CCRs). In addition to implicitly lowering the capital requirement, lowering risk weights on certain risky, illiquid assets makes it more worthwhile for banks to retain these assets and sell safe ones, for the same risk-shifting related reasons described earlier. Therefore, CRWs offer additional ammunition against the risk of fire sales compared to CCRs.

Our approach contributes to the literature by studying the joint impact of multiple types of bank regulation in a setting where banks adjust their portfolio allocation to multiple assets. Much

[7]Our model does not include incompleteness in the market for insurance against aggregate risks or other features that could make liquidity requirements optimal (Allen, 2014). We simply highlight potential negative side effects of combining capital and liquidity requirements.

[8]Similarly, exposure limits with respect to risky assets, or risk weights that increase with concentration on the balance sheet, might have good ex-ante properties but generate bad ex-post incentives.

of the literature of interactions between capital requirements and other policy tools focuses on the interaction of capital or macruprudential regulation with monetary policy. See IMF (2013) and BOE (2015) for a discussion of the literature in this area. Closer to our work, Walther (2016) and Goodhart *et al.* (2013) study the effects and design of capital and other types of bank regulation, including liquidity requirements. Both of these papers abstract away from multiple asset classes. Similarly, work on the role of cyclical capital requirements typically focuses on the time dimension rather than the cross section of assets (Kashyap and Stein, 2004; Repullo and Suarez, 2013). We emphasize the impact of the interaction of various types of bank regulation on the tradeoffs between multiple assets.

The paper proceeds as follows. In Section 2.2, we describe a basic version of the model in order to build intuition about the recapitalization problem that banks face in a crisis. In Section 2.3, we lay out a more generalized and detailed version of the model. In Section 2.4, we use the generalized model in Section 2.3 to prove various propositions related to our main results. Section 2.5 concludes.

## 2.2 Basic Model

In this section, we describe a very basic version of the model in order to show why banks are driven by risk-shifting motives when deciding how to recapitalize in a crisis. The analysis of the banks' actual decisions will be left for the more generalized model in the Section 2.3.

### 2.2.1 Setup and Assumptions

There are three periods: 0, 1, and 2. In period 0, a representative bank holds assets $A_0$, debt with face value $D_0$ that matures in period 2, and adequate regulatory capital. In period 1, asset fundamental value is shocked to $A_1$, after which some combination of asset sales and equity issuance is done in order to restore the capital ratio. Negative asset sales (i.e. asset purchases) and negative equity issuance (equity repurchases) are permitted. After these actions, bank assets are $A_{1,post}$ (implying asset sales were $A_1 - A_{1,post}$) and the face value of debt that remains until period 2 is $D_2$. In period 2, the value of assets evolves from $A_{1,post}$ to $A_2$, where the latter is uncertain.

Assets are then liquidated and debt/equity holders are paid accordingly.

We assume that all assets (including debt and equity) are priced fairly for a risk neutral investor with a discount rate of zero. This assumption implies that $E(A_2) = A_{1,post}$. In addition, the proceeds of asset sales and equity issuance in period 1 are used to buy back debt. The price at which \$1 of debt is bought back in period 1 ($\beta$) must equal the value of the debt per unit face value that remains after the action is taken:

$$\beta = \frac{E\left(\min\left[A_2, D_2\right]\right)}{D_2} \tag{2.1}$$

Finally, in period 1 banks choose the combination of equity issuance and asset sales that maximizes the expected payoff to the existing shareholders, the only restriction on asset sales being that the bank cannot short any asset.

### 2.2.2 Objective function

In period 1, the bank chooses equity issuance ($e$) and asset sales ($A_1 - A_{1,post}$) to maximize the expected payoff to existing shareholders in period 2, subject to restoring its capital ratio. Therefore, the objective function is

$$\begin{aligned}
&\left(1 - \frac{e}{E\left(\max\left[A_2 - D_2, 0\right]\right)}\right) E\left(\max\left[A_2 - D_2, 0\right]\right) \\
=&E\left(\max\left[A_2 - D_2, 0\right]\right) - e \qquad (2.2) \\
=&E\left(A_2 - \min\left[A_2, D_2\right]\right) - e \\
=&A_{1,post} - \beta D_2 - e \qquad (2.3)
\end{aligned}$$

where the last equality uses the definition of $\beta$ in (2.1). We will show that this objective function is equivalent to minimizing the price of debt $\beta$.

Since debt is bought back at a price of $\beta$ using the proceeds of asset sales and equity issuance, the following is true:

$$D_0 - D_2 = \frac{1}{\beta}\left(\left(A_1 - A_{1,post}\right) + e\right) \tag{2.4}$$

Substituting for $D_2$ into the objective (2.3), we get

$$\begin{aligned} & A_{1,post} - \beta\left(D_0 - \frac{1}{\beta}\left((A_1 - A_{1,post}) + e\right)\right) - e \\ =& A_1 - \beta D_0 \end{aligned} \tag{2.5}$$

Note that $A_1$, the fundamental value of the assets after the shock, and $D_0$, the face value of the debt in period 0, do not depend on the specific action taken. As a result, the objective function is equivalent to minimizing $\beta$, the price of the debt outstanding after the action (also the price at which the debt is repaid in period 1).

This version of the objective in (2.5) characterizes the risk shifting problem. Since all assets are priced at fair value, there is no NPV to be gained by taking any specific action. However, different actions can result in different transfers from shareholders to creditors. Shareholders want to choose the action that minimizes this transfer, which is equivalent to minimizing the price of the remaining debt, $\beta$. There is a risk-shifting motive: shareholders want to take as much risk as possible because the creditors absorb the downside risk. This result extends to the more general version of the model described in the next section but is perhaps more easily seen in this simple setting. Banks want to pursue a recapitalization strategy that minimizes the price of its remaining debt per unit face value.

## 2.3 Generalized model

In this section, we lay out a more generalized version of the model described in the previous section. This model will allow for analysis of banks' optimal recapitalization plans and will be used to establish the main results of the paper.

### 2.3.1 Setup

There are three periods: 0, 1, and 2. In period 0, there is a continuum of identical banks, each with total assets $A$ and debt $dA$. The banks hold $n$ different types of assets, the weight and risk-weight on asset $i$ being $w_i$ and $r_i$, respectively. Regulation requires that banks maintain a risk-weighted capital ratio of at least $\theta \in (0,1)$ and we assume that this requirement is binding for banks in period 0.[9] The binding capital requirement in period 0 implies:

$$\theta = \frac{A(1-d)}{\sum_{i=1}^{n} A w_i r_i} = \frac{1-d}{\bar{r}}$$
$$\iff 1 - \theta \bar{r} = d \tag{2.6}$$

where $\bar{r} = \sum_{i=1}^{n} w_i r_i$. In period 1, there is an unanticipated shock to asset fundamental values. Specifically, every asset $j$ experiences a percentage decline of $1 - \lambda_j$, where $\lambda_j \in (0,1)$. The shock causes banks' total assets to fall to $\sum_{i=1}^{n} A w_i \lambda_i = \bar{\lambda} A$, causing the capital ratio to fall under the regulatory minimum. This can be shown as follows:

$$\lambda_i < 1,\ \theta r_i < 1\ \forall i \implies \sum_{i=1}^{n} w_i \lambda_i (1 - \theta r_i) < \sum_{i=1}^{n} w_i (1 - \theta r_i) \tag{2.7}$$
$$\implies \bar{\lambda} - \theta \sum_{i=1}^{n} w_i \lambda_i r_i < 1 - \theta \bar{r} = d \tag{2.8}$$
$$\implies \frac{A(\bar{\lambda} - d)}{\sum_{i=1}^{n} A w_i \lambda_i r_i} < \theta \tag{2.9}$$

where (2.8) employs (2.6) and the left side of (2.9) represents the risk-weighted capital ratio of the bank after the unanticipated shock in period 1. Note the inclusion of $\theta r_i < 1\ \forall i$ in (2.7). We assume this is true and believe it is reasonable given the level of capital requirements and range of risk weights in the new Basel III regulations.[10]

In period 2, every asset $j$ experiences a net return of $\frac{\eta_j}{\lambda_j} - 1$, where $\{\eta_i\}_{i=1}^{n}$ are jointly distributed according to $f(\eta_1, \eta_2, \ldots \eta_n)$, marginally distributed according to $f_i(\eta_i)$, and the support of $\eta_i$ is $[0, \eta_{iH}]$ where $\eta_{iH} > 0$. We assume that in period 1, all assets are priced in a risk-neutral manner based on fundamental values with a risk-free rate of zero, which implies $E[\eta_i] = \lambda_i$. At the end of period 2, banks liquidate their assets at fundamental values, repay debt to the extent possible, and give the residual to shareholders.

[9] This assumption can be justified in the data by observing that banks generally maintain their capital ratios close to the requirement and tend to resist proposed increases in the required ratio.

[10] http://usbasel3.com/tool/

In order to recapitalize in period 1, banks undertake some combination of asset sales/purchases and equity issuance/repurchase. By assumption, they choose whatever combination maximizes the expected value of existing shareholders' equity. The dollar amount of asset $i$ sold is given by $s_i\bar{\lambda}A$, where $s_i$ can be positive (asset sale) or negative (asset purchase). There is a no shorting constraint, meaning a bank cannot sell more of an asset than it has. There is also a maximum amount of each asset that can be purchased (the market supply of the asset). The dollar amount of equity issued is given by $eA$, where $e$ can also be positive or negative. The amount of equity repurchased cannot exceed the equity value of the entire firm.

Any cash excess (deficit) from transactions in assets and equity is offset by debt repurchase (issuance). After all transactions, the outstanding debt of the bank must be weakly positive. Like the assets the banks hold, bank debt and equity are priced in a risk-neutral manner based on fundamental values. Moreover, there is a no-arbitrage condition that the price at which debt and equity are issued or repurchased must be the same as the price of outstanding debt and equity.

### 2.3.2 The bank's problem

In this section, we formally present the bank's problem. We then show how this problem collapses into the bank simply wanting to minimize the price of its debt per unit face value, just as in basic model presented in Section 2.2. Finally, we eliminate some redundant constraints to produce the most parsimonious version of the problem.

#### Statement of the problem

In period 1, the bank maximizes the expected equity value of existing shareholders in period 2. This is equivalent to solving the following problem, a general version of (2.2):

$$\max_{\{s_i\}_{i=1}^n,\, e} \; E\left( max\left[ \sum_{i=1}^{n} \frac{\eta_i}{\lambda_i}(w_i\lambda_i - s_i\bar{\lambda}) - \left( d - \frac{1}{\beta}\left( e + \bar{\lambda}\sum_{i=1}^{n} s_i \right)\right),\ 0 \right]\right) - e \tag{2.10}$$

subject to

$$\theta = \frac{\bar{\lambda}\left(1 - \sum_{i=1}^{n} s_i\right) - \left(d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n} s_i\right)\right)}{\sum_{i=1}^{n} r_i(w_i\lambda_i - s_i\bar{\lambda})} \tag{2.11}$$

$$\beta = \frac{E\left(min\left[\sum_{i=1}^{n}\frac{\eta_i}{\lambda_i}(w_i\lambda_i - s_i\bar{\lambda}),\ d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n} s_i\right)\right]\right)}{d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n} s_i\right)} \tag{2.12}$$

$$\beta d - \bar{\lambda} \leq e \leq \beta d - \bar{\lambda}\sum_{i=1}^{n} s_i \tag{2.13}$$

$$-\kappa_i \leq s_i\bar{\lambda} \leq w_i\lambda_i,\ \forall i \in \{1, 2, \ldots n\} \tag{2.14}$$

**Objective function**

The objective function (2.10) can be explained as follows. The term $w_i\lambda_i - s_i\bar{\lambda}$ represents the amount of asset $i$ that the bank holds after selling or purchasing some of the asset.[11] Multiplying this quantity by $\frac{\eta_i}{\lambda_i}$ gives the value of the bank's holdings of asset $i$ in period 2 and summing across all $i$ gives the total value of the bank's assets.

The term $e + \bar{\lambda}\sum_{i=1}^{n} s_i$ represents the net amount the bank raises in equity issuance and asset sales. By assumption, this excess or deficit goes toward debt repurchase or issuance at the price of $\beta$. Therefore, $d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n} s_i\right)$ is the face value of debt that remains after all transactions. The total value of the bank's equity is the greater of zero and the difference between total assets and remaining debt. The expected value of this quantity is taken (since $\eta_i$ is uncertain) and in order to isolate the value to existing shareholders, the amount of equity issued is subtracted.

[11]The constant $A$ (the value of the bank's initial assets) is technically a multiplier on all of the above equations but can be dropped or canceled. The problem is invariant to the original scale of the banks.

**Constraints**

Constraint (2.11) is the capital requirement that must be satisfied in period 1 after the unanticipated shock. On the right hand side, $\bar{\lambda}\left(1-\sum_{i=1}^{n} s_i\right)$ is the value of assets that remain after all sales or purchases. As discussed before, the second term in the numerator is the face value of debt that remains after all transactions, making the numerator the book value of equity in period 1. The denominator of (2.11) is the bank's cumulative risk-weighted assets after all purchases and sales: the weighted sum of the amount of each asset held, where the weight on asset $i$ is its regulatory risk weight $r_i$. Therefore, constraint (2.11) says that after all transactions in period 1, the ratio of book equity to risk-weighted assets must exceed the regulatory requirement $\theta$. The equality condition in (2.11) is by assumption, though it can be shown that if constraint was an inequality, it would be binding.[12]

Constraint (2.12) is the definition of $\beta$, the price of the bank's outstanding debt per unit of face value after all transactions. The numerator is the expectation of what debtholders will receive in period 2: the smaller of total asset value and the face value of remaining debt. This quantity is scaled by the face value of remaining debt to put $\beta$ in the correct units. By assumption, $\beta$ is also the price at which debt is purchased or issued, as seen in (2.10) and (2.11). Note that constraint (2.12) is undefined for $d-\frac{1}{\beta}\left(e+\bar{\lambda}\sum_{i=1}^{n} s_i\right)=0$. In the event this condition is true, $\beta$ is simply defined by this condition.

Constraint (2.13) defines the bounds on the choice of $e$. The right inequality states that the proceeds of equity issuance ($e$) cannot exceed the market value of debt that was originally outstanding ($\beta d$), net of the proceeds of asset sales ($\bar{\lambda}\sum_{i=1}^{n} s_i$). This constraint is equivalent to condition that the amount of debt that remains after the proceeds of equity issuance and asset sales, $d-\frac{1}{\beta}\left(e+\bar{\lambda}\sum_{i=1}^{n} s_i\right)$, must exceed zero.

The left inequality of (2.13) puts a limit on how much equity can be repurchased. This

[12]The proof involves showing that if the capital ratio exceeds $\theta$, the bank can raise shareholder value by simply issuing debt and repurchasing equity.

inequality can be restated as $-e \leq \bar{\lambda} - \beta d$, where $-e$ is the amount of equity repurchased. The right hand side is the market value of equity that is available to be repurchased after asset sales are used to repurchase debt at the price of $\beta$:

$$\bar{\lambda}\left(1 - \sum_{i=1}^{n} s_i\right) - \beta\left(d - \frac{1}{\beta}\bar{\lambda}\sum_{i=1}^{n} s_i\right)$$

which simplifies to $\bar{\lambda} - \beta d$.

Finally, constraint (2.14) puts limits on asset sales and purchases. The right inequality says that the amount of asset $i$ that is sold ($s_i\bar{\lambda}$) cannot exceed the holdings of asset $i$ after the unanticipated shock in period 1 ($w_i\lambda_i$). The left inequality says that the amount of asset $i$ that is purchased ($-s_i\bar{\lambda}$) cannot exceed market supply $\kappa_i$. One can imagine that $\kappa_i$ is large for liquid assets and small for illiquid assets.

Note that constraints (2.13) and (2.14) ensure that the numerator and denominator of $\beta$ as defined in (2.12) are weakly positive, meaning $\beta \geq 0$. In addition, the nature of the definition of $\beta$ in (2.12) implies $\beta \leq 1$. Therefore, we have $\beta \in [0, 1]$.

**Equity issuance feasibility**

In order for equity issuance to be feasible after the unanticipated shock in period 1, the amount of equity issued must be less than or equal to the total equity value to new and old shareholders after all asset sales and debt repurchases have taken place:

$$e \leq \bar{\lambda}\left(1 - \sum_{i=1}^{n} s_i\right) - \beta\left(d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n} s_i\right)\right)$$

$$\iff \bar{\lambda} \geq \beta d$$

In other words, the bank's assets must exceed the market value of its debt. Since (2.12) implies that $\beta \leq 1$, a sufficient condition for equity issuance being feasible in period 1 is $\bar{\lambda} \geq d$: after the unanticipated shock in period 1, the bank's assets must exceed the face value of its debt.

### 2.3.3 Simplifying the problem

In this section, we show how in the general model, just as in the basic model presented in section 2.2, the bank's objective is equivalent to minimizing $\beta$ (the price of the bank's outstanding debt per unit face value). We start with the fact that for any $X$ and $Y$, $max(X-Y,0) = X - min(X,Y)$. This means that the bank's objective in (2.10) can be rewritten as

$$E\left(\sum_{i=1}^{n}\frac{\eta_i}{\lambda_i}(w_i\lambda_i - s_i\bar{\lambda}) - min\left[\sum_{i=1}^{n}\frac{\eta_i}{\lambda_i}(w_i\lambda_i - s_i\bar{\lambda}), d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n}s_i\right)\right]\right) - e$$

Using (2.12) and $E[\eta_i] = \lambda_i$, this becomes

$$\bar{\lambda}\left(1 - \sum_{i=1}^{n}s_i\right) - \beta\left(d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n}s_i\right)\right) - e \tag{2.15}$$

$$= \bar{\lambda} - \beta d \tag{2.16}$$

Since $\bar{\lambda}$ and $d$ are parameters that do not depend on the bank's decision, the bank's problem can be restated as

$$\min_{\{s_i\}_{i=1}^{n},\, e} \quad \beta$$

subject to (2.11)-(2.14). Intuitively, the shareholders want to take the action that minimizes the market value of the debt (per unit of face value) because that maximizes the market value of equity, given that total asset values ($\bar{\lambda}$) are unaffected by the action taken. This is the risk-shifting motive: shareholders want to take on as much risk as possible, offloading the downside on to the creditors or equivalently, lowering the price of outstanding debt. The preceding argument proves that this is in fact the only consideration. From (2.15), asset sales reduce the market value of assets and debt by the same amount. In addition, equity issuance adds to the total equity value of the firm but is subtracted out for the existing shareholders. Therefore, asset sales/purchases and equity issuance/repurchases only affect the objective through their effects on $\beta$.

Since the definition of $\beta$ in (2.12) is self-referencing, it is useful to express it differently. Using (2.11) and noting that

$$\bar{\lambda}\left(1 - \sum_{i=1}^{n}s_i\right) = \sum_{i=1}^{n}(w_i\lambda_i - s_i\bar{\lambda}) \tag{2.17}$$

we find

$$d - \frac{1}{\beta}\left(e + \bar{\lambda}\sum_{i=1}^{n} s_i\right) = \sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda}) \tag{2.18}$$

Substituting (2.18) into (2.12), the objective function becomes

$$g(s_1, s_2, \dots s_n) = \frac{E\left(min\left[\sum_{i=1}^{n}\frac{\eta_i}{\lambda_i}(w_i\lambda_i - s_i\bar{\lambda}),\ \sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})\right]\right)}{\sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})} \tag{2.19}$$

The problem can therefore be restated as

$$\min_{\{s_i\}_{i=1}^{n}} \quad g(s_1, s_2, ..., s_n)$$

subject to (2.11), (2.13), and (2.14), with $g(s_1, s_2, ..., s_n)$ being substituted for $\beta$. To be clear, $g(\cdot)$ is the value of $\beta$ (the price per unit face value of the debt outstanding) that results from a particular choice of asset sales $\{s_1, s_2, ..., s_n\}$, assuming that the capital ratio requirement in (2.11) is met. Note that the choice of $e$ is no longer part of the bank's problem. This is because $e$ is pinned down by the choice of asset sales and (2.11), with $g(\cdot)$ in place of $\beta$.

### 2.3.4 Eliminating redundant constraints

In this section, we show that constraint (2.13) is redundant. Let $\{\{s_i\}_{i=1}^{n}, e\}$ be any feasible solution that satisfies (2.11) and (2.14), with $g(\cdot)$ in place of $\beta$. Using (2.17) and rearranging (2.11), we have

$$\begin{aligned}
e &= \left(d - \sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})\right)g(s_1, s_2, \dots s_n) - \bar{\lambda}\sum_{i=1}^{n} s_i && (2.20)\\
&= g(\cdot)d - \bar{\lambda}\sum_{i=1}^{n} s_i - g(\cdot)\sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda}) && (2.21)\\
&= g(\cdot)d - \bar{\lambda} + \left[\bar{\lambda}\left(1 - \sum_{i=1}^{n} s_i\right) - g(\cdot)\sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})\right]\\
&= g(\cdot)d - \bar{\lambda} + \left[\bar{\lambda}\left(1 - \sum_{i=1}^{n} s_i\right)(1 - g(\cdot)) + g(\cdot)\sum_{i=1}^{n}\theta r_i(w_i\lambda_i - s_i\bar{\lambda})\right] && (2.22)
\end{aligned}$$

Using $g(\cdot) \geq 0$, $\theta r_i \leq 1\ \forall i$, and (2.14), we have

$$g(\cdot)\sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda}) \geq 0$$

Combining this with (2.21) implies that $e \leq g(\cdot)d - \bar{\lambda}\sum_{i=1}^{n} s_i$, which is the right inequality of (2.13).

Using $g(\cdot) \in [0,1]$ and (2.14), the bracketed term in (2.22) is weakly positive, which implies that $e \geq g(\cdot)d - \bar{\lambda}$, which is the left inequality of (2.13). To summarize, any choice of $\{s_1, s_2, ..., s_n\}$ that satisfies (2.14) is feasible as long as $e$ is set according to (2.20).

### 2.3.5 Final statement of problem

The problem can be stated as

$$\min_{\{s_i\}_{i=1}^{n}} \quad g(s_1, s_2, \ldots s_n) = \frac{E\left(min\left[\sum_{i=1}^{n}\frac{\eta_i}{\lambda_i}(w_i\lambda_i - s_i\bar{\lambda}),\ \sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})\right]\right)}{\sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})} \tag{2.23}$$

subject to

$$\begin{gathered} -\kappa_i \leq s_i\bar{\lambda} \leq w_i\lambda_i \ \forall i \in \{1, 2, ..., n\} \\ s_j\bar{\lambda} < w_j\lambda_j \text{ for at least one } j \in \{1, 2, ..., n\} \end{gathered} \tag{2.24}$$

Note that the second constraint in (2.24) has been introduced to eliminate the possibility of selling the entire balance sheet, which would lead to zero capital and risk-weighted assets.

In summary, the bank's problem is to choose asset sales $\{s_1, s_2, ..., s_n\}$ that minimize the price per unit face value of the outstanding debt, or $g(\cdot)$. As stated earlier, this is the action that maximizes equity value, because it minimizes the transfer made to the creditors. Importantly, apart from (2.24), the problem is otherwise unconstrained. This is because for any choice of $\{s_1, s_2, ..., s_n\}$, a value of $e$ can be obtained from (2.20) that satisfies the constraints (2.11) and (2.13).

## 2.4 Results

In this section, we solve the model introduced in the previous section to establish the main results of the paper. To build intuition, we begin by analyzing the case in which banks hold just one asset. As in (Admati *et al.*, 2013), we find that banks are indifferent over all combinations of asset sales and equity issuance that restore the capital ratio. We then move to the n-asset case and show that the solution involves banks concentrating their portfolio in the asset that permits the most risk-taking. Finally, we use the n-asset case to prove the main results of the paper regarding how risk weights, liquidity requirements, and mandatory equity issuance can affect whether banks engage in fire sales when recapitalizing in a crisis.

### 2.4.1 One asset case

Using (2.23) with $n = 1$, we have

$$g(s_1) = \frac{E\left(min\left[\frac{\eta_1}{\lambda_1}\lambda_1(1-s_1),\ (1-\theta r_1)\lambda_1(1-s_1)\right]\right)}{(1-\theta r_1)\lambda_1(1-s_1)}$$
$$= \frac{E\left(min\left[\frac{\eta_1}{\lambda_1},\ 1-\theta r_1\right]\right)}{1-\theta r_1} \tag{2.25}$$

Note that $g(s_1)$ does not depend on $s_1$. Therefore, any $\{s_1, e\}$ that satisfies (2.20) and (2.24) is a valid solution. As long as the capital ratio is restored, the particular combination of asset sales and equity issuance does not affect the price of debt per unit face value and therefore does not matter.

This result is better understood by looking at the capital requirement (2.11) when $n = 1$:

$$\theta = \frac{\lambda_1(1-s_1) - \left(d - \frac{1}{\beta}(e + \lambda s_1)\right)}{r_1\lambda_1(1-s_1)}$$
$$\implies 1 - \theta r_1 = \frac{d - \frac{1}{\beta}(e + \lambda s_1)}{\lambda_1(1-s_1)}$$

The last line shows that regardless of the choice of $\{s_1, e\}$, the ratio of the face value of debt outstanding (numerator) to assets (denominator) after any asset sales and equity issuance in period 1 equals $1 - \theta r_1$. The choice of $\{s_1, e\}$ only changes the scale of the bank's balance sheet. This is the nature of the capital requirement with one asset.

Let $p$ be the probability of default in period 2. With probability $1 - p$, the value of assets exceeds the face value of debt and creditors receive \$1 for every dollar of face value. With probability $p$, creditors only recover some fraction $\mu$ of the face value of debt, where $\mu$ equals the ratio of asset value to the face value of debt. With this terminology, we can express the period 1 price of debt outstanding per unit face value as

$$g(s_1) = (1-p)(1) + pE(\mu \mid \text{default})$$

Earlier, we showed that the ratio of debt outstanding to assets at the end of period 1 is fixed at $1-\theta r_1$. This means that $p$ and $E(\mu \mid \text{default})$ will depend only on this ratio and return distribution of the asset between periods 1 and 2. They will not depend on the choice of $\{s_1, e\}$. These dependencies are clearly observed in (2.25). Different actions simply scale up or down how large the bank is but leave the fundamental riskiness of the debt (and therefore its price per unit of face value) unchanged.

### 2.4.2 Multiple asset case

In the one asset case, we showed that the price of the bank's debt per unit face value depends on the ratio of debt to assets at the end of period 1 and the return distribution of the asset between between periods 1 and 2, neither of which depend on the choice of $\{s_1, e\}$. Therefore, banks are indifferent between any choice of $\{s_1, e\}$ that satisfies the capital requirement. When there are two or more assets, this indifference result no longer holds.

The first reason is that if the assets have different risk weights, the ratio of debt to assets at the end of period 1 depends on the choice of $\{\{s_i\}_{i=1}^n, e\}$. If the bank chooses to sell assets with low risk weights (retain assets with high risk weights), the capital requirement forces the bank to hold less debt for the same level of assets. If the bank chooses to sell assets with high risk weights (retain assets with low risk weights), the bank is allowed to hold more debt for the same level of assets. The second reason is that if the assets have different return distributions, the distribution of the return of the bank's assets between periods 1 and 2 depends on the composition of the bank's portfolio at the end of period 1. This composition in turn depends on which assets the bank chooses to sell and retain in period 1.

### Characterizing the solution

We have shown that the recapitalization decision is non-trivial for banks when there is more than one asset. To find the optimal action, we formally solve the problem described by (2.23) and (2.24): choosing the combination of asset sales that minimizes the price of the bank's outstanding debt per unit face value subject to no-shorting constraints, where equity issuance is pinned down by the capital requirement according to (2.20). The bank's decision is characterized by the following proposition.

**Proposition 1.** *The solution to the recapitalization problem described by (2.23) and (2.24) involves concentrating the portfolio into one asset and selling all other assets. The asset that is retained is the solution to*

$$\arg\min_{i} \frac{E\left(min\left[\frac{\eta_i}{\lambda_i}, 1-\theta r_i\right]\right)}{1-\theta r_i}$$

*If multiple assets solve this problem, banks still choose to retain a single asset as long as there is imperfect correlation across asset returns. However, the bank is indifferent between retaining any of the assets that solve the problem.*

*Proof.* See Appendix B.1. □

The intuition for this result is fairly straightforward. Since the objective is to minimize the price of debt, the shareholders want to take on as much risk as possible. It is therefore undesirable to hold multiple assets at once, as this could only provide unwanted diversification. Even though all assets are priced fairly, the one that allows for the most risk-taking is best for shareholders because it minimizes the transfer to creditors. As a result, the optimal choice is to sell out of the other assets entirely. A similar argument holds if two assets provide equal amounts of risk when held on their own. While the solution is not unique (the bank is indifferent between holding either asset by itself), the bank does not want to hold these assets simultaneously due to the unwanted diversification effect.

In Appendix B.1, one can see in (B.1) that if only one asset ($k$) is retained, the objective function takes the form

$$\frac{E\left(min\left[\frac{\eta_k}{\lambda_k}, 1-\theta r_k\right]\right)}{1-\theta r_k} \tag{2.26}$$

Note that (2.26) does not depend $s_k$, which means (2.26) does not depend on how much of asset $k$ the bank holds after recapitalizing. While this may seem puzzling, it is because a bank that completely discards all but one asset is equivalent to the one-asset bank analyzed in the previous section. We showed that in this case, the bank is indifferent between all actions that satisfy the capital requirement.

The decision of which asset to sell involves comparing the value of (2.26) across all assets. The comparison is equivalent to the following. In period 1, suppose there are $n$ banks. Each bank holds only one asset, holds a different asset than all other banks, and is adequately capitalized. Note that these hypothetical banks are what would result from choosing to sell all assets other than one in the course of recapitalizing. The relevant question is: which bank's debt has the lowest price per unit face value? The optimal decision is to retain only the asset that this bank holds and sell all other assets. This decision can be interpreted as banks engaging in "regulatory arbitrage": concentrating the portfolio in assets that allow the most risk-taking while satisfying regulatory requirements.

It is worth noting here that Proposition 1 is quite extreme. For several reasons, it is unreasonable for banks to close out their entire holdings of all assets except for one in the process of recapitalizing (we will revisit this assumption in a later section). However, Proposition 1 is just meant to underscore the risk-shifting motives that banks face when choosing how to recapitalize in a crisis. Holding constant all of the other factors that drive banks' recapitalization decision, banks have an incentive to concentrate their portfolios in assets that allow them to take the most risk. This basic intuition leads to the main results of the paper, which we turn to next.

### The effect of risk weights

The effect of risk weights on the recapitalization decision is summarized by the following two propositions.

**Proposition 2.** *For each asset $k$, there exists $\bar{r}_k < \frac{1}{\theta}$ such that for all $r_k > \bar{r}_k$, asset $k$ is sold in the process of recapitalizing, holding fixed all other assets' risk weights.*

*Proof.* See Appendix B.2. □

The intuition for this result is as follows. Proposition 1 established that when deciding how to recapitalize in a crisis, banks choose to retain the single asset that minimizes the bank's price of debt per unit face value. Suppose the bank is evaluating the option of retaining asset $k$. As $r_k$ rises, the probability of solvency in period 2 goes up because the bank is forced to hold less debt when risk weights on asset $k$ are higher. However, this effect is ultimately second-order for creditors because the benefit of newly solvent states of the world in which creditors are made whole is offset by the fact that prior to the increase in $r_k$, creditors were just barely insolvent in these states. Creditors are benefiting only slightly in these newly solvent states in terms of the market value of debt per unit of face value.

A higher $r_k$ also does not affect the attractiveness of retaining any other asset by itself. The only first-order effect of a higher $r_k$ is that when banks holding asset $k$ are forced to hold less debt, creditors recover a greater percentage of their face value in insolvent states of period 2. This is bad for shareholders, as it raises the price of debt per unit face value and constitutes a transfer from to creditors. If $r_k$ is high enough, shareholders will eventually prefer to sell asset $k$ and retain a different asset since retaining the former forces the bank to be underleveraged.

In risk-weighted capital requirements, "risky" assets tend to be assigned higher risk weights. This fact, along with Proposition 2, leads to the first main result of the paper: the design of capital requirements, specifically the assignment of risk weights, can affect whether banks engage in fire sales of risky assets when recapitalizing in a crisis. Holdings other things constant, the higher the risk weights are on "risky" assets, the more likely it is that banks choose to sell these risky assets in a recapitalization. If these risky assets are illiquid, fire sales could occur. In other words, while high risk weights on risky assets may have ex-ante benefits, they are not necessarily a panacea because of the adverse ex-post incentives they create.

The next proposition establishes what happens when risk weights are uniform across assets.

**Proposition 3.** *Suppose the risk weights of all assets are identically equal to one. If asset j's net return*

$(\frac{\eta_j}{\lambda_j} - 1)$ *is a mean-preserving spread of the return of asset k (i.e. asset j is riskier), then the bank will not sell asset j in the process of recapitalizing.*

*Proof.* See Appendix B.3. □

Proposition 3 can be interpreted as follows: a system of uniform risk weights across assets can push banks toward the more desirable outcome of selling "safer" assets when recapitalizing in a crisis. In this paper, we denote an asset being "safer" than another if the latter's net return in period 2 is a mean-preserving spread of the former's. In other words, while a safer asset has the same expected net return as a riskier asset, it also has less volatility.

The intuition behind Proposition 3 is straightforward. When risk weights are identical, the choice of which asset to retain has no influence on the amount of leverage banks can take. Therefore, the decision for an equity-maximizing bank is simple: retain the asset whose return profile offers the most risk. Since all assets provide the same expected return of zero, an asset whose return is a mean-preserving spread of another's is attractive to shareholders because it must perform relatively well in good states of the world. In contrast, retaining a safe asset provides less upside, raises the price of the bank's debt per unit face value, and constitutes a transfers from shareholders to creditors.

**The effect of liquidity requirements**

Proposition 1 says that the solution to the recapitalization problem involves retaining just one asset and selling all of the others completely. One reason such an action would not be feasible in practice is that banks are subject to internal or regulatory liquidity requirements, whereby certain assets must be retained for liquidity purposes. In this section, we introduce a simplified version of liquidity requirements in order to understand how it changes the bank's decision from the base case in Proposition 1.

We model liquidity requirements by modifying constraint (2.24) as follows.

$$- \kappa_i \leq s_i \bar{\lambda} \leq \gamma_i w_i \lambda_i \ \forall i \in \{1, 2, ... \, n\}, \ \gamma_i \in [0, 1) \tag{2.27}$$

The modified constraint (2.27) says that asset $i$ can be sold only up to a fraction $\gamma_i$ of current holdings. One can imagine that for illiquid assets $\gamma_i \approx 1$ and for liquid assets $\gamma_i << 1$, the idea being that banks must retain a certain amount of liquid assets while there is no such requirement on illiquid assets.[13] While this setup may not be how liquidity requirements work in practice, it offers a simple way of building intuition about the problem.[14]

The next proposition examines how banks recapitalize in the presence of both capital and liquidity requirements.

**Proposition 4.** *With constraint (2.27) in place of (2.24) to reflect liquidity requirements, the solution to the bank's recapitalization problem is an extreme point of the feasible set: sell all assets but one and expand holdings of the remaining asset, both to the maximum extent allowable under (2.27).*

*Proof.* See Appendix B.4. □

The intuition underlying Proposition 4 is similar to that of Proposition 1: banks wants to take as much risk as possible. In the presence of liquidity constraints, this means expanding the holdings of one asset as much as possible while simultaneously holding as little as possible of every other asset. This produces the least diversified (riskiest) portfolio. Indeed, we can show that if there is no restriction on how much asset holdings can be built up ($\kappa = \infty$ in (2.27)), banks can achieve the same minimized objective as they do without liquidity requirements.[15]

To prove this, we take the limit of the objective (2.23) as $s_k$ converges to $-\infty$ (expanding

[13] For transaction costs reasons, it may be the case that the opposite is true: banks can sell a large amount of liquid assets and only a small amount of illiquid assets. In this paper, we assume that since banks do not have to recapitalize all at once, transaction costs have less influence on banks' decision than liquidity requirements.

[14] In practice, Basel III requires banks to hold liquid assets to match short-term liabilities. This corresponds with effective limits on selling certain assets after facing a negative shock.

[15] This thought experiment corresponds to the case in which the bank grows by issuing long-term debt (and equity to meet the capital requirement), so there is no increase in the level of the liquidity requirement.

holdings of asset $k$ to infinity).

$$\lim_{s_k \to -\infty} \frac{E\left(min\left[\sum_{i=1}^{n} \frac{\eta_i}{\lambda_i}(w_i\lambda_i - s_i\bar{\lambda}),\ \sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})\right]\right)}{\sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})}$$

$$= \lim_{s_k \to -\infty} \frac{E\left(min\left[\sum_{i\neq k} \frac{\eta_i}{\lambda_i}\frac{w_i\lambda_i - s_i\bar{\lambda}}{w_k\lambda_k - s_k\bar{\lambda}} + \frac{\eta_k}{\lambda_k},\ \sum_{i\neq k}(1-\theta r_i)\frac{w_i\lambda_i - s_i\bar{\lambda}}{w_k\lambda_k - s_k\bar{\lambda}} + 1 - \theta r_k\right]\right)}{\sum_{i\neq k}(1-\theta r_i)\frac{w_i\lambda_i - s_i\bar{\lambda}}{w_k\lambda_k - s_k\bar{\lambda}} + 1 - \theta r_k}$$

$$= \frac{E\left(min\left[\frac{\eta_k}{\lambda_k},\ 1-\theta r_k\right]\right)}{1-\theta r_k}$$

The last line is equivalent to (B.1), the objective when retaining only asset $k$ in the absence of liquidity requirements. The intuition here is that building up holdings of one asset to infinity offers the same risk profile as holding a finite amount of that asset and selling all other assets entirely. While the liquidity requirements affect the structure of the bank's portfolio, they do not affect the minimized value of the objective function.

Therefore, one of the main effects of liquidity requirements is that banks are no longer invariant to their scale, as they were in Proposition 1. Instead, banks want to expand holdings of one asset as much as they can in order to "dilute" holdings of other assets that cannot be sold because of liquidity requirements.

The limit analysis above shows that when asset holdings can be built up indefinitely ($\kappa = \infty$ in (2.27)), banks will choose to build up the same asset that is retained in the absence of liquidity requirements (see Proposition 1). However, when asset holdings cannot be built up indefinitely ($\kappa < \infty$ in (2.27)), the asset that banks choose to build up is not necessarily the asset that is retained in the absence of liquidity requirements.

Proposition 1 shows that if there are no liquidity requirements, asset $j$ is retained based on two factors only: the inherent riskiness of its gross return $\frac{\eta_j}{\lambda_j}$ and the leverage it permits through its risk weight $r_j$. With liquidity requirements, there are two additional factors at play: the extent to which an asset's holdings can be built up ($\kappa_j$) and how small the liquidity requirements of the remaining assets are ($\gamma_k \; \forall k \neq j$).

As an example, suppose there are two assets: $j$ and $k$. Asset $j$ is risky and illiquid, meaning it has a low liquidity requirement and minimal market supply. In contrast, asset $k$ is safer and more liquid, with a higher liquidity requirement and plenty of market supply. Suppose also that the assets' risk weights are the same. By Proposition 3, asset $k$ should be is sold and asset $j$ retained in the process of recapitalizing when there are no liquidity requirements. However, it is possible that this decision flips with liquidity requirements in place. The appeal of "concentrating" the portfolio in $j$ is limited by the fact that a certain amount of diversification is unavoidable: a minimum amount of asset $k$ must be held and holdings of asset $j$ cannot be built up tremendously due to limited market supply. Meanwhile, it is possible to concentrate the portfolio in asset $k$, since its holdings can be built up substantially and asset $j$ does not need to be held for liquidity purposes.

In summary, liquidity requirements have two main effects on banks' recapitalization decision. First, banks have an incentive to become very large in the process of accumulating one asset and diluting the holdings of others. Second, the desire to retain risky assets for risk-shifting purposes may be undone if there is not an abundant supply of these assets or the liquidity requirements on other asset are high. If either of these is true, banks may choose to sell risky assets en masse (a fire sale) in the process of recapitalizing in a crisis.

**Mandatory equity issuance**

In this section, we explore how banks' recapitalization decisions are affected by mandatory equity issuance when both capital and liquidity requirements are in place. The principal issue is whether mandatory equity issuance can prevent fire sales of illiquid assets that banks would otherwise engage in as part of their recapitalization decisions.

Recall that when equity issuance is not mandatory, banks can choose any combination of asset sales/purchases $\{s_i\}_{i=1}^{n}$ that complies with (2.27). Based on this choice of $\{s_i\}_{i=1}^{n}$, equity issuance is pinned down by (2.20). When there is mandatory equity issuance of $\bar{e}$, the bank's choice of $\{s_i\}_{i=1}^n$ is subject to the following additional constraint.[16]

$$\left(d - \sum_{i=1}^{n}(1-\theta r_i)(w_i\lambda_i - s_i\bar{\lambda})\right) g(s_1, s_2, \dots s_n) - \bar{\lambda}\sum_{i=1}^{n} s_i \geq \bar{e} \tag{2.28}$$

The left hand side of (2.28) is the same as the right hand side of (2.20). The first term is the market value of the debt that is repurchased and the second term is the net proceeds of asset sales. Any difference between these two values must be made up with equity issuance, which the constraint requires to be greater than $\bar{e}$.

It is difficult to articulate a general solution to the bank's problem with mandatory equity issuance due to the dependence on multiple parameter values (the issuance requirement $\bar{e}$, the liquidity requirement $\gamma_i$ for asset $i$, and the market supply $\kappa_i$ of asset $i$). Instead, we can describe the nature of the deviations that mandatory equity issuance causes when it is binding, compared to what banks choose to do when they are not subject to (2.28) (see Proposition 4). We can also describe whether there are situations in which mandatory equity issuance is ineffective (i.e. non-binding) and if so, whether banks engage in fire sales of illiquid assets in such situations.

First, we analyze the binding case. Suppose that a bank has implemented its optimal recapitalization plan according to Proposition 4, meaning the bank has built up holdings of one asset and sold the remaining, both to the maximum extent possible under (2.27). Since mandatory equity issuance is binding, the bank's choice of $e$ in (2.20) is less than $\bar{e}$ and it must issue more equity. The main impact of additional issuance is that the bank is forced to use the proceeds to purchase assets. But because the bank has already expanded holdings of its most desired asset to the maximum extent possible, the proceeds of equity issuance can only be used to "buy back" assets the bank originally wanted to sell.

While this is bad for the bank's shareholders, who are forced to have a more diversified asset portfolio, mandatory equity issuance also causes banks to engage in less selling of the assets that

[16] Note that $\bar{e}$ is expressed as a fraction of the bank's total assets.

would otherwise be sold to the maximum extent possible. If these “undesired” assets happen to be illiquid, mandatory equity issuance can reduce the severity of fire sales that might otherwise occur in the process of bank recapitalization.

It is also possible that the constraint imposed by mandatory equity issuance is not binding, i.e. banks choose to issue sufficient equity on their own. Since equity issuance equals the difference between debt repurchases and asset sales, mandatory equity issuance would be non-binding if either debt repurchases are large, the bank is a large net buyer of assets, or both. None of these circumstances rule out fire sales of illiquid assets though, since even if a bank is net buyer of assets it could be selling certain assets in large quantities.

The conclusion from this section is that mandatory equity issuance can potentially mitigate the severity of fire sales when it is binding by forcing banks to do less net selling of potentially illiquid assets. However, if the constraint is not binding, fire sales may still occur. This suggests that mandatory equity issuance amounts should be set aggressively by policymakers.

## 2.5 Conclusion

In this paper, we model how banks that are subject to capital requirements choose to recapitalize after being hit by a shock to their asset values. We then use the model to assess how policy levers such as risk weights, liquidity requirements, and mandatory equity issuance can affect whether banks engage in fire sales of illiquid assets in the process of recapitalizing.

We first show that if banks act in the interest of shareholders, they will be influenced by risk-shifting motives and choose the combination of asset sales and equity issuance that minimizes the value of their debt per unit face value. In line with this objective, banks choose to concentrate their portfolios in one asset and discard the others, as this allows for minimal diversification and maximal risk-taking. The asset that is retained is the one that provides the best tradeoff between risk and allowable leverage, a function of the asset’s regulatory risk weight.

Based on this result, we show that the assignment of risk weights can affect whether banks engage in fire sales when recapitalizing. If the risk weight on a particular asset is sufficiently high, banks will choose to sell it because retaining it forces them to be underleveraged, limiting the amount of risk that can be shifted on to creditors. Since high risk-weight assets are likely to be illiquid, fire sales can result. However, fire sales may be prevented if the risk weights on all assets are the same because banks will retain the asset that has the greatest underlying risk and sell assets that are safer and probably more liquid.

When liquidity requirements are introduced, banks again choose to concentrate their portfolios into one asset but instead of selling the others completely, they are sold to the maximum extent that the liquidity requirements allow. This has two main effects on the bank recapitalization process. First, banks have an incentive to build up holdings of the desired asset as much as possible in order to dilute the holdings of the less desired assets that must be held because of liquidity requirements. Second, the appeal of retaining a risky, illiquid asset is diminished because banks are forced to hold a certain amount of safe, liquid assets, creating unwanted diversification. To avoid this, banks might instead choose to build up holdings of a different asset and sell the illiquid asset in a fire sale.

Finally, mandatory equity issuance can potentially mitigate the severity of fire sales when it is binding because it forces banks to do less net selling of potentially illiquid assets. However, if the circumstances are such that banks are issuing sufficient amounts of equity on their own, fire sales may still occur. This issue can potentially be addressed by setting mandatory equity issuance amounts aggressively.

Overall, our model suggests that regulations can impact how banks choose to recapitalize in a crisis. While assigning high risk weights to risky, illiquid assets may have favorable ex-ante incentives, doing so might generate unintended ex-post incentives for banks to engage in fire sales. In contrast, a system of uniform risk weights probably reduces the risk of fire sales. These findings are interesting in light of the fact that the new Basel III accords both raise the risk weights on a variety of risky, illiquid assets and introduce a simple leverage requirement (which is equivalent to uniform risk weights). According to the results of this paper, these two policies may actually have different effects on the tendency of banks to engage in fire sales when recapitalizing.

Our model also suggests that the interaction between different types of regulations can affect bank decision making. For example, it is possible that banks choose not to engage in fire sales when just capital requirements are in place but do engage in fire sales when liquidity requirements are introduced.

Finally, a policy prescription of our model is cyclical risk weights (CRWs) for illiquid assets. In addition to implicitly lowering the capital requirement after a shock occurs, CRWs would also lower risk weights on certain risky, illiquid assets, making it more worthwhile for banks to retain these assets in the process of recapitalizing. CRWs therefore offer additional ammunition against the risk of fire sales compared to cyclical capital requirements.

# Chapter 3

# Expectations vs. fundamentals-driven bank runs: When should bailouts be permitted?[1]

## 3.1 Introduction

The recent financial crisis saw governments and central banks undertake a range of unusual and, in some cases, unprecedented actions that could be characterized as "bailing out" financial institutions and investors. Many of these actions remain controversial and have led to calls for restricting policy makers' ability to intervene in future crises. Some restrictions of this type have already been put into place. For example, the Dodd-Frank Act in the United States requires any future Federal Reserve emergency lending programs to be approved by the Secretary of the Treasury, imposes stricter collateral and disclosure requirements on these programs, and prohibits programs that are designed to aid a particular financial institution. In addition, the Act prohibits the Treasury from issuing the type of guarantees offered to money market mutual funds beginning in September 2008. These legal changes raise an important question: When is it desirable to restrict policy makers' ability to intervene in a future crisis? While there has been much debate about the effects of such restrictions in policy circles, no clear principles have emerged to guide

[1]Co-authored with Todd Keister

these decisions. One common view holds that the desirability of restricting intervention depends critically on the underlying cause of a financial crisis. Gorton (2010) argues that the recent crisis was – at its heart – a run on certain elements of the financial system, similar in structure to the events that plagued the U.S. banking system in the 19th century. In such an event, many investors withdraw their funds from banks and other financial institutions in a short period of time, placing severe strain on the financial system. Lacker (2008) proposes a simple rule to guide decisions about whether intervention should be allowed that focuses on the underlying cause of these runs:

> Researchers have found it useful to distinguish between what I'll call 'fundamental' and 'non-fundamental' runs. . . . This distinction is important because the two types of runs have very different policy implications. Preventing a non-fundamental run avoids the cost of unnecessary early asset liquidation, and in some models can rationalize government or central bank intervention. In contrast, in the case of runs driven by fundamentals, the liquidation inefficiencies are largely unavoidable and government support interferes with market discipline and distorts market prices.

In other words, Lacker (2008) argues that intervention may be useful when runs on the financial system are self-fulfilling in nature, caused by shifts in investors' expectations. In particular, if the economy has multiple equilibria, allowing intervention may help eliminate undesirable equilibria and thereby prevent a run from occurring. If, however, the economy has a unique equilibrium and runs are instead driven by deteriorating economic fundamentals, restricting policy makers from intervening is claimed to lead to better outcomes.

Support for this view can be found in the growing literature on bank runs and financial crises. In the classic paper of Diamond and Dybvig (1983), for example, a bank run is non-fundamental in nature; depositors who are not in immediate need of funds will run on their bank only if they expect other depositors to do so. In their setting, intervention in the form of deposit insurance is desirable if it can remove the strategic complementarity in depositors' actions and ensure that no run occurs. This pattern – where bank runs are driven by agents' expectations and where allowing intervention may be desirable – can be found in many subsequent papers; examples include Chang and Velasco (2000), Cooper and Kempf (2013) and Keister (2016), to name only a few. Other papers in the literature, in contrast, study environments where a crisis results from a fundamental shock and have the property that restricting intervention, if feasible, would generate a superior outcome by eliminating the incentive distortions that arise when investors anticipate being rescued in the event of a crisis. See, for example, Farhi and Tirole (2012) and Chari and Kehoe (2013) for environments with these features.

While the results in these papers are consistent with the view that allowing intervention may be desirable if runs are caused by shifting expectations but is otherwise undesirable, none of the papers directly *test* this view. The models studied differ across papers along a number of dimensions, making it difficult to isolate the precise source(s) of the differing policy prescriptions. In this paper, we investigate the desirability of restricting intervention using a model in which an equilibrium bank run may be driven by either expectations or fundamentals, depending on parameter values. By including both possibilities in a unified framework, we are able to study the extent to which the desirability of restricting intervention depends on the underlying cause of a crisis and the extent to which it depends on other factors.

Our model is in the tradition of Diamond and Dybvig (1983) and builds most closely on that in Keister (2016), where a bank run can occur when depositors' actions are coordinated on an extrinsic " sunspot" variable. We extend the model by introducing intrinsic uncertainty: the level of fundamental withdrawal demand is random. We say that a bank run in this expanded setting is driven by expectations when depositors' behavior depends on the sunspot variable and, hence, is driven in part by their beliefs about the actions of other depositors. In contrast, we say that a bank run is driven by fundamentals if a run necessarily occurs whenever fundamental withdrawal demand is high, independent of the sunspot variable. We ask whether the desirability of restricting intervention in this setting depends critically on which form a run takes, that is, on whether runs are driven by expectations or by fundamentals.

We show that the optimal policy regime in our model depends on a basic tradeoff between incentives and insurance. When banks and depositors anticipate that policy makers will intervene in the event of a crisis, they have less incentive to provision for bad outcomes. In response, banks increase their short-term liabilities, which distorts the allocation of resources and tends to make the financial system more susceptible to a run. At the same time, however, intervention can provide an important source of risk sharing in the economy. By mitigating the potential losses depositors suffer during a crisis, a "bailout" can both smooth depositors' consumption across states and encourage them to leave their funds in the financial system rather than trying to withdraw. Thus, while the incentive distortion associated with intervention tends to make the financial system more fragile and lower welfare, the insurance effect tends to raise welfare and promote stability. Importantly, this same tradeoff arises regardless of whether runs in the model are driven by expectations or by fundamentals.

The desirability of restricting intervention depends on which of these two effects dominates. If policy makers are able to eliminate the incentive distortion through effective regulation and supervision of banks, then allowing intervention is always optimal. If regulation is imperfect and the risk-sharing benefit from intervention is absent, in contrast, it is optimal to prohibit intervention. In between these extreme cases, we show that allowing intervention is optimal whenever regulation is sufficiently effective for the insurance effect to dominate. The precise cutoff point will depend on the specific features of the economy, including whether runs are driven by expectations or by fundamentals. However, the same tradeoff between incentives and insurance arises in both cases and the same basic principle should guide the policy choice. In this sense, our model provides meaningful policy advice that applies regardless of the underlying cause of these crises.

In the next section, we present the model and discuss the distinction between fundamental and non-fundamental runs in our framework. In Section 3, we study equilibrium outcomes when policy makers are restricted from intervening during a crisis. In section 4, we study equilibrium when intervention is allowed, highlighting both the resulting incentive distortion and the insurance benefit that arise. We compare these outcomes in Section 5, deriving conditions under which each regime is optimal and illustrating these conditions with a series of examples. Finally, in Section 6, we offer some concluding remarks that relate our results to the long-standing debate about the role of self-fulfilling expectations in financial crises.

## 3.2 The Model

Our model builds on that in Keister (2016), which is a version of the Diamond and Dybvig (1983) model augmented to include fiscal policy and a public good. We introduce aggregate uncertainty about the level of fundamental withdrawal demand to the model so that we can study runs caused by fundamental shocks in addition to runs triggered by shifts in expectations.

### 3.2.1 The environment

There are three time periods, $t = 0, 1, 2$. Each of a continuum of depositors, indexed by $i \in [0,1]$, is endowed with one unit of the good at $t = 0$ and has preferences given by

$$U(c_1, c_2, g; \omega_i) = u\left(c_1 + \mathbb{I}_{(\omega_i = 2)} c_2\right) + v(g),$$

where $c_t$ is consumption of the private good in period $t$, $\mathbb{I}$ is the indicator function, and $g$ is the level of public good. The preference type of depositor $i$, denoted $\omega_i$, is a binomial random variable with support $\Omega = \{1, 2\}$. If $\omega_i = 1$, depositor $i$ is *impatient* and only cares about consumption at $t = 1$, while if $\omega_i = 2$ she is *patient* and can consume at either $t = 1$ or $t = 2$. A depositor's type $\omega_i$ is revealed to her in period 1 and is private information. We assume the functions $u$ and $v$ to be of the constant relative risk-aversion form, with

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma} \quad \text{and} \quad v(g) = \delta \frac{g^{1-\gamma}}{1-\gamma}. \tag{3.1}$$

The parameter $\delta \geq 0$ measures the relative importance of the public good and will be a key factor in determining the potential insurance benefit from intervention. As in Diamond and Dybvig (1983), the coefficient of relative risk-aversion $\gamma$ is assumed to be greater than one.

At the beginning of period 1, the aggregate state of the economy is realized. This state has two components. The fundamental state ($L$ or $H$) determines the fraction $\pi$ of depositors who are impatient, with $\pi_L < \pi_H$. Conditional on the realized value of $\pi$, each depositor faces the same probability of being impatient. The " sunspot" state ($\alpha$ or $\beta$) is independent of the fundamental state and has no effect on preferences or technologies, but may serve to coordinate depositors' expectations in equilibrium. We denote the full state of the economy by

$$s \in S = \{L_\alpha, L_\beta, H_\alpha, H_\beta\}$$

and the probability of state $s$ by $q_s$.

There is a single, constant-returns-to-scale technology for transforming endowments into private consumption in the later periods. A unit of the good invested in period 0 yields $R > 1$ units in period 2, but only one unit in period 1. This investment technology is operated by a set of banks in which depositors pool resources to insure individual liquidity risk. Each bank is large enough that the fraction of its depositors who are impatient will equal the economy-wide average $\pi_s$ with probability 1, but small enough that its deposits are a negligible fraction of the aggregate endowment. Banks operate to maximize their depositors' expected utility at all times.

Depositors are isolated from each other in periods 1 and 2 and no trade can occur among them. Upon learning her preference type, each depositor chooses to withdraw in either period 1 or period 2. Depositors who choose to withdraw in period 1 arrive at their bank one at a time in a randomly-determined order and each exits the banking location before the next depositor arrives. As in Wallace (1988) and Wallace (1990), this sequential-service constraint implies that the payment made to a depositor can only depend on the information received by the bank up to the point at which she withdraws; we discuss the implications of this constraint in detail below.

There is also a linear technology for transforming units of the private good into units of the public good in period 1. Without any loss of generality, we assume the transformation rate is one-for-one. This technology is available to all agents, but the fact that both depositors and banks are small relative to the overall economy implies that there is no private incentive to provide the public good. Instead, there is a benevolent policy maker who has the the ability to tax banks in period 1 and can use the revenue from this tax to produce the public good. The objective of the policy maker is to maximize the equal-weighted sum of individual expected utilities,

$$W = \int_0^1 E\left[U\left(c_1(i), c_2(i), g; \omega_i\right)\right] di. \tag{3.2}$$

Note that while banks and the policy maker both aim to maximize depositor welfare, a key difference is that each bank only cares about its own depositors while the policy maker cares about all depositors in the economy.

We follow Ennis and Keister (2009) and Ennis and Keister (2010) in assuming that banks cannot commit to future actions. This inability to commit implies that they are unable to use the type of suspension of convertibility plans discussed in Diamond and Dybvig (1983) or the type of run-proof contracts studied in Cooper and Ross (1998) to eliminate undesirable equilibria. Instead, the payment given to each depositor who withdraws in period 1 will always be chosen as a best response to the current situation. The policy maker is also unable to commit to future plans and will choose the tax policy to maximize the objective 3.2 at each point in time in reaction to the situation at hand.

Depositors observe the realization of the state of nature at the beginning of period 1 and can, therefore, condition their withdrawal behavior on this information. Banks do not observe the state at this point and must make inferences about it from the flow of withdrawals.[2] In the equilibria we study below, a bank will be able to infer that the fundamental state is $H$ whenever the measure of $t = 1$ withdrawals goes above $\pi_L$. To simplify the analysis, we allow banks to observe the sunspot state at this same point. In other words, after a fraction $\pi_L$ of depositors have withdrawn, banks will learn the full state and, therefore, will know whether any surge in withdrawals has an expectations-driven component.[3] We place no restrictions on the payments a bank can make to its depositors other than those imposed by the information structure and sequential service

[2]This inference problem has been studied in related settings by Green and Lin (2003), Peck *et al.* (2003), Andolfatto *et al.* (2007) and Ennis and Keister (2010), and Sultanum (2014), among others.

[3]If banks and the policy maker did not observe the sunspot state, their reaction to a surge of withdrawals at $t = 1$ in the type of equilibria we study here would occur in two stages, the first when the fundamental state is inferred (after $\pi_L$ withdrawals) and the second when the sunspot state is inferred (after $\pi_H$ withdrawals). This two-stage response would imply that different types of expectations-driven runs are possible. Patient depositors may, for example, run until the first reaction and then stop, or they may run until the second reaction and then stop. While the possibility of expectations-driven bank runs occurring in distinct waves is interesting (see Ennis and Keister (2010) for a detailed analysis), our focus here is on comparing the policy implications of expectations-driven vs. fundamentals-driven runs. Assuming that the sunspot state is revealed after $\pi_L$ withdrawals simplifies the analysis by allowing us to focus on a single type of expectations-driven run. The results we present below would be qualitatively unchanged if we instead allowed for a two-stage response and choose to focus only on equilibria in which a run stops after the first policy response.

constraint described above. In particular, a bank is always free to adjust the payment it gives to its remaining depositors and will choose to do so when this new information arrives. We assume the policy maker observes the same information as banks about withdrawal behavior and the sunspot state.

### 3.2.2 Intervention and Regulation

We study two policy regimes. In the *no intervention* regime, the policy maker collects taxes and provides the public good at the beginning of period 1, before any withdrawals have occurred. Once withdrawals begin, any further fiscal policy is prohibited. In the regime *with intervention*, in contrast, the policy maker is able to learn the state $s$ before collecting taxes. The policy maker will respond to this information by adjusting tax rates and the level of the public good. In particular, the policy maker will generally respond to a crisis by lowering taxes, thereby "bailing out" banks and their depositors. Figure 3.1 depicts the timeline of events under each policy regime.

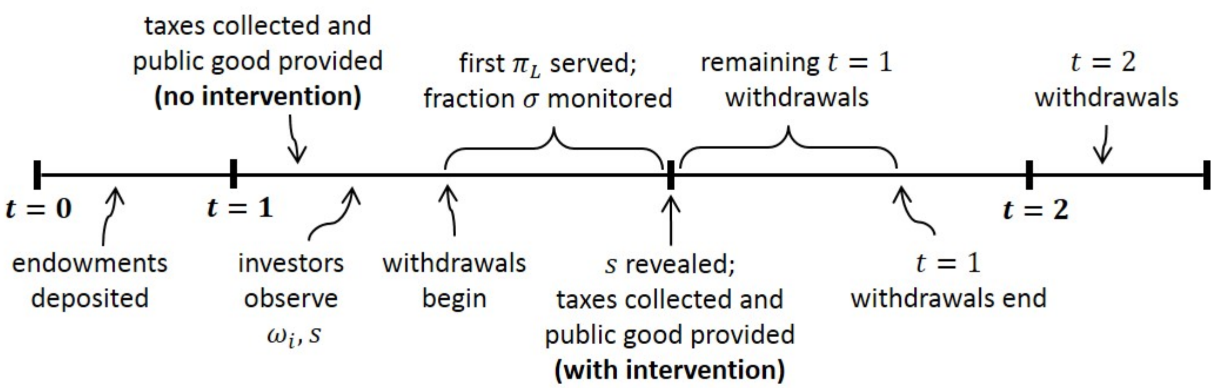



**Figure 3.1:** *Timeline of events*

We also give the policy maker a regulatory tool for mitigating potential incentive distortions. We show below that once the state has been fully revealed and any intervention has taken place, no such distortions arise and there is no role for regulation. As the first $\pi_L$ withdrawals take place, however, the policy maker may wish to influence banks' choices. We assume the policy maker is able to encounter a fraction $\sigma \in [0, 1]$ of these depositors immediately after they have withdrawn from the bank and before they have consumed. When the policy maker encounters a depositor, he can observe the quantity of goods she holds and can confiscate some of these goods, if desired. Confiscated goods are rebated back to all banks in a lump-sum fashion. The identities of the depositors who will encounter the policy maker are determined randomly but, as each depositor withdraws, the bank observes whether or not she will be monitored. The bank can forecast the maximum amount of consumption allowed by the policy maker and will, in equilibrium, choose to give monitored depositors exactly that amount, which may differ from the level of consumption given to non-monitored depositors. In this way, the policy maker's ability to monitor some withdrawals effectively places a cap on the amount these depositors will receive from their bank.

We interpret funds that will be withdrawn from a bank before the state is revealed as representing the bank's short-term liabilities. The activity of monitoring depositors is intended to represent, within the context of our model, a range of regulatory and supervisory activities that aim to limit such liabilities in practice. The Basel III accords, for example, introduce a Liquidity Coverage Ratio requirement that limits the short-term liabilities of a bank to be no larger than the quantity of safe, liquid assets it holds.[4] The parameter $\sigma$ in our model represents the policy maker's ability to use these types of regulatory and supervisory powers effectively. When $\sigma = 1$, we say that prudential regulation is perfectly effective: the policy maker can completely control the amount of funds withdrawn from the banking system before the state is revealed. Having $\sigma < 1$ represents an environment where writing effective regulation is difficult or where banks can partially evade regulations by, for example, designing new legal or accounting structures. In the analysis below, we study how the effectiveness of regulation impacts the desirability of allowing the policy maker to intervene.

### 3.2.3 Runs and fragility

Each depositor chooses a strategy that lists the period in which she will withdraw (1 or 2) for each possible realization of her preference type $\omega_i$ and the state $s$,

[4]See BCBS (2013) for a detailed discussion of this requirement.

$$y_i : \Omega \times S \to \{1, 2\}. \tag{3.3}$$

Let $y$ denote a profile of withdrawal strategies for all depositors. An equilibrium of the model is a profile of withdrawal strategies, together with strategies for each bank and the policy maker, such that every agent is best responding to the strategies of others. Because the strategy sets of banks and the policy maker are more complex, we discuss them in the context of each policy regime separately in Sections 3.3 and 3.4. In this section, we discuss the types of withdrawal strategies that depositors may play in equilibrium.

Because depositors only care about $t = 1$ consumption when they are impatient, withdrawing at $t = 2$ is a strictly dominated action in this case and any equilibrium strategy profile will have $y_i(1, s) = 1$ for all $s$. The interesting question is how depositors will behave in each state when they are patient. We focus on symmetric equilibria, in which all depositors follow the same strategy, and on equilibria in which patient depositors choose to wait until period 2 to withdraw when the fundamental state is $L$. The latter restriction serves only to simplify the presentation; we focus on crises that occur when the fundamental shock is bad and not when it is good. We also impose a normalization on the sunspot variable to eliminate equilibria that are equivalent up to a relabelling of the sunspot states. In particular, we study equilibria in which the measure of withdrawals at $t = 1$ is at least as large in state $H_\beta$ as in state $H_\alpha$. In other words, we assume that depositors potentially view $\alpha$ to be the "good" sunspot state and $\beta$ the "bad" state rather than the other way around.[5] Formally, while an individual depositor can follow any strategy 3.3, we only study equilibria in which the profile of withdrawal strategies lies in the set

$$Y = \left\{ \begin{array}{c} y : y_i(\omega_i, L) = \omega_i \text{ for all } i \text{ and} \\ \\ \lambda\left(y_i\left(2, H_\beta\right) = 1\right) \geq \lambda\left(y_i\left(2, H_\alpha\right) = 1\right) \end{array} \right\}, \tag{3.4}$$

where $\lambda$ is the measure of depositors following a strategy with the indicated property.

[5] Focusing on the opposite case, where the measure of early withdrawals is weakly larger in state $H_\alpha$ than in state $H_\beta$, would lead to exactly the same results if the probabilities of states $\alpha$ and $\beta$ are reversed. What matters is the set of possible probability distributions over actions and not the labels of the states.

We refer to an event in which some patient depositors choose to withdraw in period 1 as a *run*. Note that the number of early withdrawals is large in a run for two distinct reasons: a higher-than-normal fraction of the population is impatient in state $H$ and some patient depositors are also withdrawing early. In this way, a run in this model consists of a shock to fundamentals whose effect is amplified by the (endogenous) decisions of depositors.

In this setting, two distinct types of runs may arise. We say that a run is driven by expectations if patient depositors' withdrawal behavior depends on the realization of the sunspot variable. In contrast, a run is driven by fundamentals if each depositor's optimal action is independent of the actions of other depositors and, hence, of the sunspot variable. We introduce the following definitions to formalize this distinction.

**Definition 1:** An economy is *weakly fragile* if there is an equilibrium in which depositors play strategy profile

$$y^E : y_i\left(\omega_i, s\right) = \left\{ \begin{array}{c} \omega_i \\ 1 \end{array} \right\} \text{ for } s = \left\{ \begin{array}{c} L,\, H_\alpha \\ H_\beta \end{array} \right\} \text{ for all } i. \tag{3.5}$$

In other words, we say that an economy is weakly fragile if there exists an equilibrium in which all depositors condition their withdrawal decisions in fundamental state $H$ on the realization of the sunspot variable. In this sense, a weakly-fragile economy is susceptible to an expectations-driven bank run. In contrast, we will say that an economy is strongly fragile if a run necessarily occurs whenever the realization of withdrawal demand is high.

**Definition 2:** An economy is *strongly fragile* if the only equilibrium profile of withdrawal strategies $y \in Y$ is

$$y^F : y_i\left(\omega_i, s\right) = \left\{ \begin{array}{c} \omega_i \\ 1 \end{array} \right\} \text{ for } s = \left\{ \begin{array}{c} L \\ H \end{array} \right\} \text{ for all } i. \tag{3.6}$$

When an economy is strongly fragile, the expectations-driven run specified in 3.5 is inconsistent with equilibrium because withdrawing early is a dominant action for patient depositors when the fundamental state is $H$. Instead, depositors necessarily follow 3.6 in equilibrium and bank runs are driven solely by fundamentals. Lastly, if there is no equilibrium in which patient depositors withdraw early in some state, we say that the economy is not fragile.

**Definition 3:** An economy is *not fragile* if the only equilibrium profile of withdrawal strategies $y \in Y$ is the no-run profile

$$y^N : y_i\left(\omega_i, s\right) = \omega_i \text{ for all } s, i. \tag{3.7}$$

We show in the analysis below that, under a given policy regime, an economy fits into exactly one of these three categories, which we refer to as the *fragility type* of the economy under that regime.

In the next two sections, we study fragility and equilibrium allocations under the two different policy regimes. In Section 5, we then ask when the policy maker should be allowed to intervene and when intervention should be prohibited. Of particular interest is the extent to which the answer to this question depends on the fragility type of the economy, that is, the extent to which the desirability of intervention depends on whether the economy is susceptible to runs driven by expectations or by fundamentals.

## 3.3 Equilibrium with no intervention

In this section, we study equilibrium outcomes under the policy regime with no intervention, in which taxes are collected and the public good is provided at the beginning of $t = 1$ (as shown in Figure 3.1). In this regime, the same amount of tax $\tau$ will be collected from each bank and the same level of the public good will be provided in all states, that is

$$g_s = \tau \text{ for all } s. \tag{3.8}$$

We begin the analysis of equilibrium by finding the best responses of banks and the policy maker to an arbitrary profile of withdrawal strategies $y$ and to each other's actions. With these responses in hand, we then ask what profiles $y$ are part of an equilibrium in a given economy.

### 3.3.1 The best-response allocation

Given a profile of withdrawal strategies for its depositors, bank $j$ will allocate its available resources across depositors to maximize the sum of their expected utilities, taking as given the actions of other banks and the policy maker. In principle, a bank can distribute its resources in any way that is consistent with depositors' withdrawal decisions and its own information set. We can, however, simplify matters considerably by determining the general form an efficient response to any strategy profile $y$ must take. A bank knows that at least a fraction $\pi_L$ of its depositors will withdraw in period 1 in both states. As the first $\pi_L$ withdrawals take place, therefore, the bank is unable to make any inference about the state and will choose to give the same level of consumption to each non-monitored depositor who withdraws; let $c_1^j$ denote this amount for bank $j$. Similarly, the bank will choose to give an amount $\hat{c}_1^j$ to each monitored depositor who withdraws.

The bank will be able to infer the fundamental state after $\pi_L$ withdrawals have been made by observing whether or not withdrawals continue. It will also observe the sunspot state at this point and will thus know both what fraction of its depositors are impatient and whether or not a run is underway. The bank can use this information to calculate the fraction of its remaining depositors who are impatient, which we denote $\hat{\pi}_s$. We assume that, once the state has been revealed, each bank is able to efficiently allocate its available resources among its remaining depositors, even if a run is underway. In particular, we assume that the remaining patient depositors do not withdraw early, but instead withdraw in period 2.[6] The efficient allocation of bank $j$'s remaining resources gives a common amount of consumption, denoted $c_{1s}^j$, to each remaining impatient depositor in period 1 and a common amount $c_{2s}^j$ to each remaining patient depositor in period 2. These amounts will be chosen to maximize the average utility of those depositors who have not yet

[6]None of our results depend on this assumption. The issue of how banks and policy makers react to a run, and how this reaction affects the behavior of those depositors who have not yet withdrawn, is quite interesting. Ennis and Keister (2010) show how a model similar to ours can be used to study this interplay between the actions of depositors and the reactions of policy makers. The outcome we study here, where a run ends after $\pi_L$ withdrawals, is one equilibrium that would emerge in such a setting. Focusing on this one outcome allows us to simplify the notation and focus more clearly on the distinction between expectations-driven and fundamentals-driven runs.

withdrawn.[7]

This reasoning shows that a best-response strategy for bank $j$ can be summarized by a vector $\left(c_1^j, \hat{c}_1^j, \left\{c_{1s}^j, c_{2s}^j\right\}_{s\in S}\right)$. We can derive the elements of this vector by working backward, starting with the allocation of the bank's remaining resources after it learns the state.

**Post-crisis payments.** Let $\psi_s^j$ denote the quantity of resources available to bank $j$, in per-depositor terms, after a fraction $\pi_L$ of its depositors have withdrawn. The bank will distribute these resources to solve

$$V\left(\psi_s^j; \hat{\pi}_s\right) \equiv \max_{\left\{c_{1s}^j, c_{2s}^j\right\}} \left(1-\pi_L\right)\left(\hat{\pi}_s u\left(c_{1s}^j\right) + \left(1-\hat{\pi}_s\right) u\left(c_{2s}^j\right)\right) \tag{3.9}$$

subject to the resource constraint

$$\left(1-\pi_L\right)\left(\hat{\pi}_s c_{1s}^j + \left(1-\hat{\pi}_s\right)\frac{c_{2s}^j}{R}\right) \leq \psi_s^j$$

and appropriate non-negativity conditions. Letting $\mu_s^j$ denote the multiplier associated with the resource constraint, the solution to this problem is characterized by the conditions

$$u'\left(c_{1s}^j\right) = Ru'\left(c_{2s}^j\right) = \mu_s^j. \tag{3.10}$$

**Early payments.** As the first $\pi_L$ depositors withdraw, bank $j$ is unable to make any inference about the state. The bank will choose the amount it gives to each monitored depositor, $\hat{c}_1^j$, and to each non-monitored depositor, $c_1^j$, to maximize

$$\pi_L\left[\sigma u\left(\min\left\{\hat{c}_1^j, \tilde{c}_1\right\}\right) + \left(1-\sigma\right) u\left(c_1^j\right)\right] + \sum_{s\in S} q_s V\left(1-\tau-\pi_L\left(\sigma\hat{c}_1^j + \left(1-\sigma\right)c_1^j\right); \hat{\pi}_s\right),$$

where $\tilde{c}_1$ denotes the cap for the consumption of monitored depositors set by the policy maker, which bank $j$ takes as given. The min term in this expression shows that any resources above the cap will be confiscated from these depositors. Looking first at the optimal choice for non-monitored

[7] The fact that this allocation is efficient implies that there is no role for regulation in improving the allocation of resources among the remaining $(1-\pi_L)$ depositors under either policy regime. For this reason, our assumption that the policy maker monitors a fraction $\sigma$ of only the first $\pi_L$ depositors to withdraw is without any loss of generality.

depositors, it is characterized by the first-order condition

$$u'\left(c_1^j\right)=\sum_{s\in S} q_s\mu_s^j. \tag{3.11}$$

This condition says that the bank will allocate resources to equate the marginal utility of a non-monitored depositor to the expected marginal utility from private consumption for the remaining $(1-\pi_L)$ depositors. In the absence of the cap $\tilde{c}_1$, the first-order condition for the consumption of monitored depositors would be identical to 3.11. The bank's optimal choice is, therefore, to give each monitored depositor the lesser of $c_1^j$, as defined in 3.11, and the cap set by the policy maker,

$$\hat{c}_1^j=\min\left\{c_1^j,\tilde{c}_1\right\}. \tag{3.12}$$

Since all banks face the same optimization problem, they will all choose the same levels of $c_1^j$ and $\hat{c}_1^j$. As a result, all banks will have the same level of resources $\psi_s^j$ available in a given state after taxes have been collected and the first $\pi_L$ withdrawals have been made. This fact, in turn, implies that they all face the same optimization problem 3.9 and will choose the same values of $\left(c_{1s}^j, c_{2s}^j\right)$ in each state. We can, therefore, simplify the notation slightly by omitting the $j$ subscripts when referring to the best-response payments $\left(c_1,\hat{c}_1,\{c_{1s},c_{2s}\}_{s\in S}\right)$.

**Prudential regulation.** Like the banks, the policy maker is unable to make any inference about the state $s$ as the first $\pi_L$ withdrawals are made. When he encounters one of these depositors, the policy maker will choose to confiscate any resources she has above some cutoff amount $\tilde{c}_1$. The optimal cutoff value maximizes

$$\sigma\pi_L u\left(\tilde{c}_1\right)+\sum_{s\in S} q_s V\left(1-\tau-\pi_L\left(\sigma\tilde{c}_1+(1-\sigma)\,c_1\right);\hat{\pi}_s\right). \tag{3.13}$$

The policy maker recognizes that any confiscated resources will be rebated lump-sum to banks and, therefore, banks' remaining resources per depositor, $\psi$, will depend on the actual consumption levels of both monitored depositors, $\tilde{c}_1$, and non-monitored depositors, $c_1$.[8] The solution to this

[8]Recall, however, that the decision rule 3.12 ensures that no funds are actually confiscated in equilibrium.

problem is characterized by the first-order condition

$$u'\left(\tilde{c}_1\right) = \sum_{s\in S} q_s \mu_s, \tag{3.14}$$

which is exactly the same as the condition governing an individual bank's choice in 3.11. In other words, in the policy regime with no intervention, banks' incentives are not distorted; the early payments $c_1$ are set at exactly the level a benevolent policy maker would choose,

$$c_1\left(y\right) = \tilde{c}_1\left(y\right) \quad \text{for all } y, \tag{3.15}$$

and the regulatory policy is never binding. In the remainder of this section, we use the relationship in 3.15 to simplify the notation by using $c_1$ to represent the consumption of both monitored and non-monitored depositors.

**The tax rate.** When choosing the tax rate at the beginning of $t = 1$, the policy maker recognizes that banks will allocate the resources available to them as described above and that prudential regulation will be non-binding. Taking banks' allocation rules into account and using 3.15, we can write the policy maker's objective as

$$\pi_L u\left(c_1\left(\tau\right)\right) + \sum_{s\in S} q_s V\left(1 - \tau - \pi_L c_1\left(\tau\right); \hat{\pi}_s\right) + v\left(\tau\right),$$

where the notation indicates that the payment $c_1$ will depend on the tax rate $\tau$, as will banks' remaining resources after the state has been revealed. The first-order condition characterizing the policy maker's optimal choice is

$$\pi_L u'\left(c_1\left(\tau\right)\right)\frac{dc_1\left(\tau\right)}{d\tau} - \sum_{s\in S} q_s\mu_s\left(1 + \pi_L\frac{dc_1\left(\tau\right)}{d\tau}\right) + v'\left(\tau\right) = 0.$$

Using banks' decision rule for choosing $c_1$ in 3.11, this condition simplifies to

$$v'\left(\tau\right) = \sum_{s\in S} q_s\mu_s. \tag{3.16}$$

In other words, when the policy maker chooses the tax rate at the beginning of the period, the optimal choice equates the marginal value of public consumption with the expected marginal value of private consumption.[9]

For any profile $y$ of withdrawal strategies, the discussion above shows how the best responses of banks and the policy maker are summarized by the vector

$$\mathbf{c}^{NI}(y) \equiv \left(c_1^{NI}, \tilde{c}_1^{NI}, \left\{c_{1s}^{NI}, c_{2s}^{NI}\right\}_{s\in S}, g^{NI}\right),$$

which we refer to as the *best-response allocation* associated with $y$ under the policy regime with no intervention. The elements of this allocation are completely characterized by equations 3.8, 3.10, 3.11, 3.15, 3.16, and the resource constraint in each state. We provide an explicit derivation of this allocation in appendix C.1. With these best responses in hand, we next ask what profiles $y$ emerge as equilibria under this policy regime.

### 3.3.2 Fragility

A profile of withdrawal strategies $y^*$ is part of an equilibrium under the policy regime with no intervention if each depositor is choosing the strategy $y_i^*$ that maximizes her own expected utility, taking as given the strategies of other depositors and the allocation $\mathbf{c}^{NI}(y^*)$ that results from the best-responses of banks and the policy maker to those strategies. In Section 3.2.3, we defined the fragility type of an economy based on which withdrawal strategy(ies) are part of an equilibrium. Our first proposition determines which of these types applies to a given economy.[10]

**Proposition 5.** *Under the policy regime with no intervention, the economy is:*

(*a*) *weakly fragile if and only if* $c_{2H_\alpha}^{NI}\left(y^E\right) \geq c_1^{NI}\left(y^E\right) \geq c_{2H_\beta}^{NI}\left(y^E\right)$,

(*b*) *strongly fragile if and only if* $c_1^{NI}\left(y^E\right) > c_{2H_\alpha}^{NI}\left(y^E\right)$, *and*

(*c*) *not fragile if and only if* $c_1^{NI}\left(y^E\right) < c_{2H_\beta}^{NI}\left(y^E\right)$.

*Proof.* See Appendix C.2.1. □

[9]Notice that, while the policy maker can use $\tau$ to influence banks' choice of $c_1$, as well as his own future choice of $\tilde{c}_1$, the term $dc_1/d\tau$ does not appear in 3.16. This fact reflects an envelope result: $c_1$ and $\tilde{c}_1$ are already being set efficiently from the policy maker's current point of view. Hence, there is no benefit in deviating from 3.16 in an attempt to influence these choices.

[10]Proofs of selected propositions are provided in appendix C.2.

This result shows that determining the fragility type of a given economy only requires calculating the best-response allocation to the single strategy profile $y^E$ defined in 3.5. If this profile together with the best responses of banks and the policy maker, $\mathbf{c}^{NI}\left(y^E\right)$, form an equilibrium, then the economy is weakly fragile by definition. If not, the proposition provides a simple test for determining whether the economy is strongly fragile or not fragile. In particular, if an individual patient depositor would prefer to withdraw early in state $H_\alpha$, even though the sunspot state is " good" and she expects other patient depositors to wait until $t = 2$, then any equilibrium must feature all patient depositors withdrawing early whenever the fundamental state is $H$. Conversely, if an individual patient depositor would prefer to wait until $t = 2$ in state $H_\beta$ even though the sunspot state is " bad" and she expects all other patient depositors to withdraw early, then patient investors will never withdraw early in equilibrium and the economy is not fragile.

The next result shows that the fragility type of an economy under this regime does not depend on the regulation parameter $\sigma$ nor on the desirability of the public good.

**Proposition 6.** *Under the policy regime with no intervention, the fragility type of an economy is independent of the parameters $\sigma$ and $\delta$.*

The first part of this result is trivial: since prudential regulation is never binding under this regime, the entire allocation $\mathbf{c}^{NI}(y)$ is independent of the fraction $\sigma$ of monitored depositors for any strategy profile $y$. The second part of the result follows from the functional form in 3.1, which implies that preferences over private consumption across states of nature are homothetic. An increase in the parameter $\delta$ would, therefore, raise consumption of the public good while lowering consumption of the private good in each state in proportion, leaving the ratios $c_1^{NI}(y) / c_{2s}^{NI}(y)$ unchanged for any $s$ and any $y$.[11] Depositors' withdrawal incentives are thus independent of the size of the public sector under this policy regime.

Using Proposition 5, it is straightforward to find examples of economies that are strongly fragile under the policy regime with no intervention, as well as economies that are weakly fragile and not fragile. For each of these economies, our interest is in determining whether welfare would

[11] This fact is easily verified using the expressions for the best-response allocation $\mathbf{c}^{NI}$ in appendix C.1.

be increased by allowing the policy maker to intervene by adjusting tax rates after the state has been revealed. As discussed in the Introduction, one view holds that such intervention tends to be desirable when the economy is weakly fragile, but is undesirable when the economy is either strongly fragile or not fragile. To test the validity of this view, we next characterize equilibrium outcomes under a policy regime with intervention.

## 3.4 Equilibrium with intervention

Now suppose the policy maker collects taxes later in period 1, after a fraction $\pi_L$ of depositors have withdrawn. (See Figure 3.1 in Section 3.2.2.) At this point, the policy maker has learned the state and thus knows both the level of fundamental withdrawal demand and whether a run has occurred. The benefit of acting at this later point is that the level of taxes can be state-contingent, which allows for risk sharing between the public and private sectors. The cost is that the policy maker will be tempted to set tax rates in a way that, from an ex ante point of view, will distort banks' incentives to provision for bad outcomes. We analyze equilibrium in the model with such intervention in this section, then study the desirability of allowing intervention in Section 3.5.

### 3.4.1 Bailouts

After a fraction $\pi_L$ of depositors have withdrawn, the policy maker observes whether or not withdrawals stop. If they do, the policy maker is able to infer that the fundamental state is $L$. In this case, we assume the policy maker chooses a single tax rate $\tau_L$ and collects this tax per unit of deposits from all banks. If withdrawals continue past $\pi_L$, however, the policy maker infers that the fundamental state is $H$. The policy maker then observes the sunspot state and the financial condition of each bank before choosing a tax rate $\tau_s^j$ for bank $j$. All tax rates are chosen with the objective of maximizing 3.2 given the current situation and anticipating that each bank will allocate its after-tax resources to solve 3.9. The difference

$$\tau_L - \tau_s^j$$

can be interpreted as the "bailout" of bank $j$ in states $s = H_\alpha, H_\beta$. When fundamental withdrawal demand is high, the policy maker will tend to cut production of the public good in order to help mitigate the decline in private consumption of the remaining depositors in the banking system. In principle, however, this bailout can be either positive or negative; a bank in better-than-average condition might be required to pay a higher-than-normal tax to make up for the poor condition of other banks.[12]

### 3.4.2 The best response allocation

We characterize equilibrium under this regime following the same steps as in Section 3.3. For a given profile $y$ of withdrawal strategies, we first determine the best responses of banks and the policy maker to this profile and to each other's actions. With these responses in hand, we then ask whether the strategy $y_i$ is a best response for depositor $i$ to the strategies of other depositors, banks, and the policy maker.

After a fraction $\pi_L$ of depositors have withdrawn and taxes have been collected, each bank will again allocate its remaining resources to solve the problem in 3.9 and, as before, this allocation is characterized by the first-order conditions in 3.10. We begin, therefore, by studying how the policy maker will intervene, then work backward to determine the consumption of the first $\pi_L$ depositors who withdraw.

**Choosing tax rates.** In state $s$, the policy maker will choose the tax rate $\tau_s^j$ per unit of deposits in bank $j$ to maximize

$$\int V\left(1-\tau_s^j-\pi_L c_1^j;\hat{\pi}_s\right)d\phi\left(j\right)+v\left(\tau_s\right),$$

where $\phi$ represents the distribution of investors across banks and $\tau_s$ denotes total tax revenue in state $s$, that is,

$$\tau_s\equiv\int\tau_s^j d\phi\left(j\right).$$

The tax rate must be the same for all banks in fundamental state $L$, but may differ across banks

[12]The assumption that the policy maker does not set bank-specific tax rates in fundamental state $L$ is designed to ensure that banks have an incentive to provision for $t=2$ withdrawals in normal times. It can be justified in different ways, for example, by assuming that the detailed monitoring needed to accurately determine a bank's financial condition is only worthwhile in state $H$, or by appealing to reputational considerations that would arise in normal times in a more fully dynamic model. For our purposes, the important thing is that the policy maker's inability to commit creates a distortion in banks' incentives with respect to those states where a crisis occurs.

in state $H$. The solution will, therefore, equate the marginal value of public consumption in fundamental state $L$ to the marginal value of private consumption averaged across banks,

$$v'\left(\tau_L\right) = \int \mu_s^j d\phi\left(j\right).$$

The marginal value of public consumption in fundamental state $H$, in contrast, will be set equal to the marginal value of private consumption in every bank $j$ ,

$$v'\left(\tau_s\right) = \mu_s^j \text{ for all } j, \text{ for } s = H_\alpha, H_\beta. \tag{3.17}$$

In other words, when a crisis occurs, the policy maker will set the tax rate $\tau_s^j$ to equalize the consumption levels of the remaining depositors across banks, meaning that a bank that is in worse financial condition (because it set $c_1^j$ higher and gave away more resources to the first $\pi_L$ depositors) will receive a larger bailout. As a result, the resources available to bank $j$ after taxes have been collected in a crisis state will depend on aggregate economic conditions and not on the bank's own actions. Specifically, we have

$$\psi_s^j = 1 - \tau_s - \pi_L \bar{c}_1 \quad \text{for all } j, \text{ for } s = H_\alpha, H_\beta, \tag{3.18}$$

where $\bar{c}_1$ is defined to be the average early payment across all banks and all depositors,

$$\bar{c}_1 \equiv \int \left(\sigma \hat{c}_1^j + \left(1-\sigma\right) c_1^j\right) d\phi\left(j\right).$$

The incentive problems caused by this bailout policy are clear: a bank with fewer remaining resources (because it chose a higher value of $c_1^j$) will be charged a lower tax, effectively receiving a larger "bailout". This bailout policy will lead all banks to set $c_1^j$ too high from a social point of view.

Notice that this problem arises even when $\delta = 0$ and there is no value associated with the public good. In that case, the policy maker will set $\tau_L = 0$ and collect no revenue in normal times. When a crisis occurs, total tax revenue $\tau_s$ will be set to zero, but the policy maker will still choose to intervene by taxing banks that have more resources than average and making transfers to banks that have fewer resources than average. In equilibrium, of course, all banks will make the same choices and no taxes/transfers will occur. Nevertheless, the fact that these transfers would occur off the equilibrium path of play affects banks' decisions on the equilibrium path, as we show below.

**Early payments.** As the first $\pi_L$ withdrawals take place, bank $j$ will choose the amount it gives to each monitored depositor, $\hat{c}_1^j$, and to each non-monitored depositor, $c_1^j$, to maximize

$$\pi_L\left[\sigma u\left(\min\left\{\hat{c}_1^j, \tilde{c}_1\right\}\right)+(1-\sigma) u\left(c_1^j\right)\right]+q_L V\left(1-\tau_L-\pi_L\left(\sigma \hat{c}_1^j+(1-\sigma) c_1^j\right) ; \hat{\pi}_L\right) \tag{3.19}$$
$$+\sum_{s=H_\alpha, H_\beta} q_s V\left(1-\tau_s-\pi_L \bar{c}_1 ; \hat{\pi}_s\right).$$

Since there are no bailouts in state $L$, the bank recognizes that giving an extra unit of resources to the first $\pi_L$ depositors will leave one unit less for the remaining depositors in that state. However, when the fundamental state is $H$, the policy maker will intervene in such a way that the bank's remaining resources will be given by 3.18, independent of its choice of $c_1^j$. As a result, the terms on the second line of 3.19 are fixed from the individual bank's point of view and the first-order condition characterizing the solution to this problem is

$$u'\left(c_1^j\right)=q_L \mu_L^j. \tag{3.20}$$

Comparing 3.20 with 3.11 shows the distortion created by intervention: bank $j$ no longer has an incentive to provision for the fundamental state $H$. Instead, the bank will balance the marginal value of resources for the earliest withdrawals against the marginal value of resources for later withdrawals in fundamental state $L$ only. As a result, the bank will tend to set $c_1^j$ too high from a social point of view. For monitored depositors, the bank's optimal choice again follows 3.12; it will give these agents the lesser of $c_1^j$, now defined in 3.20, and the cap $\tilde{c}_1$ set by the policy maker.

As above, all banks face the same decision problem and will choose the same values of $c_1^j$. Together with the bailout policy in 3.18, this fact implies that all banks also face the same decision problem in choosing the later payments $\left(c_{1s}^j, c_{2s}^j\right)$ and will again select the same values. We can, therefore, omit the $j$ subscripts to simplify the notation in what follows.

**Prudential regulation.** When the policy maker encounters one of the first $\pi_L$ depositors to withdraw, he will again choose the cutoff value $\tilde{c}_1$ to maximize 3.13, with the adjustment that tax revenue $\tau_s$ now varies across states. The key difference between the policy maker's objective function and that of an individual bank in 3.19 is that the policy maker recognizes that giving a unit of resources to one of the first $\pi_L$ depositors decreases the resources available for the remaining depositors in all states, whereas the intervention policy in 3.18 makes this effect external to an individual bank when the fundamental state is $H$. The first-order condition that characterizes the policy maker's optimal choice is again given by 3.14, which shows how prudential regulation is now used to correct the distortion created by intervention. When a depositor is monitored by the policy maker, her marginal utility of consumption is equated to the expected future marginal value of consumption, taking all states into account, which is precisely what an individual bank chooses to do when there is no intervention and incentives are not distorted.

The best-response allocation under the policy regime with intervention, denoted

$$\mathbf{c}^I(y) \equiv \left(c_1^I, \tilde{c}_1^I, \left\{c_{1s}^I, c_{2s}^I, g_s^I\right\}_{s\in S}\right),$$

is characterized by equations 3.10, 3.14, 3.17, 3.20, and the resource constraint in each state. We provide an explicit derivation of the allocation in appendix C.1. It is straightforward to show that prudential regulation is always active in this allocation, that is, the policy maker's cap $\tilde{c}_1$ is strictly lower than the consumption of non-monitored depositors $c_1$,

$$\tilde{c}_1^I(y) < c_1^I(y) \text{ for all } y \in Y. \tag{3.21}$$

### 3.4.3 Fragility

We now use the allocation $\mathbf{c}^I$ to identify conditions under which an economy is susceptible to runs driven by either expectations or fundamentals under the policy regime with intervention. We begin with a characterization result similar to Proposition 5. As in Keister (2016), we assume the states in which intervention occurs are relatively rare, with

$$q_{H_\alpha} + q_{H_\beta} < \frac{R-1}{R}, \tag{3.22}$$

which simplifies the analysis by placing an upper bound on the size of the incentive distortion. For notational convenience, we define

$$\mathcal{E}\left(\mathbf{c}^{I}\left(y\right)\right) \equiv \sigma u\left(\tilde{c}_{1}^{I}\left(y\right)\right) + \left(1-\sigma\right) u\left(c_{1}^{I}\left(y\right)\right), \tag{3.23}$$

which represents the expected utility of a depositor who is among the first $\pi_L$ withdrawals before she knows whether or not she will be monitored. We then have the following result.

**Proposition 7.** *Under the policy regime with intervention, the economy is:*

($a$) *weakly fragile if and only if* $u\left(c_{2H_\alpha}^{I}\left(y^E\right)\right) \geq \mathcal{E}\left(\mathbf{c}^{I}\left(y^E\right)\right) \geq u\left(c_{2H_\beta}^{I}\left(y^E\right)\right)$,

($b$) *strongly fragile if and only if* $\mathcal{E}\left(\mathbf{c}^{I}\left(y^E\right)\right) > u\left(c_{2H_\alpha}^{I}\left(y^E\right)\right)$, *and*

($c$) *not fragile if and only if* $\mathcal{E}\left(\mathbf{c}^{I}\left(y^E\right)\right) < u\left(c_{2H_\beta}^{I}\left(y^E\right)\right)$.

*Proof.* See Appendix C.2.2. □

As with Proposition 5 in Section 3.3, this result demonstrates that every economy has a unique fragility type under a given policy regime and that determining this type only requires calculating the best-response allocation for the single strategy profile $y^E$ defined in 3.5.

The next two propositions study how the fragility type of an economy depends on the effectiveness of regulation, measured by the parameter $\sigma$, and on the importance of the public good, measured by $\delta$. Recall that Proposition 6 showed the fragility type of an economy to be independent of these two parameters under the policy regime with no intervention. These relationships change when intervention is allowed. Let $e$ denote the vector of all parameter values except $\sigma$, so that $e = \left(R, \gamma, \delta, \{q_s, \pi_s\}_{s\in S}\right)$ and an economy is defined by the pair $(e, \sigma)$. Then we have the following result.

**Proposition 8.** *Under the policy regime with intervention, the fragility type of an economy* $(e, \sigma)$ *is weakly decreasing in* $\sigma$.

*Proof.* See Appendix C.2.3. □

In other words, more effective regulation promotes financial stability when the prospect of intervention distorts banks' incentives. The intuition for this result is straightforward. The first-order condition 3.20 illustrates how intervention leads banks to increase their short-term liabilities by offering relatively large payments to the non-monitored depositors who withdraw before the policy reaction occurs. Condition 3.21 shows that the policy maker will cap the consumption of monitored depositors at a lower level. An increase in the fraction of depositors who are monitored thus tends to make withdrawing early less attractive for patient investors. At the same time, the smaller payments made to monitored depositors imply that banks will have more resources left after the first $\pi_L$ withdrawals have been made, which also makes waiting to withdraw at $t = 2$ more attractive. For both of these reasons, more effective regulation lowers the incentive for a patient depositor to run and thus tends to reduce fragility.

The next result highlights the insurance benefit of bailouts: when regulation is sufficiently effective, financial fragility will be lower in economies where the public sector is larger. For this result, we need to impose a fairly weak condition on parameter values:

$$q_{H_\alpha} > \frac{1}{R}\left(1 - \frac{\pi_H(1-\pi_L)}{\pi_L(1-\pi_H)}\left(\frac{(1-\pi_L)R^{\frac{1-\gamma}{\gamma}}}{(\pi_H-\pi_L)+(1-\pi_H)R^{\frac{1-\gamma}{\gamma}}}\right)^{\gamma}\right) \equiv \bar{q}_{H_\alpha}. \tag{3.24}$$

In many economies, the lower bound $\bar{q}_{H_\alpha}$ is negative and this condition is automatically satisfied. In some cases, however (when $R$ is very large, for example), this condition sets a small, positive floor on the probability $q_{H_\alpha}$.

**Proposition 9.** *Under the policy regime with intervention, if 3.24 holds, then for any $e$ there exists $\bar{\sigma} < 1$ such that the fragility type of all economies $(e, \sigma)$ with $\sigma > \bar{\sigma}$ is weakly decreasing in $\delta$.*

*Proof.* See Appendix C.2.4. □

When $\delta$ is higher, the public sector is larger and, as a result, the policy maker will choose bailouts that are larger relative to the level of private consumption. These larger bailouts decrease the losses suffered by investors who are not among the first $\pi_L$ to withdraw and, therefore, tend to lower the incentive for patient depositors to withdraw early. However, there is an offsetting effect: because the larger bailout payments mitigate the effects of a crisis, the policy maker will choose to allow a higher level of consumption for monitored depositors who withdraw before the policy reaction. This fact makes withdrawing early more attractive and tends to increase the incentive for patient depositors to run. In general, either effect can dominate and increasing the parameter $\delta$ can either increase or decrease fragility. Proposition 9 demonstrates that when regulation is sufficiently effective and 3.24 holds, however, the first effect always dominates and having a larger public sector will (weakly) decrease financial fragility.

## 3.5 Comparing Policy Regimes

The analysis in the previous two sections has illustrated the costs and benefits of allowing the policy maker to intervene during a crisis. We now turn to the question of when the benefits outweigh the costs, providing two analytical results followed by some illustrative examples. We first study the case where regulation is very effective, that is, the parameter $\sigma$ is close to one. We show that, in this case, allowing intervention is always desirable, regardless of the fragility type of the economy under each regime. We then study the case where $\delta = 0$, meaning that depositors get no utility from the public good. In this case, we show that there is no insurance benefit from allowing intervention and, as a result, intervention is never desirable. Away from these two limiting cases, either of the two effects can dominate. We use a series of examples to show that intervention tends to be desirable when it improves the economy's fragility type, but can be desirable even if it does not because the increased risk sharing between private and public consumption may more than compensate for the distorted allocation of private consumption.

### 3.5.1 When regulation is very effective

Our first result identifies situations where regulation is effective enough to guarantee that the insurance benefit from intervention outweighs the incentive costs. Specifically, assume investors value the public good $(\delta > 0)$ and fix all parameter values except the effectiveness of prudential regulation $\sigma$. When $\sigma$ is close enough to 1, allowing intervention is always desirable.

**Proposition 10.** *Assume 3.24 holds. For any $e$ with $\delta > 0$, there exists $\overline{\sigma} < 1$ such that allowing intervention strictly increases equilibrium welfare for all economies $(e, \sigma)$ with $\sigma > \overline{\sigma}$.*

*Proof.* See Appendix C.2.5. □

The intuition for this result can be seen in two steps. First, imagine that we hold depositors' withdrawal behavior fixed. When private consumption levels vary across states, an efficient allocation of resources requires public consumption levels to vary across states as well. By collecting higher taxes in good states and lower taxes in bad states, the policy maker helps smooth depositors' private consumption, which raises expected utility. In addition, this type of consumption smoothing lowers the incentive for patient depositors to withdraw early. In fact, the proof of Proposition 10 (see appendix C.2) shows that when $\sigma$ is close enough to one, allowing intervention weakly decreases fragility relative to the regime with no intervention. In other words, when regulation is sufficiently effective, allowing intervention improves both the allocation of resources conditional on depositor behavior *and* depositors' equilibrium withdrawal behavior; hence, it is always desirable.

In a model of expectations-driven runs, Keister (2016) shows that allowing bailouts is always desirable when policy makers can completely offset the associated incentive distortion using Pigouvian taxes. Proposition 10 shows that this type of result obtains even when prudential regulation is somewhat imperfect and, more importantly, regardless of whether runs are driven by expectations or fundamentals.

#### 3.5.2 When the insurance benefit is absent

Our next result focuses on economies where $\delta = 0$, that is, depositors do not value the public good. The policy maker can still collect taxes and monitor some withdrawals, but there is no longer a potential gain from sharing risk between the public and private sectors because the optimal amount of public consumption is zero. In this case, if the incentive distortions associated with bailouts cannot be fully corrected through regulation (that is, $\sigma < 1$), allowing intervention is undesirable.[13]

**Proposition 11.**

For any economy with $\delta = 0$ and $\sigma < 1$, allowing intervention strictly decreases equilibrium welfare.

[13] If regulation is perfectly effective $(\sigma = 1)$, the two policy regimes lead to exactly the same outcome when $\delta = 0$. In this case, the incentive distortion created by intervention is completely corrected through regulation, leaving the allocation of consumption across depositors unchanged.

*Proof.* See Appendix C.2.6. □

This result highlights the importance of the insurance benefit of bailouts in our setting. When this benefit is absent, allowing intervention still distorts banks' incentives because the policy maker is able to reallocate resources across banks following a crisis. This distortion leads to a misallocation of resources and lowers depositors' welfare if regulation is imperfect. In this special case, our model yields the same prescription as others in the literature in which bailouts distort incentives but do not generate any ex ante benefits; see, for example, Farhi and Tirole (2012) and Chari and Kehoe (2013). In this way, Proposition 11 demonstrates that the desirability of prohibiting intervention in these frameworks stems not from the assumptions about what causes a crisis (fundamentals vs. expectations), but rather from the fact that there is no insurance benefit from bailouts that could potentially offset the distortion in incentives.[14]

### 3.5.3 Examples

Propositions 10 and 11 identify situations in which one of the two competing effects – incentives or insurance – is clearly dominant and thus determines the optimal policy choice. In between these limiting cases, interesting patterns arise. We illustrate some of these patterns using a series of three related examples.

**An economy that is weakly fragile with no intervention.** For our first example, we set $R = 1.05$, $\pi_L = 0.45$, $\pi_H = 0.55$, $q_{H_\alpha} = q_{H_\beta} = 0.02$ and $\gamma = 4$. At these values, the economy is weakly fragile under the policy regime with no intervention for all $(\sigma, \delta)$ pairs.[15] Panel (a) of Figure 3.2 depicts the fragility type of the economy under the regime with intervention. For a broad range of $(\sigma, \delta)$ pairs in the middle of the panel, the economy is also weakly fragile under this regime. If $\sigma$ and $\delta$ are both large enough, however, the run equilibrium is eliminated and the economy is no longer fragile. If $\sigma$ and $\delta$ are small enough, in contrast, allowing intervention makes withdrawing early a dominant strategy for patient depositors and the economy is strongly fragile.

[14] This type of insurance benefit of bailouts also appears, in different settings, in Cooper *et al.* (2008), Green (2010) and Bianchi (2013).

[15] Recall that Proposition 2 shows the fragility type of an economy under the regime with no intervention to be independent of $\sigma$ and $\delta$.

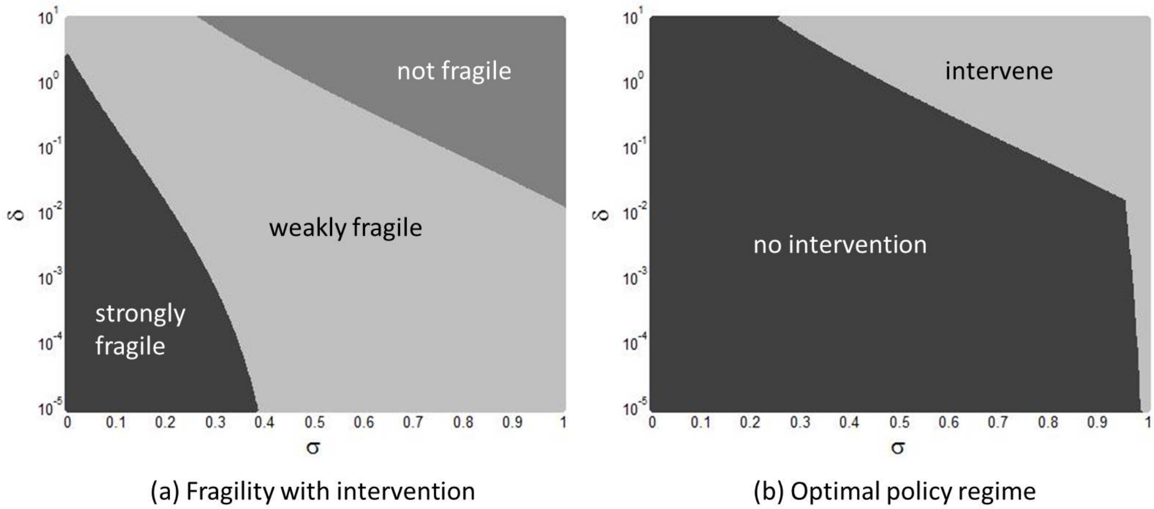
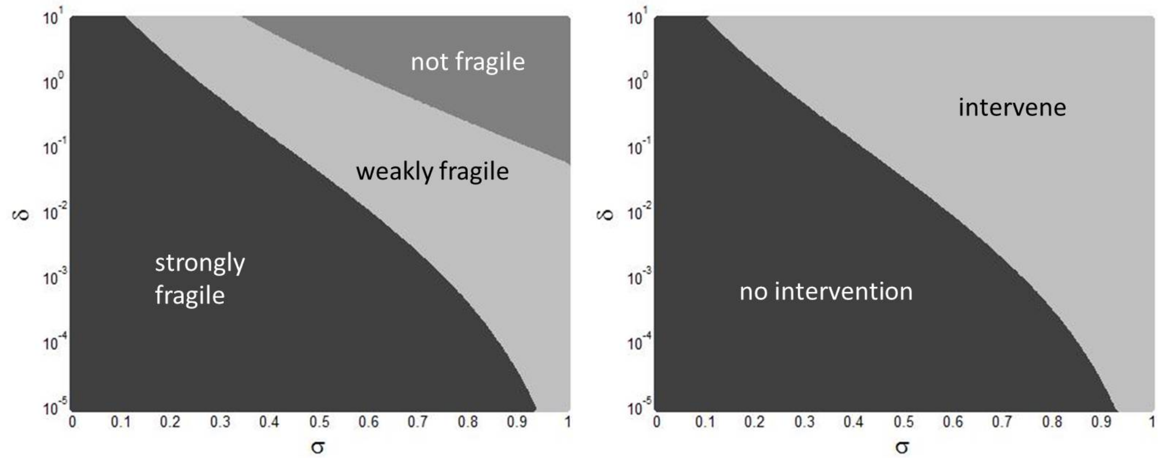



**Figure 3.2:** *An economy that is weakly fragile with no intervention*

Panel (b) of the figure shows which policy regime generates higher welfare. Allowing intervention is desirable in this example in two situations. First, if allowing intervention eliminates the run equilibrium and makes the economy not fragile, then doing so is always desirable. Second, even if allowing intervention leaves the economy weakly fragile, it is desirable whenever $\sigma$ is close enough to one, as established in Proposition 10.

**An economy that is strongly fragile with no intervention.** Now suppose $\pi_H$ is raised to 0.65. This larger value for the fundamental shock makes the economy strongly fragile under the policy regime with no intervention. Panel (a) of Figure 3.3 shows the fragility type of the economy when intervention is allowed. If $\sigma$ and $\delta$ are low enough, the economy remains strongly fragile. For these cases, panel (b) of the figure indicates that intervention is undesirable. When $\sigma$ and $\delta$ are higher, however, the fragility type of the economy improves under the regime with intervention, becoming either weakly fragile or, if $\sigma$ and $\delta$ are high enough, not fragile. In both of these cases, panel (b) of the figure indicates that allowing intervention raises welfare.

The example in Figure 3.2 showed that allowing intervention may be desirable because it eliminates a bad equilibrium, moving the economy from weakly fragile to not fragile. The example in Figure 3.3 shows that allowing intervention may be desirable because it *introduces* a better equilibrium. In this case, the economy with no intervention has a unique equilibrium profile of

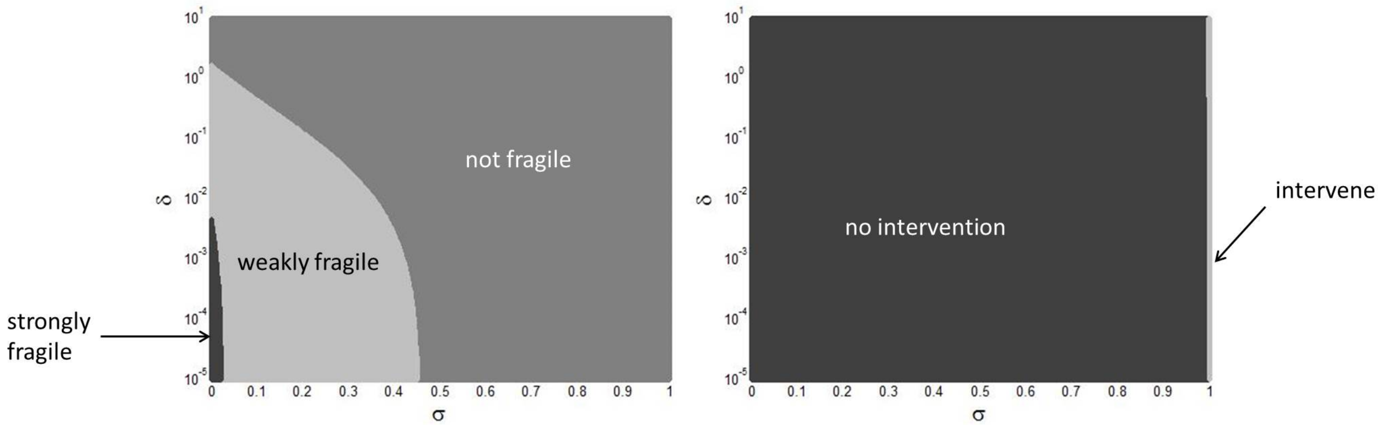



(a) Fragility with intervention (b) Optimal policy regime

**Figure 3.3:** *An economy that is strongly fragile with no intervention*

withdrawal strategies $y^* \in Y$. Bank runs in this equilibrium are driven by fundamentals, which might tempt one to conclude that runs are inevitable and that allowing intervention cannot change the level of fragility or improve welfare. However, as the figure shows, allowing intervention in this case can introduce an equilibrium in which patient depositors only run in state $H_\beta$, rather than in both $H_\alpha$ and $H_\beta$. In this new equilibrium, where bank runs are driven by expectations, depositors have higher expected utility. If $\delta$ and $\sigma$ are larger still, allowing intervention can eliminate runs entirely. This second example illustrates the importance of recognizing that whether runs are driven by fundamentals or expectations can depend on the policy regime in place. Even when runs are driven by fundamentals under one regime, it is possible for their incidence to be lessened or even eliminated under another regime.

**An economy that is not fragile with no intervention.** Figure 3.4 presents the results when $\pi_H$ is lowered back to 0.55 and $\gamma$ is lowered to 2. The smaller coefficient of relative risk aversion leads banks to provide less liquidity insurance and, in this example, makes the economy not fragile under the policy regime with no intervention. Panel (a) of the figure shows how, in terms of fragility, allowing intervention can only make the situation worse in this case. If $\sigma$ and $\delta$ are high enough the economy remains not fragile under this regime; otherwise it can become weakly or even strongly fragile. Panel (b) of the figure shows that, in this case, prohibiting intervention is the optimal policy for the vast majority of $(\sigma, \delta)$ pairs. However, in line with Proposition 10, allowing intervention is desirable if $\sigma$ is very close to 1.




(a) Fragility with intervention

(b) Optimal policy regime

**Figure 3.4:** *An economy that is not fragile with no intervention*

Taken together, these three examples present a clear pattern. Allowing intervention tends to reduce fragility and raise welfare in the upper-right corner of the graphs, where the insurance benefit is significant and regulation is effective in mitigating the incentive distortion. Prohibiting intervention tends to be desirable in the lower-left corner, where the potential for risk-sharing is small and regulation is ineffective. While the precise boundary between these two areas depends on the particulars of the economy, including whether runs are driven by expectations or by fundamentals when there is no intervention, the same general pattern arises in each case. The examples thus illustrate how the key tradeoff facing policy makers, as well as the factors that should guide the decision to allow or prohibit intervention, are independent of the underlying cause of bank runs.

## 3.6 Concluding Remarks

Policy makers and academics around the world are currently engaged in a wide-ranging discussion about how to best reform banking and financial regulation in light of recent experience. There is widespread agreement that the anticipation of being bailed out in the event of a crisis distorts the incentives of financial institutions and their investors, leading them to take actions that are socially inefficient and may, in addition, leave the economy more susceptible to a crisis. There is no consensus, however, about the best way to design a policy regime to mitigate these problems.

A number of recent papers examine bailout policy in models that include moral hazard concerns and account for the possible time inconsistency of policy makers' objectives.[16] Each of these papers makes some assumption about the underlying causes of a crisis: it either is the unique equilibrium outcome following some real shock to the economy or it arises, in part, from the self-fulfilling beliefs of agents in the model. There is a long-standing debate about which of these two approaches best captures the complex array of forces that combine to generate real-world financial crises.[17] Financial crises are infrequent events and there is a limited amount of available data that can be used to distinguish between the two approaches. Existing empirical work focuses on establishing a correlation between economic fundamentals and the occurrence of banking panics. Miron (1986), Gorton (1988) and others argue that such a correlation implies that runs are caused by shifts in these fundamentals. Ennis (2003) points out, however, that models of self-fulfilling bank runs will tend to generate this same type of correlation under reasonable equilibrium selection rules, so that the presence of this correlation alone cannot be used to distinguish between the two views. Moreover, establishing the importance (or unimportance) of self-fulfilling beliefs in causing a run requires answering a counterfactual question: would an individual depositor have withdrawn even if she expected other depositors to remain invested? Answering such questions on the basis of data from observed crises is intrinsically difficult.[18]

This ongoing debate would seem to present a serious hindrance to using such models for policy analysis. Without knowing the underlying cause of observed crises, how can one decide

---

[16] See, for example, Bianchi (2013), Chari and Kehoe (2013), Farhi and Tirole (2012), Green (2010), and Keister (2016).

[17] See Kindleberger (1978), Gorton (1988), Allen and Gale (1998), Allen and Gale (2007), and Goldstein and Pauzner (2005) for different views of this debate.

[18] Some authors have argued that the degree to which depositors discriminate between banks during a panic provides evidence on the underlying cause of the event. See, for example, Saunders and Wilson (1996), Calomiris and Mason (1997), Calomiris and Mason (2003), Schumacher (2000), and Chen *et al.* (2010). However, the Ennis (2003) critique again applies: all but the simplest models of self-fulfilling runs will tend to generate the same correlations as a model of fundamentals-based runs.

which type of model should be used to evaluate alternative policy regimes? We have shown how, in some cases, it is possible to perform meaningful policy analysis without taking a stand on the question of whether financial crises are driven by expectations or fundamentals. We constructed a model in which, depending on parameter values, a bank run may be driven by either of these two forces. We used this model to evaluate the desirability of allowing policy makers to intervene in the event of a crisis and provide bailouts. We showed that the same broad policy prescription comes out of the model regardless of whether runs are driven by expectations or fundamentals. In particular, intervention should be permitted only when prudential regulation and supervision are sufficiently effective that the insurance benefit from bailouts outweighs the resulting incentive distortion.

While our focus in this paper is on a single policy issue, we also aim to make a more general point. Much effort has been devoted to trying to determine the extent to which financial crises can be caused by self-fulfilling beliefs. This work has generated important insights, but has not led to a definitive answer to this difficult question. The lack of a clear answer does not imply, however, that the insights gained from this work cannot be used to inform the current policy debate. Our analysis here shows how these insights can be useful in studying one particular policy issue. Future work could examine other issues in banking and financial stability policy, or could seek to identify conditions under which a more general invariance result might hold.

# References


ACHARYA, V. V. and STEFFEN, S. (2015). The “greatest” carry trade ever? Understanding eurozone bank risks. *Journal of Financial Economics*, **115** (2), 215–236.

— and VISWANATHAN, S. (2011). Leverage, Moral Hazard, and Liquidity. *Journal of Finance*, **66** (1), 99–138.

ADMATI, A. R., DEMARZO, P. M. P., HELLWIG, M. F. and PFLEIDERER, P. C. (2013). The leverage ratchet effect. *Working paper, Stanford University*.

ALLEN, F. (2014). How Should Bank Liquidity be Regulated? *Speech at Federal Reserve Bank of Atlanta*.

— and GALE, D. (1998). Optimal Financial Crises. *The Journal of Finance*, **53** (4), 1245–1284.

— and — (2007). *Understanding Financial Crises*. Oxford University Press.

ANDOLFATTO, D., NOSAL, E. and WALLACE, N. (2007). The role of independence in the Green-Lin Diamond-Dybvig model. *Journal of Economic Theory*, **137** (1), 709–715.

BAHAJ, S. and MALHERBE, F. (2016). A positive analysis of bank behaviour under capital requirements. *Unpublished*.

BECKER, B. and IVASHINA, V. (2014). Financial Repression in the European Sovereign Debt Crisis. *Unpublished*.

BIANCHI, J. (2013). Efficient Bailouts? *NBER Working Paper*.

BOE (2015). *One Bank Research Agenda*. Tech. rep., Bank of England.

BROWNING, M. and COLLADO, M. D. (2001). The Response of Expenditures to Anticipated Income Changes : Panel Data Estimates. *American Economic Review*, **91** (3), 681–692.

BRUNNERMEIER, M. K. (2009). Deciphering the Liquidity and Credit Crunch 2007–2008. *Journal of Economic Perspectives*, **23** (1), 77–100.

CALOMIRIS, C. W. and MASON, J. R. (1997). Contagion and Bank Failures During the Great Depression: The June 1932 Chicago Banking Panic. *The American Economic Review*, **87** (5), 863–883.

— and — (2003). Fundamentals , Panics , and Bank Distress During the Depression. *American Economic Review*, **93**, 1615–1647.

CARROLL, C. D. (1992). Buffer Stock Saving and the Permanent Income Hypothesis. *Unpubllished paper*.

Chang, R. and Velasco, A. S. (2000). Financial Fragility and the Exchange Rate Regime. *Journal of Economic Theory*, **92**, 1–34.

Chari, V. V. and Kehoe, P. J. (2013). Bailouts, Time Inconsistency, and Optimal Regulation. *NBER Working Paper*.

Chen, Q., Goldstein, I. and Jiang, W. (2010). Payoff complementarities and financial fragility: Evidence from mutual fund outflows. *Journal of Financial Economics*, **97**, 239–262.

Cooper, R. and Kempf, H. (2013). Deposit Insurance and Orderly Liquidation without Commitment: Can we Sleep Well? *NBER Working Paper*.

—, — and Peled, D. (2008). Is it is or is it ain't my obligation? Regional debt in a fiscal federation. *International Economic Review*, **49** (4), 1469–1504.

— and Ross, T. W. (1998). Liquidity costs and investment distortions. *Journal of Monetary Economics*, **41**, 27–38.

Crosignani, M. (2015). Why are banks not recapitalized during crises? *Unpublished*.

Deaton, A. (1991). Saving and Liquidity Constraints. *Econometrica*, **59** (5), 1221–1248.

Dell'Ariccia, G., Laeven, L. and Suarez, G. A. (2016). Bank Leverage and Monetary Policy's Risk-Taking Channel: Evidence from the United States. *The Journal of Finance*.

Di Maggio, M., Kermani, A. and Ramcharan, R. (2015). Monetary Policy Pass-Through: Household Consumption and Voluntary Deleveraging. *Working Paper*.

Diamond, D. W. and Dybvig, P. H. (1983). Bank Runs , Deposit Insurance , and Liquidity. *Journal of Political Economy*, **91** (3), 401–419.

— and Rajan, R. G. (2011). Fear of Fire Sales, Illiquidity Seeking, and Credit Freezes. *The Quarterly Journal of Economics*, **126** (2), 557–591.

Ennis, H. M. (2003). Economic Fundamentals and Bank Runs. *Federal Reserve Bank of Richmond Economic Quarterly*, **89**, 55–71.

— and Keister, T. (2009). Bank Runs and Institutions: The Perils of Intervention. *American Economic Review*, **99** (4), 1588–1607.

— and — (2010). Banking panics and policy responses. *Journal of Monetary Economics*, **57**, 404–419.

Farhi, E. and Tirole, J. (2012). Collective Moral Hazard, Maturity Mismatch, and Systemic Bailouts. *American Economic Review*, **102** (1), 60–93.

French, K. R., Baily, M. N., Campbell, J. Y., Cochrane, J. H., Diamond, D. W., Duffie, D., Kashyap, A. K., Mishkin, F. S., Rajan, R. G., Scharfstein, D. S., Shiller, R. J., Song, H., Matthew, S., Slaughter, J., Stein, J. C. and Stulz, R. M. (2010). *The Squam Lake Report: Fixing the Financial System*. Princeton University Press.

Fuster, A. and Willen, P. S. (2013). Payment Size, Negative Equity, and Mortgage Default. *NBER Working Paper*.

Goldstein, I. and Pauzner, A. (2005). Demand-deposit contracts and the probability of bank runs. *Journal of Finance*, **60** (3), 1293–1327.

Goodhart, C. A. E., Kashyap, A. K., Tsomocos, D. P. and Vardoulakisd, A. P. (2013). An Integrated Framework for Analyzing Multiple Financial Regulations. *International Journal of Central Banking*.

Gorton, G. (1988). Banking Panics and Business Cycles. *Oxford Economic Papers*, **40**, 751–781.

— (2010). *Slapped by the Invisible Hand: The Panic of 2007*. Oxford University Press.

Green, E. J. (2010). Bailouts. *Federal Reserve Bank of Richmond Economic Quarterly*, **1**, 11–32.

— and Lin, P. (2003). Implementing efficient allocations in a model of financial intermediation. *Journal of Economic Theory*, **109**, 1–23.

Hanson, S. G., Kashyap, A. K. and Stein, J. C. (2011). A Macroprudential Approach to Financial Regulation. *Journal of Economic Perspectives*, **25** (1), 3–28.

Hellmann, T. F., Murdock, K. C. and Stiglitz, J. E. (2000). Moral Hazard in Banking , and Prudential Liberalization , Regulation : Are Capital Requirements Enough ? *American Economic Review*, **90** (1), 147–165.

Hsieh, C. T. (2003). Do consumers react to anticipated income changes? Evidence from the Alaska permanent fund. *American Economic Review*, **93** (1), 397–405.

IMF (2013). *The Interaction of Monetary and Macroprudential Policies*. Tech. rep., International Monetary Fund.

Johnson, D. S., Parker, J. A. and Souleles, N. S. (2006). Household expenditure and the income tax rebates of 2001. *American Economic Review*, **96** (5), 1589–1610.

Kashyap, A. K. and Stein, J. C. (2004). Cyclical implications of the Basel II capital standards. *Economic Perspectives*, **28** (1), 18–32.

Keister, T. (2016). Bailouts and financial fragility. *Review of Economic Studies*, **83** (2), 704–736.

Keys, B. J., Piskorski, T., Seru, A. and Yao, V. (2014). Mortgage Rates, Household Balance Sheets, and the Real Economy. *NBER Working Paper*.

Kindleberger, C. P. (1978). *Manias, Panics, and Crashes: A History of Financial Crises*. Basic Books.

Lacker, J. M. (2008). Financial Stability and Central Banks. *European Economics and Financial Centre Seminar*.

Miron, J. A. (1986). American Economic Association Financial Panics, the Seasonality of the Nominal Interest Rate, and the Founding of the. *Source: The American Economic Review*, **76** (1), 125–140.

Parker, J., Souleles, N. S., Johnson, D. S. and McClelland, R. (2013). Consumer Spending and the Economic Stimulus Payments of 2008. *American Economic Review*, **103** (6), 2530–2553.

Parker, J. A. (2015). Why Don't Households Smooth Consumption? Evidence from a 25 Million Dollar Experiment. *Working paper*.

Peck, J., Shell, K., Journal, S., February, N., Peck, J. and Shell, K. (2003). Equilibrium Bank Runs. *Journal of Political Economy*, **111** (1), 103–123.

Repullo, R. and Suarez, J. (2013). The procyclical effects of bank capital regulation. *Review of Financial Studies*, **26** (2), 452–490.

Saunders, A. and Wilson, B. (1996). Contagious Bank Runs: Evidence from the 1929–1933 Period. *Journal of Financial Intermediation*, **5**, 409–423.

Schumacher, L. (2000). Bank runs and currency run in a system without a safety net: Argentina and the 'tequila" shock. *Journal of Monetary Economics*, **46**, 257–277.

Shleifer, A. and Vishny, R. (2011). Fire Sales in Finance and Macroeconomics. *Journal of Economic Perspectives*, **25** (1), 29–48.

Souleles, N. S. (1999). The Response of Household Comsumption to Income Tax Refunds. *American Economic Review*, **89** (4), 947–958.

Stiglitz, J. E. and Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. *The American Economic Review*, **71** (3), 393–410.

Sultanum, B. (2014). Optimal Diamond-Dybvig mechanism in large economies with aggregate uncertainty. *Journal of Economic Dynamics and Control*, **40**, 95–102.

Wallace, N. (1988). Another attempt to explain an illiquid banking system: The Diamond and Dybvig model with sequential service taken seriously. *Federal Reserve Bank of Minneapolis Quarterly Review*, **12** (Fall), 3–16.

— (1990). A banking model in which partial suspension is best. *Federal Reserve Bank of Minneapolis Quarterly Review*, **14** (Fall), 11–23.

Walther, A. (2016). Jointly Optimal Regulation of Bank Capital and Liquidity. *Journal of Money, Credit and Banking*, **48** (2-3), 415–448.

Zeldes, S. P. (1989). Consumption and Liquidity Constraints: An Empirical Investigation. *Journal of Political Economy*, **97** (21), 305–346.

# Appendix A

# Appendix to Chapter 1

## A.1 Changes in mortgage payments at reset

In this appendix, I derive the ratios of the post–reset and pre-reset mortgage payments for each amortization group. We begin with some notation. Let *pre* and *post* represent the time periods before and after reset, respectively. Let the monthly payment, $P$, and the rate on which the payment $P$ is based, $r$, be defined for *pre* and *post*. Define $\eta = \frac{r_{post}}{r_{pre}}$ as the ratio of the post-reset to the pre-reset rate. Let $B_0$ be the original mortgage balance and $B_{60}$ be the outstanding mortgage balance 60 months after origination (which is the reset date for a 5-year hybrid ARM).

### A.1.1 Zero-zero group

Payments for this group are the most straightforward, as they are just interest-only both before and after reset. This implies

$$
\begin{aligned}
P_{post} &= \frac{r_{post}}{12} B_{60} \\
P_{pre} &= \frac{r_{pre}}{12} B_0 \\
\implies \frac{P_{post}}{P_{pre}} &= \frac{r_{post}}{r_{pre}} = \eta
\end{aligned}
$$

where the last line uses the fact that $B_{60} = B_0$ when payments are interest-only before reset. For the zero-zero group, the ratio of the post-reset and pre-reset mortgage payments is simply the ratio the mortgage rates, $\eta$.

### A.1.2 Positive-positive group

The post-reset and pre-reset mortgage payments for this group are given by the annuity formulas for mortgage payments. First define $g(r, m)$ as follows

$$g(r,m) = 1 - \left(1 + \frac{r}{12}\right)^{-m} < 1$$

where $r, m > 0$. The payments are given below. The post-reset payment uses $r_{post}$, the outstanding mortgage balance after 5 years ($B_{60}$), and a term of 25 years (300 months) while the pre-reset payment uses $r_{pre}$, the original mortgage balance ($B_0$), and a term of 30 years (360 months).

$$\begin{aligned}
P_{post} &= \frac{\frac{r_{post}}{12}}{g(r_{post}, 300)} B_{60} \\
P_{pre} &= \frac{\frac{r_{pre}}{12}}{g(r_{pre}, 360)} B_0 \\
\implies \frac{P_{post}}{P_{pre}} &= \frac{\frac{r_{post}}{g(r_{post},300)}}{\frac{r_{pre}}{g(r_{pre},300)}} \left[\frac{B_{60}}{B_0} \frac{g(r_{pre}, 360)}{g(r_{pre}, 300)}\right] \\
&= \eta \frac{g(r_{pre}, 300)}{g(r_{post}, 300)} \left[\frac{B_{60}}{B_0} \frac{g(r_{pre}, 360)}{g(r_{pre}, 300)}\right]
\end{aligned}$$

I now show that the bracketed term in the last line equals 1. Positive amortization implies the following evolution of $B$ for $1 \leq n \leq 60$ (i.e. before reset):

$$\begin{aligned}
B_n &= B_{n-1} - \left(P_{pre} - \frac{r_{pre}}{12} B_{n-1}\right) \\
&= B_{n-1}\left(1 + \frac{r_{pre}}{12}\right) - P_{pre} \\
&= B_0 \left(1 + \frac{r_{pre}}{12}\right)^n - P_{pre} \sum_{i=0}^{n-1} \left(1 + \frac{r_{pre}}{12}\right)^i
\end{aligned}$$

where the last line iterates the previous line back to $B_0$. Since $P_{pre}$ is the 30-year annuity payment whose present value equals $B_0$, we have

$$B_0 = P_{pre} \sum_{i=1}^{360} \left(1 + \frac{r_{pre}}{12}\right)^{-i} = \frac{g(r_{pre}, 360)}{\frac{r_{pre}}{12}} P_{pre}$$

Substituting the first equality into the previous line gives

$$\begin{aligned} B_n &= P_{pre}\left(\sum_{i=1}^{360}\left(1+\frac{r_{pre}}{12}\right)^{n-i} - \sum_{i=0}^{n-1}\left(1+\frac{r_{pre}}{12}\right)^{i}\right) \\ &= P_{pre}\sum_{i=1}^{360-n}\left(1+\frac{r_{pre}}{12}\right)^{-i} \\ &= \frac{g(r_{pre}, 360-n)}{\frac{r_{pre}}{12}} P_{pre} \end{aligned}$$

where the last line comes from the fact that the previous line represents the present value of the annuity $P_{pre}$ with term $360-n$. This implies

$$\begin{aligned} \frac{B_{60}}{B_0} &= \frac{g(r_{pre}, 300)}{g(r_{pre}, 360)} \\ \implies 1 &= \frac{B_{60}}{B_0}\frac{g(r_{pre}, 360)}{g(r_{pre}, 300)} \end{aligned}$$

Going back to the beginning, we have

$$\frac{P_{post}}{P_{pre}} = \eta\frac{g(r_{pre}, 300)}{g(r_{post}, 300)} > \eta \iff r_{post} < r_{pre}$$

The equality indicates that the ratio of the post-reset to pre-reset payments for the positive-positive group equals $\eta$ times a multiplier, which is shown to be greater than 1 if interest rates fall (this is due to $\frac{\partial g}{\partial r} > 0$). In other words, the positive-positive group experience a smaller reduction in monthly payments at reset compared to the zero-zero group for the same reduction in interest rates.

### A.1.3 Zero-positive group

$$\begin{aligned} P_{post} &= \frac{\frac{r_{post}}{12}}{g(r_{post}, 300)} B_{60} \\ P_{pre} &= \frac{r_{pre}}{12} B_0 \\ \implies \frac{P_{post}}{P_{pre}} &= \eta\frac{1}{g(r_{post}, 300)} \\ &> \eta\frac{g(r_{pre}, 300)}{g(r_{post}, 300)} \end{aligned}$$

where the last line uses $g(r, m) < 1$. The last inequality shows that households in the zero-positive group experience a smaller payment reduction than the positive-positive group, and therefore the smallest payment reduction overall. This is because while these housholds' interest rate goes down at reset, they have to start paying a principal portion as well. Based on the last equality, whether the ratio of payments is greater or less than 1 will depend on the whether $\eta > g(r_{post}, 300)$.

# Appendix B

# Appendix to Chapter 2

## B.1 Proof of Proposition 1

**Proposition 1.** *The solution to the recapitalization problem described by (2.23) and (2.24) involves concentrating the portfolio into one asset and selling all other assets. The asset that is retained is the solution to*

$$\arg\min_{i} \frac{E\left(min\left[\frac{\eta_i}{\lambda_i}, 1-\theta r_i\right]\right)}{1-\theta r_i}$$

*If multiple assets solve this problem, banks still choose to retain a single asset as long as there is imperfect correlation across asset returns. However, the bank is indifferent between retaining any of the assets that solve the problem.*

*Proof.* The first step of the proof is establishing the following lemma.

**Lemma 1.** *Let $x, y \in \mathbb{R}_n^{++}$ and let $\alpha$ be in the n-dimensional unit simplex. Then*

$$\frac{x_j}{y_j} = \min\left\{\frac{x_i}{y_i}\right\} \implies \frac{x_j}{y_j} \leq \frac{\sum_{i=1}^{n} \alpha_i x_i}{\sum_{i=1}^{n} \alpha_i y_i}$$

*with strict inequality if $\alpha_j \neq 1$ and $\frac{x_j}{y_j}$ is a unique minimum.*

*Proof.* The proof is by induction. First, we prove the result for $n = 2$. Without loss of generality, suppose $\frac{x_1}{y_1} \leq \frac{x_2}{y_2}$. Based on the conditions of the Lemma, $\alpha_1 \in [0, 1]$. If $\alpha_1 = 1$, then the result is trivially true. Suppose $\alpha_1 \in (0,1)$. This implies

$$\frac{\alpha_1 x_1}{\alpha_1 y_1} \leq \frac{\alpha_2 x_2}{\alpha_2 y_2}$$

$$\implies \alpha_1 x_1 \alpha_2 y_2 \leq \alpha_1 y_1 \alpha_2 x_2$$

$$\implies \alpha_1 x_1 (\alpha_1 y_1 + \alpha_2 y_2) \leq \alpha_1 y_1 (\alpha_1 x_1 + \alpha_2 x_2)$$

$$\implies \frac{x_1}{y_1} \leq \frac{\alpha_1 x_1 + \alpha_2 x_2}{\alpha_1 y_1 + \alpha_2 y_2}$$

and the result holds. Note that if $\frac{x_1}{y_1}$ is the unique minimum ($\frac{x_1}{y_1} < \frac{x_2}{y_2}$), the result would hold with strict inequality.

Now suppose $\alpha_1 = 0$ (which implies $\alpha_2 = 1$). Then we have

$$\frac{x_1}{y_1} \leq \frac{x_2}{y_2} = \frac{\alpha_1 x_1 + \alpha_2 x_2}{\alpha_1 y_1 + \alpha_2 y_2}$$

and the result still holds, noting again that if $\frac{x_1}{y_1}$ is the unique minimum, the result would hold with strict inequality.

Now suppose that the result is true for $n$. With $x, y \in \mathbb{R}^{++}_{n+1}$, $\frac{x_j}{y_j} = \min\{\frac{x_i}{y_i}\}$, and $\alpha$ in the $(n+1)$-dimensional unit simplex, we must show that the result is true for $n+1$:

$$\frac{x_j}{y_j} \leq \frac{\sum_{i=1}^{n+1} \alpha_i x_i}{\sum_{i=1}^{n+1} \alpha_i y_i}$$

Again, the result is trivially true for $\alpha_j = 1$. Suppose that $\alpha_j \in [0,1)$. We start with

$$\frac{\sum_{i=1}^{n+1} \alpha_i x_i}{\sum_{i=1}^{n+1} \alpha_i y_i} = \frac{\sum_{i \neq j} \alpha_i x_i + \alpha_j x_j}{\sum_{i \neq j} \alpha_i y_i + \alpha_j y_j}$$

Given the assumptions that $\frac{x_j}{y_j} = \min \left\{ \frac{x_i}{y_i} \right\}$ and the result is true for $n$, we have

$$\frac{x_j}{y_j} = \min \left\{ \frac{x_i}{y_i} \right\} \leq \min_{i \neq j} \left\{ \frac{x_i}{y_i} \right\} \leq \frac{\sum_{i \neq j} \frac{\alpha_i}{1-\alpha_j} x_i}{\sum_{i \neq j} \frac{\alpha_i}{1-\alpha_j} y_i}$$

with the first inequality being strict if $\frac{x_j}{y_j}$ is the unique minimum. Since we established that the result is true for $n = 2$, we can use the outer terms of the above to show

$$\frac{x_j}{y_j} \leq \frac{(1-\alpha_j) \sum_{i \neq j} \frac{\alpha_i}{1-\alpha_j} x_i + \alpha_j x_j}{(1-\alpha_j) \sum_{i \neq j} \frac{\alpha_i}{1-\alpha_j} y_i + \alpha_j y_j} = \frac{\sum_{i=1}^{n+1} \alpha_i x_i}{\sum_{i=1}^{n+1} \alpha_i y_i}$$

with the inequality being strict if $\frac{x_j}{y_j}$ is the unique minimum. This completes the proof of the lemma. □

The proof continues as follows. When only asset $j$ is retained, $s_i \bar{\lambda} = w_i \lambda_i$ for all $i \neq j$. Using this, the objective function (2.23) when only asset $j$ is retained simplifies to

$$\frac{E\left( min\left[ \frac{\eta_j}{\lambda_j}, 1 - \theta r_j \right] \right)}{1 - \theta r_j} \tag{B.1}$$

Therefore, if the bank wants to minimize the objective while retaining only one asset, that asset must be contained in the set $S$ defined below:

$$S = \left\{ j \in \{1, 2, ..., n\} : j = \arg\min_i \frac{E\left( min\left[ \frac{\eta_i}{\lambda_i}, 1 - \theta r_i \right] \right)}{1 - \theta r_i} \right\} \tag{B.2}$$

The set $S$ is non-empty and may contain multiple elements. Let $k$ be any element in $S$. We will now show that the objective function when retaining only asset $k$ is strictly less than the objective when the bank holds any portfolio containing more than just one asset. Clearly, if the banks holds a portfolio with just one asset, the objective will either be the same (if this asset is in $S$) or higher (if this asset is not in $S$). In other words, the solution to the recapitalization problem involves retaining only asset $k$, where $k$ is any element of $S$.

Let $\{s_1, s_2, ..., s_n\}$ be a decision of the bank in which more than one asset is retained. Denote the bank's portfolio weight on asset $j$ after all sales and purchases by $\alpha_j$, defined below:

$$\alpha_j = \frac{w_j\lambda_j - s_j\bar{\lambda}}{\sum_{i=1}^{n}(w_i\lambda_i - s_i\bar{\lambda})} \tag{B.3}$$

where $\alpha$ is in the n-dimensional unit simplex because of (2.24). With this terminology, the condition that more than one asset is retained in the portfolio is equivalent to $\alpha_i \in (0,1)$ for at least two assets.

Now, we divide the numerator and denominator of bank's objective (2.23) by $\sum_{i=1}^{n}(w_i\lambda_i - s_i\bar{\lambda})$ to get

$$g(s_1, s_2, ..., s_n) = \frac{E\left(min\left[\sum_{i=1}^{n}\alpha_i\frac{\eta_i}{\lambda_i},\ \sum_{i=1}^{n}\alpha_i(1-\theta r_i)\right]\right)}{\sum_{i=1}^{n}\alpha_i(1-\theta r_i)} \tag{B.4}$$

Our goal is to show that if $\alpha_i \in (0,1)$ for at least two assets, then (B.1) for asset $k$, where $k \in S$ defined in (B.2), is strictly less than (B.4) . Since the function $min(x,y)$ is concave and $\alpha$ is in the n-dimensional unit simplex, Jensen's inequality implies

$$min\left[\sum_{i=1}^{n}\alpha_i\frac{\eta_i}{\lambda_i},\ \sum_{i=1}^{n}\alpha_i(1-\theta r_i)\right] \geq \sum_{i=1}^{n}\alpha_i min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right] \tag{B.5}$$

$$\implies \frac{min\left[\sum_{i=1}^{n}\alpha_i\frac{\eta_i}{\lambda_i},\ \sum_{i=1}^{n}\alpha_i(1-\theta r_i)\right]}{\sum_{i=1}^{n}\alpha_i(1-\theta r_i)} \geq \frac{\sum_{i=1}^{n}\alpha_i min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right]}{\sum_{i=1}^{n}\alpha_i(1-\theta r_i)} \tag{B.6}$$

$$\implies g(s_1, s_2, ..., s_n) \geq \frac{\sum_{i=1}^{n}\alpha_i E\left(min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right]\right)}{\sum_{i=1}^{n}\alpha_i(1-\theta r_i)} \tag{B.7}$$

Note that (B.6) is true because $\theta r_i < 1\ \forall i$ and $\alpha_i > 0$ for at least two assets. The expression in (B.7) takes expectations of both sides of (B.6) and uses (B.4).

In (B.5), the inequality is generally weak but it will be useful to understand when it would be strong. Define the sets $D_1(\eta, \alpha)$ and $D_2(\eta, \alpha)$ as follows:

$$\begin{aligned} D_1(\eta, \alpha) &= \left\{ i \in \{1, , ..., n\} : \frac{\eta_i}{\lambda_i} < 1 - \theta r_i \text{ and } \alpha_i \in (0,1) \right\} \\ D_2(\eta, \alpha) &= \left\{ i \in \{1, , ..., n\} : \frac{\eta_i}{\lambda_i} > 1 - \theta r_i \text{ and } \alpha_i \in (0,1) \right\} \end{aligned} \tag{B.8}$$

The set $D_1$ is the set of assets that have positive weights in portfolio $\alpha$ and for a particular realization of $\eta$, a well-capitalized bank in period 1 holding only that asset defaults in period 2. This is because as can be seen from (B.1), the ratio of assets to debt in period 2 for a well-capitalized bank (capital ratio equals $\theta$) in period 1 holding only asset $x$ is $\frac{1}{1-\theta r_x} \frac{\eta_x}{\lambda_x}$. Similarly, the set $D_2$ is the set of assets that have positive weights in portfolio $\alpha$ and for a particular realization of $\eta$, a well-capitalized bank in period 1 holding only that asset does not default in period 2.

Based on the nature of the min function in (B.5), it is clear that the inequality in (B.5) would be strong if both $D_1$ and $D_2$ were non-empty for all possible $\alpha$'s. This would be true if, for every possible portfolio $\alpha$ and for a particular realization of $\eta$, there exists two particular assets with positive weight. For one asset, a well-capitalized bank holding just that asset in period 1 defaults in period 2. For the other asset, a well-capitalized bank holding just that asset does not default in period 2.

This may not be true for a particular realization of $\eta$. However, in (B.7), expectations of (B.6) are taken over all possible realizations of $\eta$. Therefore, the inequality in (B.7) would be strict if, for all possible portfolios $\alpha$, there exists a realization of $\eta$ such that the sets $D_1$ and $D_2$ are both non-empty. This condition can be stated more concisely as $D_3 \neq \emptyset$, where $D_3$ is defined below in (B.9).

$$D_3 = \{\{\eta\} : D_1(\eta, \alpha) \neq \emptyset \text{ and } D_2(\eta, \alpha) \neq \emptyset \; \forall \alpha\} \tag{B.9}$$

The condition $D_3 \neq \emptyset$ can be thought of as imposing imperfect correlation across asset returns. The condition is fairly weak and will be useful in establishing the uniqueness of the solution in certain corner cases (see the end of the proof).

Now, define the following:

$$x_i = E\left(min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right]\right)$$
$$y_i = 1-\theta r_i$$

which imply the following:

$$\frac{x_i}{y_i} = \frac{E\left(min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right]\right)}{1-\theta r_i}$$
$$\frac{\sum_{i=1}^{n}\alpha_i x_i}{\sum_{i=1}^{n}\alpha_i y_i} = \frac{\sum_{i=1}^{n}\alpha_i E\left(min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right]\right)}{\sum_{i=1}^{n}\alpha_i(1-\theta r_i)} \tag{B.10}$$

Note that $\frac{x_i}{y_i}$ takes the same form as (B.1), the objective function when retaining only one asset. In addition, $x_i, y_i \in \mathbb{R}_n^{++}$ and $\alpha$ defined in (B.3) is in the n-dimensional unit simplex. Furthermore, according to (B.2), we assumed that among the options of retaining each asset by itself, retaining asset $k \in S$ minimizes the bank's objective function: $\frac{x_k}{y_k} = \min\left\{\frac{x_i}{y_i}\right\}$. With all of these conditions, we can demonstrate the following.

$$\frac{E\left(min\left[\frac{\eta_k}{\lambda_k},\ 1-\theta r_k\right]\right)}{1-\theta r_k} = \frac{x_k}{y_k} \leq \frac{\sum_{i=1}^{n}\alpha_i x_i}{\sum_{i=1}^{n}\alpha_i y_i} = \frac{\sum_{i=1}^{n}\alpha_i E\left(min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right]\right)}{\sum_{i=1}^{n}\alpha_i(1-\theta r_i)} \leq g(s_1, s_2, ..., s_n)$$

where the equalities are from (B.10), the first inequality is from Lemma 1, and the second inequality is from (B.7).

Note that $\alpha_k \neq 1$ because at least two assets in the portfolio have positive weight. Therefore, if $S$ contains only one element or equivalently, if the solution to (B.2) is unique, the first inequality would be strict based on Lemma 1 and the proof is complete. The intuition here is that if retaining asset $k$ by itself is strictly preferred to retaining any other asset by itself, retaining asset $k$ must be superior to any portfolio of two or more assets, since the portfolio can be thought of as a weighted average of retaining the constituent assets by themselves. This logic holds even if the assets are perfectly correlated.

If the set $S$ has multiple elements, the first inequality is not strict. In this case, the bank is indifferent between retaining different individual assets by themselves. However, if the set $D_3$ as defined in (B.9) is non-empty, the second inequality is strict and the proof is complete. This condition can be thought of as imposing imperfect correlation across the returns of the assets contained in $S$. The intuition is that imperfect correlation creates diversification and lowers risk. As a result, it is undesirable for the bank to hold two assets at once, even if they are both contained in $S$.

The only case in which the proof breaks down is if $S$ contains multiple elements and the assets in $S$ are almost perfectly correlated (i.e. $D_3$ as defined in (B.9) is empty), a situation that would only occur if two assets were close to identical. Even in this case, the solution prescribed in Proposition 1 is valid, just not necessarily unique.

□

## B.2 Proof of Proposition 2

**Proposition 2.** *For each asset $k$, there exists $\bar{r}_k < \frac{1}{\theta}$ such that for all $r_k > \bar{r}_k$, asset $k$ is sold in the process of recapitalizing, holding fixed all other assets' risk weights.*

*Proof.* Expression (2.26) is the value of the bank's objective if it chooses to retain asset $k$. Note that (2.26) only depends on $r_k$, not any of the other assets' risk weights. Also, note that (2.26) converges to 1 as $r_k$ converges to $\frac{1}{\theta}$ from below:

$$\lim_{r_k \to \frac{1}{\theta}^-} \frac{E\left(min\left[\frac{\eta_k}{\lambda_k},\, 1-\theta r_k\right]\right)}{1-\theta r_k} = 1$$

Note also that since the objective is the price of debt per unit face value, its value must be less than one. Therefore, if we can show that (2.26) is strictly increasing in $r_k$, there exists some $\bar{r}_k < \frac{1}{\theta}$ above which (2.26) will exceed the value of the objective when retaining any other asset. For any $r_k \in (\bar{r}_k, \frac{1}{\theta})$, asset $k$ will be sold and a different asset will be retained.

We begin by rewriting (2.26) as follows

$$g(\cdot) = \frac{1}{1-\theta r_k} \int_{-\infty}^{\infty} min\left[\frac{\eta_k}{\lambda_k},\ 1-\theta r_k\right] f_k(\eta_k) d\eta_k \tag{B.11}$$
$$= 1 - F_k(\lambda_k(1-\theta r_k)) + \frac{1}{\lambda_k(1-\theta r_k)} \left( \int_{-\infty}^{\lambda_k(1-\theta r_k)} \eta_k f_k(\eta_k) d\eta_k \right)$$

then differentiating with respect to $r_k$

$$\frac{\partial g(\cdot)}{\partial r_k} = f_k(\lambda_k(1-\theta r_k))\theta\lambda_k + \frac{1}{(\lambda_k(1-\theta r_k))^2} \Bigg( \lambda_k(1-\theta r_k)(-\theta\lambda_k)\lambda_k(1-\theta r_k) f_k(\lambda_k(1-\theta r_k))$$
$$+ \theta\lambda_k \int_{-\infty}^{\lambda_k(1-\theta r_k)} \eta_k f_k(\eta_k) d\eta_k \Bigg)$$
$$= \frac{\theta\lambda_k}{(\lambda_k(1-\theta r_k))^2} \left( \int_{0}^{\lambda_k(1-\theta r_k)} \eta_k f_k(\eta_k) d\eta_k \right) > 0$$

because $\theta$, $\lambda_k$, and $1-\theta r_k$ are all strictly greater than zero. This completes the proof.

□

## B.3 Proof of Proposition 3

**Proposition 3.** *Suppose the risk weights of all assets are identically equal to one. If asset j's net return* $(\frac{\eta_j}{\lambda_j} - 1)$ *is a mean-preserving spread of the return of asset k (i.e. asset j is riskier), then the bank will not sell asset j in the process of recapitalizing.*

*Proof.* Suppose that the return of asset $j$ ($R_j = \frac{\eta_j}{\lambda_j} - 1$) is a mean-preserving spread of the return of asset $k$ ($R_k = \frac{\eta_k}{\lambda_k} - 1$). According to (2.26) with $r_j = r_k = 1$, asset $k$ is sold in the process of recapitalizing if

$$\frac{E\left(min\left[\frac{\eta_k}{\lambda_k},\ 1-\theta\right]\right)}{1-\theta} \geq \frac{E\left(min\left[\frac{\eta_j}{\lambda_j},\ 1-\theta\right]\right)}{1-\theta}$$
$$\iff E\left(min\left[R_k,\ -\theta\right]\right) \geq E\left(min\left[R_j,\ -\theta\right]\right)$$
$$\iff E\left(h(R_k)\right) \geq E\left(h(R_j)\right) \tag{B.12}$$

where the function $h(\cdot)$ is defined as $h(x) = \min[x, -\theta]$.

Since $R_j$ is a mean-preserving spread of $R_k$, it must be the case that $E\left(u(R_k)\right) \geq E\left(u(R_j)\right)$ for any concave, non-decreasing function $u(\cdot)$. Since $h(\cdot)$ in (B.12) is concave and non-decreasing, the proof is complete. □

## B.4 Proof of Proposition 4

**Proposition 4.** *With constraint (2.27) in place of (2.24) to reflect liquidity requirements, the solution to the bank's recapitalization problem is an extreme point of the feasible set: sell all assets but one and expand holdings of the remaining asset, both to the maximum extent allowable under (2.27).*

*Proof.* The proof has three steps. The first step is showing that any feasible solution can be represented as a convex combination of the extreme points of the feasible set.[1] The second step is showing that the objective function is quasiconcave.[2] Combining the first two steps, the third step is showing that the value of the objective at any non-extreme feasible solution must exceed the value of the objective at some extreme point of the feasible set. Therefore, the objective must achieve its minimum at an extreme point.

Let the feasible set for $s = \{s_1, s_2, ..., s_n\}$ be given by $\Omega_s$. The set $\Omega_s$ is defined by (2.27) and is clearly closed, bounded, and therefore compact. In addition, $\Omega_s$ is an n-dimensional box and is therefore convex.

The objective function can be restated in terms of the vector $\alpha$ in the n-dimensional unit simplex, as in (B.4), with $\alpha$ defined in (B.3). Furthermore, based on (B.3), the function that maps $s$ to $\alpha$ is a linear fractional function of the form

$$\alpha = \frac{As + b}{c's + d}$$

where $A = -\bar{\lambda} I_n$, $b_i = w_i \lambda_i$, $c_i = -\bar{\lambda}$, and $d = \bar{\lambda}$. Since linear fractional functions preserve

[1] An extreme point of a set $S$ is any point in $S$ which does not lie in any open line segment joining two points of $S$.

[2] A function $f(x)$ is quasiconcave if and only if for every $t \in [0, 1]$ and any $x_1, x_2$ in the domain of $f(\cdot)$, $f(tx_1 + (1 - t)x_2) \geq \min\{f(x_1), f(x_2)\}$.

convexity of sets, it follows that the feasible set for $\alpha$, $\Omega_\alpha$, is also convex. Moreover, the function that maps $s$ to $\alpha$ is also continuous, which implies that $\Omega_\alpha$ is also compact.

The Krein-Millman theorem states that if $\Omega_\alpha$ is convex and compact, $\Omega_\alpha$ is the convex hull of its extreme points, given by the set $e(\Omega_\alpha)$. By the definition of a convex hull, this implies that for every non-extreme feasible solution $\alpha^f$ to the problem in $\Omega_\alpha$, there exists $t \in (0,1)$ and $\alpha^1, \alpha^2 \in e(\Omega_\alpha)$ such that $\alpha^f = t\alpha^1 + (1-t)\alpha^2$. Note that the definition of extreme points implies that the set $e(\Omega_\alpha)$ contains all the portfolios that result from accumulating one asset while selling all other assets, both to the maximum extent possible under (2.27).

Below, we show that the objective function (B.4) evaluated at any non-extreme feasible solution $\alpha^f$ is weakly greater than the minimum of the objective evaluated at all of the extreme points of $\Omega_\alpha$.

$$\begin{aligned}
g(\alpha^f) &= g(t\alpha^1 + (1-t)\alpha^2) \\
&= \frac{E\left(min\left[\sum_{i=1}^{n}(t\alpha_i^1 + (1-t)\alpha_i^2)\frac{\eta_i}{\lambda_i},\ \sum_{i=1}^{n}(t\alpha_i^1 + (1-t)\alpha_i^2)(1-\theta r_i)\right]\right)}{\sum_{i=1}^{n}(t\alpha_i^1 + (1-t)\alpha_i^2)(1-\theta r_i)} \\
&\geq \frac{E\left(\sum_{i=1}^{n}(t\alpha_i^1 + (1-t)\alpha_i^2)min\left[\frac{\eta_i}{\lambda_i},\ 1-\theta r_i\right]\right)}{t\sum_{i=1}^{n}\alpha_i^1(1-\theta r_i) + (1-t)\sum_{i=1}^{n}\alpha_i^2(1-\theta r_i)} \qquad \text{(B.13)} \\
&= \frac{tE\left(\sum_{i=1}^{n}min\left[\alpha_i^1\frac{\eta_i}{\lambda_i},\ \alpha_i^1(1-\theta r_i)\right]\right) + (1-t)E\left(\sum_{i=1}^{n}min\left[\alpha_i^2\frac{\eta_i}{\lambda_i},\ \alpha_i^2(1-\theta r_i)\right]\right)}{t\sum_{i=1}^{n}\alpha_i^1(1-\theta r_i) + (1-t)\sum_{i=1}^{n}\alpha_i^2(1-\theta r_i)} \qquad \text{(B.14)}
\end{aligned}$$

where (B.13) uses the fact that the function $min(x,y)$ is concave and $t\alpha_i^1 + (1-t)\alpha_i^2$ is in the n-dimensional unit simplex to apply Jensen's inequality. Note that for the same reason discussed at length in Appendix B.1 (see the discussion of equations (B.8) and (B.9)), the inequality in (B.13) would be strict under fairly weak conditions that amount to imperfect correlation across assets.

Let $g_n(\cdot)$ and $g_d(\cdot)$ represent the numerator and denominator, respectively, of the objective (B.4). Then, continuing from (B.14) we have

$$g(a^f) \geq \frac{tg_n(\alpha^1) + (1-t)g_n(\alpha^2)}{tg_d(\alpha^1) + (1-t)g_d(\alpha^2)}$$

Since $t \in (0,1)$, we can use Lemma 1 to show

$$g(a^f) \geq \min\{g(\alpha^1), g(\alpha^2)\} \geq \min_{\alpha \in e(\Omega_\alpha)} \{g(\alpha)\}$$

The first inequality is equivalent to the objective function $g(\cdot)$ being quasiconcave. The second inequality is strict if one of the extreme points uniquely minimizes the objective among all extreme points.

This completes the proof. To summarize, we have shown that the objective function evaluated at any non-extreme, feasible solution is weakly greater than the minimum objective across all extreme points of the feasible set. This means that at least one of the extreme points minimizes the objective. Moreover, if assets are imperfectly correlated or one of the extreme points uniquely minimizes the objective among all extreme points, no non-extreme point minimizes the objective. Finally, if multiple extreme points minimize the objective, they all represent valid solutions.

□

# Appendix C

# Appendix to Chapter 3

In appendix C.1, we provide an explicit derivation of the allocation of resources that results from the best responses of banks and the policy maker to an arbitrary profile of depositors' withdrawal strategies under each of the two policy regimes. In appendix C.2, we use this derivation to provide proofs of selected propositions from the paper.

## C.1 Best-Response Allocations

For any profile of withdrawal strategies $y \in Y$, let $\hat{\pi}_s(y)$ denote the fraction of the remaining depositors who are impatient after $\pi_L$ withdrawals have been made. Since we focus on equilibria in which there is no panic when the fundamental state is $L$, the first $\pi_L$ withdrawals in that state represent all of the impatient depositors and we have

$$\hat{\pi}_L(y) = 0. \tag{C.1}$$

When the fundamental state is $H$, the first $\pi_L$ withdrawals may be a mix of patient and impatient investors. Using the assumption that $\beta$ is the "bad" sunspot state, as introduced in 3.4, we have

$$\frac{\pi_H - \pi_L}{1 - \pi_L} \leq \hat{\pi}_{H_\alpha}(y) \leq \hat{\pi}_{H_\beta}(y) \leq \pi_H \tag{C.2}$$

for all $y \in Y$. Given the values of $\hat{\pi}_s$ associated with a particular profile $y$, we can derive the best responses of banks and the policy maker under each regime as follows.

### C.1.1 No intervention

Under the policy regime with no intervention, the best-responses of banks and the policy maker are characterized by equations 3.8, 3.10, 3.11, 3.15, 3.16, and the resource constraint in each state. It can be shown that these same conditions also characterize the solution to the problem of choosing an allocation vector **c** to maximize 3.2 subject to the basic resource constraints

$$\pi_L\left(\sigma\tilde{c}_1+(1-\sigma)\,c_1\right)+(1-\pi_L)\left(\hat{\pi}_s(y)\,c_{1s}+(1-\hat{\pi}_s(y))\,\frac{c_{2s}}{R}\right)+g\leq 1$$

for all $s\in S$. Using the functional form 3.1, the solution to this problem can be shown to be

$$\tilde{c}_1^{NI}(y) = c_1^{NI}(y)=\frac{1}{\pi_L+\delta^{\frac{1}{\gamma}}+\bar{x}(y)^{\frac{1}{\gamma}}}, \tag{C.3}$$

$$c_{1s}^{NI}(y) = \left(\frac{\bar{x}(y)}{x_s(y)}\right)^{\frac{1}{\gamma}}c_1^{NI}(y)\quad\text{and}\quad c_{2s}^{NI}(y)=R^{\frac{1}{\gamma}}c_{1s}^{NI}(y)\quad\text{for all } s, \tag{C.4}$$

$$g(y) = \delta^{\frac{1}{\gamma}}c_1^{NI}(y), \tag{C.5}$$

where

$$x_s(y) \equiv \left((1-\pi_L)\left(\hat{\pi}_s(y)+(1-\hat{\pi}_s(y))\,R^{\frac{1-\gamma}{\gamma}}\right)\right)^{\gamma}\quad\text{and} \tag{C.6}$$

$$\bar{x}(y) \equiv \sum_{s\in S}q_s x_s(y). \tag{C.7}$$

This solution depends on depositors' withdrawal strategies $y$ only through the values of $\hat{\pi}_s(y)$.

### C.1.2 With intervention

Under the policy regime with intervention, the best-responses of banks and the policy maker are characterized by equations 3.10, 3.14, 3.17, and 3.20, together with the resource constraint in each state,

$$\pi_L\left(\sigma\tilde{c}_1+(1-\sigma)\,c_1\right)+(1-\pi_L)\left(\hat{\pi}_s(y)\,c_{1s}+(1-\hat{\pi}_s(y))\,\frac{c_{2s}}{R}\right)+g_s\leq 1.$$

Again using the functional form 3.1, the unique solution to these equations can be shown to be

$$\tilde{c}_1^{I}(y) = \left(\sigma\pi_L+(1-\sigma)\,\pi_L\left(\frac{\overline{z}(y)}{q_L z_L(y)}\right)^{\frac{1}{\gamma}}+\overline{z}(y)^{\frac{1}{\gamma}}\right)^{-1}, \tag{C.8}$$

$$c_1^{I}(y) = \left(\frac{\overline{z}(y)}{q_L z_L(y)}\right)^{\frac{1}{\gamma}}\tilde{c}_1^{I}(y), \tag{C.9}$$

$$c_{1s}^{I}(y) = \left(\frac{\bar{z}(y)}{z_s(y)}\right)^{\frac{1}{\gamma}} \tilde{c}_1^{I}(y) \quad \text{and} \quad c_{2s}^{I}(y) = R^{\frac{1}{\gamma}} c_{1s}^{I}(y) \quad \text{for all } s, \tag{C.10}$$

$$g_s(y) = \left(\delta \frac{\bar{z}(y)}{z_s(y)}\right)^{\frac{1}{\gamma}} \tilde{c}_1^{I}(y), \tag{C.11}$$

where

$$z_s(y) \equiv \left((x_s(y))^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}\right)^{\gamma}, \tag{C.12}$$

$$\bar{z}(y) \equiv \sum_{s\in S} q_s z_s(y) \tag{C.13}$$

and $x_s(y)$ is defined in C.6.

## C.2 Proofs of Selected Propositions

### C.2.1 Proof of Proposition 5

**Proposition 5.** *Under the policy regime with no intervention, the economy is:*

$(a)$ *weakly fragile if and only if* $c_{2H_\alpha}^{NI}(y^E) \geq c_1^{NI}(y^E) \geq c_{2H_\beta}^{NI}(y^E)$,

$(b)$ *strongly fragile if and only if* $c_1^{NI}(y^E) > c_{2H_\alpha}^{NI}(y^E)$, *and*

$(c)$ *not fragile if and only if* $c_1^{NI}(y^E) < c_{2H_\beta}^{NI}(y^E)$.

*Proof.* For part $(a)$, recall that the economy is defined to be weakly fragile if there exists an equilibrium in which depositors follow the strategy profile $y^E$ in 3.5. Consider the decision problem of depositor $i$ if she expects all other depositors to follow this profile. Her best response clearly requires withdrawing at $t = 1$ when she is impatient. When she is patient, withdrawing at $t = 1$ in state $s$ is part of a best response if and only if $c_1^{NI}(y^E) \geq c_{2s}^{NI}(y^E)$ holds, while withdrawing at $t = 2$ is part of a best response if and only if this inequality is reversed. The definitions in C.6 and C.7, together with C.1 and the inequalities in C.2, imply that $\bar{x}(y) \geq x_L(y)$ holds for any $y \in Y$. Using C.4, we then have

$$c_1^{NI}(y) < c_{2L}^{NI}(y) \tag{C.14}$$

for any $y \in Y$ and, hence, the depositor will always choose to wait until $t = 2$ when patient and the fundamental state is $L$. Therefore, the strategy in profile $y^E$, where she withdraws early in state $H_\beta$ but not in $H_\alpha$, represents a best response if and only if the two inequalities in part $(a)$ of the proposition hold. In this case, there is an equilibrium in which all depositors follow $y^E$ and, hence, the economy is weakly fragile if and only if these inequalities hold.

Before moving to parts $(b)$ and $(c)$ of the proposition, we establish some useful inequalities. Note that the definitions of $y^E$ and $\hat{\pi}_s$ imply

$$\hat{\pi}_{H_\alpha}\left(y^E\right) = \frac{\pi_H - \pi_L}{1-\pi_L} \quad \text{and} \quad \hat{\pi}_{H_\beta}\left(y^E\right) = \pi_H. \tag{C.15}$$

The inequalities in C.2 thus imply that, of all strategy profiles in the set $Y$, $y^E$ has the minimum proportion of remaining investors who are impatient in state $H_\alpha$ and the maximum proportion in state $H_\beta$. Using the definition of $x_s(y)$ in C.6, it follows that for any $y \in Y$, we have

$$x_L(y) = x_L\left(y^E\right) < x_{H_\alpha}\left(y^E\right) \quad \text{and} \tag{C.16}$$

$$x_{H_\alpha}\left(y^E\right) \leq x_{H_\alpha}(y) \leq x_{H_\beta}(y) \leq x_{H_\beta}\left(y^E\right). \tag{C.17}$$

We can use the definition of $\bar{x}(y)$ in C.7 to write

$$\frac{\bar{x}(y)}{x_{H_\alpha}(y)} = q_L \frac{x_L(y)}{x_{H_\alpha}(y)} + q_{H_\alpha} + q_{H_\beta}\frac{x_{H_\beta}(y)}{x_{H_\alpha}(y)}$$

and

$$\frac{\bar{x}(y)}{x_{H_\beta}(y)} = q_L \frac{x_L(y)}{x_{H_\beta}(y)} + q_{H_\alpha}\frac{x_{H_\alpha}(y)}{x_{H_\beta}(y)} + q_{H_\beta}.$$

Using C.16 and C.17, it is then straightforward to show that for all $y \in Y$, we have

$$\frac{\bar{x}(y)}{x_{H_\alpha}(y)} \leq \frac{\bar{x}(y^E)}{x_{H_\alpha}(y^E)} \quad \text{and} \quad \frac{\bar{x}(y)}{x_{H_\beta}(y)} \geq \frac{\bar{x}(y^E)}{x_{H_\beta}(y^E)}.$$

In addition, the middle inequality in C.17 implies

$$\frac{\bar{x}(y)}{x_{H_\beta}(y)} \leq \frac{\bar{x}(y)}{x_{H_\alpha}(y)}$$

for any $y \in Y$. Combining the two previous lines, we have

$$\frac{\bar{x}(y^E)}{x_{H_\beta}(y^E)} \leq \frac{\bar{x}(y)}{x_{H_\beta}(y)} \leq \frac{\bar{x}(y)}{x_{H_\alpha}(y)} \leq \frac{\bar{x}(y^E)}{x_{H_\alpha}(y^E)} \tag{C.18}$$

for any $y \in Y$.

Now suppose the inequality in part $(b)$ of the proposition holds. Then the expression for the best-response allocation $\mathbf{c}^{NI}$ in C.4 implies

$$\frac{\bar{x}\left(y^E\right)}{x_{H_\alpha}\left(y^E\right)} < \frac{1}{R}.$$

Using the two right-most inequalities in C.18 together with C.4, it then follows that

$$c_1^{NI}\left(y\right) > c_{2s}^{NI}\left(y\right)$$

holds for $s = \left(H_\alpha, H_\beta\right)$ and for all $y \in Y$. In other words, if an investor's best response when all other investors are playing $y^E$ requires withdrawing early in state $H_\alpha$, then her best response to **any** strategy profile in $Y$ will involve withdrawing early whenever the fundamental state is $H$. As a result, $y^F$ is the only possible equilibrium strategy profile in $Y$. The fact that $y^F$ is indeed an equilibrium profile follows from these inequalities together with C.14. We have, therefore, shown that $y^F$ is the unique equilibrium strategy profile and the economy is strongly fragile.

For the converse, suppose the economy is strongly fragile. Then $y^E$ is not an equilibrium strategy profile and one of the two inequalities in part $(a)$ of the proposition must be violated. Using C.4, the fact that $y^F$ is an equilibrium strategy profile implies

$$\frac{\bar{x}\left(y^F\right)}{x_{H_\alpha}\left(y^F\right)} < \frac{1}{R}.$$

It is straightforward to show that $\bar{x}\left(y^F\right) > \bar{x}\left(y^E\right)$ and $x_{H_\alpha}\left(y^F\right) = x_{H_\beta}\left(y^E\right)$. Together with the previous line, these two conditions imply

$$\frac{\bar{x}\left(y^E\right)}{x_{H_\beta}\left(y^E\right)} < \frac{1}{R}.$$

Using C.4 again, we then have

$$c_1^{NI}\left(y^E\right) > c_{2H_\beta}^{NI}\left(y^E\right).$$

In other words, when the economy is strongly fragile, the second inequality in part $(a)$ of the proposition is satisfied. The first inequality on that line must, therefore, be violated, which establishes that the inequality in part $(b)$ of the proposition holds.

Now suppose the inequality in part $(c)$ of the proposition holds. Again using the expression for the best-response allocation $\mathbf{c}^{NI}$ in C.4, this inequality implies

$$\frac{\bar{x}\left(y^E\right)}{x_{H_\beta}\left(y^E\right)} > \frac{1}{R}.$$

Using C.4 and the two left-most inequalities in C.18, together with C.14, it follows that

$$c_1^{NI}\left(y\right) < c_{2s}^{NI}\left(y\right)$$

for all $s$ and for all $y \in Y$. In other words, if an investor's best response when all other investors are playing $y^E$ involves waiting until $t = 2$ in state $H_\beta$ if she is patient, then her best response to **any** profile in $Y$ will be to wait until $t = 2$ in all states if she is patient. We have, therefore, established that $y^N$, defined in 3.7, is the unique equilibrium strategy profile and the economy is not fragile.

Finally, for the converse, note that if $y^N$ is the unique equilibrium strategy profile, it follows immediately from parts $(a)$ and $(b)$ of the proposition that the inequality in part $(c)$ must hold.

□

### C.2.2 Proof of Proposition 7

**Proposition 7.** *Under the policy regime with intervention, the economy is:*

$(a)$ *weakly fragile if and only if* $u\left(c_{2H_\alpha}^I\left(y^E\right)\right) \geq \mathcal{E}\left(\mathbf{c}^I\left(y^E\right)\right) \geq u\left(c_{2H_\beta}^I\left(y^E\right)\right)$,

$(b)$ *strongly fragile if and only if* $\mathcal{E}\left(\mathbf{c}^I\left(y^E\right)\right) > u\left(c_{2H_\alpha}^I\left(y^E\right)\right)$, *and*

$(c)$ *not fragile if and only if* $\mathcal{E}\left(\mathbf{c}^I\left(y^E\right)\right) < u\left(c_{2H_\beta}^I\left(y^E\right)\right)$.

*Proof.* The proof is broadly similar to that of Proposition 5, but with some important differences. For part $(a)$, consider the decision problem of depositor $i$ if she expects all other depositors to follow $y^E$ in 3.5. Her best response clearly requires withdrawing at $t = 1$ when she is impatient. When she is patient, withdrawing at $t = 1$ in state $s$ is part of a best response if and only if

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E\right)\right) \geq u\left(c_{2s}^I\left(y^E\right)\right)$$

holds, while withdrawing at $t = 2$ is part of a best response if and only if this inequality is reversed. Using 3.22 together with the definition 3.23 and the expressions for the best-response allocation $\mathbf{c}^I$ in C.9 - C.10, it is straightforward to show that

$$\mathcal{E}\left(\mathbf{c}^I(y)\right) \leq u\left(c_{2L}^I(y)\right) \tag{C.19}$$

holds for any $y \in Y$ and, hence, the depositor will always choose to wait until $t = 2$ when she is patient and the fundamental state is $L$. The strategy in profile $y^E$, under which she withdraws early in state $H_\beta$ but not in $H_\alpha$, then represents a best response if and only if the two inequalities in part $(a)$ of the proposition hold. In this case, there is an equilibrium in which investors follow $y^E$ and, hence, the economy is weakly fragile if and only if these two inequalities hold.

Next, note that the definition of $z_s(y)$ in C.12 combined with C.16 and C.17 implies that for any $y \in Y$, we have

$$z_L(y) = z_L\left(y^E\right) < z_{H_\alpha}\left(y^E\right) \quad \text{and} \tag{C.20}$$

$$z_{H_\alpha}\left(y^E\right) \leq z_{H_\alpha}(y) \leq z_{H_\beta}(y) \leq z_{H_\beta}\left(y^E\right). \tag{C.21}$$

In addition, using the same steps that led to C.18, we can show

$$\frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)} \leq \frac{\bar{z}(y)}{z_{H_\beta}(y)} \leq \frac{\bar{z}(y)}{z_{H_\alpha}(y)} \leq \frac{\bar{z}\left(y^E\right)}{z_{H_\alpha}\left(y^E\right)}. \tag{C.22}$$

Suppose the inequality in part $(b)$ of the proposition holds. Using 3.23 together with the expressions for the best-response allocation $\mathbf{c}^I$ in C.9 - C.10, and recalling that $\gamma > 1$, this inequality implies

$$\sigma + (1-\sigma)\left(\frac{\bar{z}\left(y^E\right)}{q_L z_L\left(y^E\right)}\right)^{\frac{1-\gamma}{\gamma}} < R^{\frac{1-\gamma}{\gamma}}\left(\frac{\bar{z}\left(y^E\right)}{z_{H_\alpha}\left(y^E\right)}\right)^{\frac{1-\gamma}{\gamma}}$$

or

$$\sigma\left(\frac{\bar{z}\left(y^E\right)}{z_{H_\alpha}\left(y^E\right)}\right)^{\frac{\gamma-1}{\gamma}} + (1-\sigma)\left(\frac{q_L z_L\left(y^E\right)}{z_{H_\alpha}\left(y^E\right)}\right)^{\frac{\gamma-1}{\gamma}} < R^{\frac{1-\gamma}{\gamma}}.$$

Combined with C.20 and C.21, we then have

$$\sigma\left(\frac{\bar{z}(y)}{z_s(y)}\right)^{\frac{\gamma-1}{\gamma}} + (1-\sigma)\left(\frac{q_L z_L(y)}{z_s(y)}\right)^{\frac{\gamma-1}{\gamma}} < R^{\frac{1-\gamma}{\gamma}}$$

for $s = \left(H_\alpha, H_\beta\right)$ and for all $y \in Y$. Again using 3.23 and C.9 - C.10, this inequality implies that

$$\mathcal{E}\left(\mathbf{c}^I(y)\right) > u\left(c_{2s}^I(y)\right)$$

holds for $s = \left(H_\alpha, H_\beta\right)$ and for all $y \in Y$. In other words, if an investor's best response when all other investors are playing $y^E$ requires withdrawing early in state $H_\alpha$, then her best response to **any** strategy profile in $Y$ will involve withdrawing early whenever the fundamental state is $H$. As a result, $y^F$ is the only possible equilibrium strategy profile in $Y$. The fact that $y^F$ is indeed an equilibrium profile follows from these inequalities together with C.19. We have, therefore, shown that $y^F$ is the unique equilibrium strategy profile and the economy is strongly fragile.

For the converse, suppose the economy is strongly fragile. Then $y^E$ is not an equilibrium strategy profile and one of the two inequalities in part $(a)$ of the proposition must be violated. Using C.9 - C.10, the fact that $y^F$ is an equilibrium strategy profile implies

$$\sigma\left(\frac{\bar{z}\left(y^F\right)}{z_{H_\alpha}\left(y^F\right)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(\frac{q_L z_L\left(y^F\right)}{z_{H_\alpha}\left(y^F\right)}\right)^{\frac{\gamma-1}{\gamma}}<R^{\frac{1-\gamma}{\gamma}}.$$

Using the definitions in C.12 and C.13, it is easy to show that $\bar{z}\left(y^F\right)>\bar{z}\left(y^E\right), z_L\left(y^F\right)=z_L\left(y^E\right)$ and $z_{H_\alpha}\left(y^F\right)=z_{H_\beta}\left(y^E\right)$. Together with the previous line, these three conditions imply

$$\sigma\left(\frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(\frac{q_L z_L\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}\right)^{\frac{\gamma-1}{\gamma}}<R^{\frac{1-\gamma}{\gamma}}.$$

Using C.9 and C.10 again, we then have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E\right)\right)>u\left(c^I_{2H_\beta}(y)\right).$$

In other words, when the economy is strongly fragile, the second inequality in part $(a)$ of the proposition is satisfied. The first inequality on that line must, therefore, be violated, which establishes that the inequality in part $(b)$ of the proposition holds.

Now suppose the inequality in part $(c)$ of the proposition holds. Again using 3.23 together with C.9 and C.10, this inequality implies

$$\sigma+(1-\sigma)\left(\frac{\bar{z}\left(y^E\right)}{q_L z_L\left(y^E\right)}\right)^{\frac{1-\gamma}{\gamma}}>R^{\frac{1-\gamma}{\gamma}}\left(\frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}\right)^{\frac{1-\gamma}{\gamma}}$$

or

$$\sigma\left(\frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(\frac{q_L z_L\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}\right)^{\frac{\gamma-1}{\gamma}}>R^{\frac{1-\gamma}{\gamma}}.$$

Using the inequalities in C.20 - C.22, we then have

$$\sigma\left(\frac{\bar{z}(y)}{z_s(y)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(\frac{q_L z_L(y)}{z_s(y)}\right)^{\frac{\gamma-1}{\gamma}}>R^{\frac{1-\gamma}{\gamma}}$$

for $s=(H_\alpha,H_\beta)$ and for all $y\in\Upsilon$. Using C.9 and C.10 together with the definition 3.23 and the inequality in C.19, we then have

$$\mathcal{E}\left(\mathbf{c}^I(y)\right)<u\left(c^I_{2s}(y)\right)$$

for all $s$ and for all $y\in\Upsilon$. In other words, if an investor's best response when all other investors are playing $y^E$ involves waiting until $t=2$ in state $H_\beta$ if she is patient, then her best response to **any** profile in $\Upsilon$ will be to wait until $t=2$ in all states if she is patient. This fact establishes that $y^N$ is the unique equilibrium strategy profile and the economy is not fragile.

Finally, for the converse, note that if $y^N$ is the unique equilibrium strategy profile, it follows immediately from parts $(a)$ and $(b)$ of the proposition that the inequality in part $(c)$ must hold.

□

### C.2.3 Proof of Proposition 8

**Proposition 8.** *Under the policy regime with intervention, the fragility type of an economy $(e,\sigma)$ is weakly decreasing in $\sigma$.*

*Proof.* To establish this result, we need to show that for any $e$ and any $\sigma'>\sigma$,

$(a)$ if $(e,\sigma)$ is not fragile, then $(e,\sigma')$ is not fragile, and

$(b)$ if $(e,\sigma)$ is weakly fragile, then $(e,\sigma')$ is either weakly fragile or not fragile.

For part $(a)$, if $(e,\sigma)$ is not fragile, then from Proposition 7 we have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;\sigma\right)\right)<u\left(c^I_{2H_\beta}\left(y^E;\sigma\right)\right).$$

Using the definition in 3.23 and the expressions for the best-response allocation $\mathbf{c}^I$ in C.9 and C.10, this inequality is equivalent to

$$\sigma+(1-\sigma)\left(\frac{\bar{z}\left(y^E\right)}{q_L z_L\left(y^E\right)}\right)^{\frac{1-\gamma}{\gamma}}>R^{\frac{1-\gamma}{\gamma}}\left(\frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}\right)^{\frac{1-\gamma}{\gamma}}.$$

The definitions of $z_s$ and $\bar{z}$ in C.12 and C.13 show that these terms are independent of $\sigma$. Moreover, $\bar{z}\left(y^E\right)>q_L z_L\left(y^E\right)$ and $\gamma>1$ imply that the left-hand side of this inequality is strictly increasing in $\sigma$. Therefore, for any $\sigma' > \sigma$ we have

$$\sigma' + (1-\sigma') \left( \frac{\bar{z}\left(y^E\right)}{q_L z_L\left(y^E\right)} \right)^{\frac{1-\gamma}{\gamma}} > R^{\frac{1-\gamma}{\gamma}} \left( \frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)} \right)^{\frac{1-\gamma}{\gamma}}.$$

Again using 3.23, C.9, and C.10, this inequality implies

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;\sigma'\right)\right) < u\left(c^I_{2H_\beta}\left(y^E;\sigma'\right)\right),$$

which establishes that economy $(e,\sigma')$ is not fragile as well.

The argument for part $(b)$ is similar. If $(e,\sigma)$ is weakly fragile, then we have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;\sigma\right)\right) \leq u\left(c^I_{2H_\alpha}\left(y^E;\sigma\right)\right).$$

Following the same steps used in part $(a)$ then shows that for any any $\sigma' > \sigma$, we have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;\sigma'\right)\right) < u\left(c^I_{2H_\alpha}\left(y^E;\sigma'\right)\right).$$

This inequality establishes that the economy $(e,\sigma')$ is not strongly fragile, implying that it is either weakly fragile or not fragile, as desired.

□

### C.2.4 Proof of Proposition 9

**Proposition 9.** *Under the policy regime with intervention, if 3.24 holds, then for any e there exists $\bar{\sigma} < 1$ such that the fragility type of all economies $(e,\sigma)$ with $\sigma > \bar{\sigma}$ is weakly decreasing in $\delta$.*

*Proof.* Let $e'$ denote a vector of parameters that differs from $e$ only in the parameter $\delta$, with $\delta' > \delta$. To establish the result, we need to show there exists $\bar{\sigma} < 1$ such that $\sigma > \bar{\sigma}$ implies

$(a)$ if $(e,\sigma)$ is not fragile, then $(e',\sigma)$ is not fragile, and

$(b)$ if $(e,\sigma)$ is weakly fragile, then $(e',\sigma)$ is either weakly fragile or not fragile.

The proof is divided into three steps.

**Step (i)** : Establish part $(a)$ . If $(e,\sigma)$ is not fragile, then from Proposition 7 we have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;e\right)\right) < u\left(c^I_{2H_\beta}\left(y^E;e\right)\right).$$

Using the definition in 3.23 and dividing both sides by $u\left(c^I_{2H_\beta}\right)$, this inequality can be written as[1]

$$\sigma\frac{u\left(\tilde{c}^I_1\left(y^E;e\right)\right)}{u\left(c^I_{2H_\beta}\left(y^E;e\right)\right)}+(1-\sigma)\frac{u\left(c^I_1\left(y^E;e\right)\right)}{u\left(c^I_{2H_\beta}\left(y^E;e\right)\right)}>1.$$

Using the expressions in C.9 and C.10, this inequality reduces to

$$\sigma\left(R\frac{\bar{z}\left(y^E;e\right)}{z_{H_\beta}\left(y^E;e\right)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(R\frac{q_L z_L\left(y^E;e\right)}{z_{H_\beta}\left(y^E;e\right)}\right)^{\frac{\gamma-1}{\gamma}}>1. \tag{C.23}$$

Using the definitions in C.12 and C.13, the ratio of $\bar{z}(y)$ to $z_{H_\beta}(y)$ for any $y$ can be written as

$$\frac{\bar{z}(y)}{z_{H_\beta}(y)}=q_L\left(\frac{x_L(y)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}{x_{H_\beta}(y)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}\right)^{\gamma}+q_{H_\alpha}\left(\frac{x_{H_\alpha}(y)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}{x_{H_\beta}(y)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}\right)^{\gamma}+q_{H_\beta}.$$

The definitions in C.6 show that $x_s(y)$ is independent of $\delta$ for all $s$. It is then straightforward to show that C.17 implies this expression is strictly increasing in $\delta$ for any $y$. Therefore, since $e'$ differs from $e$ only in that $\delta'>\delta$, we have

$$R\frac{\bar{z}\left(y^E;e'\right)}{z_{H_\beta}\left(y^E;e'\right)}>R\frac{\bar{z}\left(y^E;e\right)}{z_{H_\beta}\left(y^E;e\right)}. \tag{C.24}$$

The same steps can be used to show that the ratio of $z_L(y)$ to $z_{H_\beta}(y)$ is strictly increasing in $\delta$ for any $y$. Combining this fact with C.23 and C.24 yields

$$\sigma\left(R\frac{\bar{z}\left(y^E;e'\right)}{z_{H_\beta}\left(y^E;e'\right)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(R\frac{q_L z_L\left(y^E;e'\right)}{z_{H_\beta}\left(y^E;e'\right)}\right)^{\frac{\gamma-1}{\gamma}}>1.$$

Using C.9 and C.10 together with the definition in 3.23, we then have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;e'\right)\right)<u\left(c^I_{2H_\beta}\left(y^E;e'\right)\right),$$

which establishes that the economy $(e',\sigma)$ is also not fragile. Note that no restriction on $\sigma$ is required for this step of the proof.

**Step (ii)** : Establish a useful intermediate result: If 3.24 holds and the ratio $R\bar{z}\left(y^E\right)/z_{H_\alpha}\left(y^E\right)$ is greater than 1 for some value of $\delta$, then it is greater than 1 for all $\delta'>\delta$. We establish this result by showing that whenever this ratio is smaller than 1, the ratio is a strictly increasing function of

[1]Recall that $\gamma>1$ implies $u(c)$ is a negative number, which is why the inequality reverses direction in this step.

$\delta$. Since the ratio is a continuously differentiable function of $\delta$, its value can never cross 1 from above as $\delta$ is increased.

We begin this step by using the definitions in C.12 and C.13 to show that $R\bar{z}\left(y^E\right)/z_{H_\alpha}\left(y^E\right) < 1$ is equivalent to[2]

$$\frac{\bar{z}}{z_{H_\alpha}} = q_L\left(\frac{(x_L)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}{(x_{H_\alpha})^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}\right)^{\gamma} + q_{H_\alpha} + q_{H_\beta}\left(\frac{\left(x_{H_\beta}\right)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}{(x_{H_\alpha})^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}\right)^{\gamma} < \frac{1}{R}. \tag{C.25}$$

This expression is a differentiable function of $\delta$ for all $\delta > 0$ and we can write the derivative as

$$\frac{d}{d\delta}\left(\frac{\bar{z}}{z_{H_\alpha}}\right) = \delta^{\frac{1-\gamma}{\gamma}}\left(z_{H_\alpha}\right)^{-\frac{2}{\gamma}}\left(\begin{array}{c} q_L\left(\frac{z_L}{z_{H_\alpha}}\right)^{\frac{\gamma-1}{\gamma}}\left((x_{H_\alpha})^{\frac{1}{\gamma}} - (x_L)^{\frac{1}{\gamma}}\right) \\ -q_{H_\beta}\left(\frac{z_{H_\beta}}{z_{H_\alpha}}\right)^{\frac{\gamma-1}{\gamma}}\left(\left(x_{H_\beta}\right)^{\frac{1}{\gamma}} - (x_{H_\alpha})^{\frac{1}{\gamma}}\right)\end{array}\right). \tag{C.26}$$

In general, the sign of this expression can be either positive or negative. Our interest, however, is in signing the expression when condition C.25 holds. We can rewrite C.25 as

$$q_{H_\beta}\frac{z_{H_\beta}}{z_{H_\alpha}} < \frac{1}{R} - q_{H_\alpha} - q_L\frac{z_L}{z_{H_\alpha}}.$$

Combined with C.26, we then have

$$\frac{d}{d\delta}\left(\frac{\bar{z}}{z_{H_\alpha}}\right) > \delta^{\frac{1-\gamma}{\gamma}}\left(z_{H_\alpha}\right)^{-\frac{2}{\gamma}}\left(\begin{array}{c} q_L\left(\frac{z_L}{z_{H_\alpha}}\right)^{\frac{\gamma-1}{\gamma}}\left((x_{H_\alpha})^{\frac{1}{\gamma}} - (x_L)^{\frac{1}{\gamma}}\right) \\ -\left[\frac{1}{R} - q_{H_\alpha} - q_L\frac{z_L}{z_{H_\alpha}}\right]\left(\frac{z_{H_\beta}}{z_{H_\alpha}}\right)^{-\frac{1}{\gamma}}\left(\left(x_{H_\beta}\right)^{\frac{1}{\gamma}} - (x_{H_\alpha})^{\frac{1}{\gamma}}\right)\end{array}\right) \tag{C.27}$$

The inequalities in C.20 and C.21 imply

$$\left(\frac{z_L}{z_{H_\alpha}}\right)^{-\frac{1}{\gamma}} > \left(\frac{z_{H_\beta}}{z_{H_\alpha}}\right)^{-\frac{1}{\gamma}}.$$

Using this inequality to replace the penultimate term in C.27 and simplifying, we have

$$\frac{d}{d\delta}\left(\frac{\bar{z}}{z_{H_\alpha}}\right) > \delta^{\frac{1-\gamma}{\gamma}}\left(z_{H_\alpha}\right)^{-\frac{2}{\gamma}}\left(\begin{array}{c} q_L\left(\frac{z_L}{z_{H_\alpha}}\right)^{\frac{\gamma-1}{\gamma}}\left(\left(x_{H_\beta}\right)^{\frac{1}{\gamma}} - (x_L)^{\frac{1}{\gamma}}\right) \\ -\left[\frac{1}{R} - q_{H_\alpha}\right]\left(\frac{z_L}{z_{H_\alpha}}\right)^{-\frac{1}{\gamma}}\left(\left(x_{H_\beta}\right)^{\frac{1}{\gamma}} - (x_{H_\alpha})^{\frac{1}{\gamma}}\right)\end{array}\right). \tag{C.28}$$

Note that 3.22 implies $q_L \geq 1/R$ and, therefore, a sufficient condition for the derivative in C.28 to

[2]The terms $z_s$ and $x_s$ are all evaluated at the strategy profile $y^E$ throughout this step. We omit this dependence from the notation here to save space.

be positive is

$$\frac{1}{R}\left(\frac{z_L}{z_{H_\alpha}}\right)^{\frac{\gamma-1}{\gamma}}\left(\left(x_{H_\beta}\right)^{\frac{1}{\gamma}}-\left(x_L\right)^{\frac{1}{\gamma}}\right)>\left[\frac{1}{R}-q_{H_\alpha}\right]\left(\frac{z_L}{z_{H_\alpha}}\right)^{-\frac{1}{\gamma}}\left(\left(x_{H_\beta}\right)^{\frac{1}{\gamma}}-\left(x_{H_\alpha}\right)^{\frac{1}{\gamma}}\right).$$

Using the definitions C.6 and C.12 together with C.1 and C.15, it is straightforward to show that this inequality is equivalent to 3.24. In other words, as long as 3.24 holds, we have established that the ratio $R\bar{z}\left(y^E\right)/z_{H_\alpha}\left(y^E\right)$ strictly increasing in $\delta$ whenever the value of the ratio is less than 1. If the ratio is greater than 1 for some value of $\delta$, therefore, it must be greater than 1 for all $\delta' > \delta$ since continuity implies that it cannot cross 1 from above as $\delta$ is increased.

**Step (iii)** : Establish part $(b)$. If the economy $(e,\sigma)$ is weakly fragile, then from Proposition 7 we have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;e\right)\right)\leq u\left(c^I_{2H_\alpha}\left(y^E;e\right)\right).$$

Following the same approach as in step (i) above, this inequality can be written as

$$\sigma\left(R\frac{\bar{z}\left(y^E;e\right)}{z_{H_\alpha}\left(y^E;e\right)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(R\frac{q_L z_L\left(y^E;e\right)}{z_{H_\alpha}\left(y^E;e\right)}\right)^{\frac{\gamma-1}{\gamma}}\geq 1. \tag{C.29}$$

It follows immediately from the definition of $\bar{z}$ in C.13 that $\bar{z}(y) > q_L z_L(y)$ holds for any $y$. If the inequality in C.29 holds, therefore, it must be the case that

$$R\frac{\bar{z}\left(y^E;e\right)}{z_{H_\alpha}\left(y^E;e\right)}>1.$$

The result from step (ii) above together with $\delta' > \delta$ then implies

$$R\frac{\bar{z}\left(y^E;e'\right)}{z_{H_\alpha}\left(y^E;e'\right)}>1.$$

It follows from continuity that we can find $\bar{\sigma} < 1$ such that if $\sigma > \bar{\sigma}$, we have

$$\sigma\left(R\frac{\bar{z}\left(y^E;e'\right)}{z_{H_\alpha}\left(y^E;e'\right)}\right)^{\frac{\gamma-1}{\gamma}}+(1-\sigma)\left(R\frac{q_L z_L\left(y^E;e'\right)}{z_{H_\alpha}\left(y^E;e'\right)}\right)^{\frac{\gamma-1}{\gamma}}>1.$$

Using 3.23, C.9 and C.10, we then have

$$\mathcal{E}\left(\mathbf{c}^I\left(y^E;e'\right)\right)<u\left(c^I_{2H_\alpha}\left(y^E;e'\right)\right).$$

By Proposition 7, this inequality demonstrates that the economy $(e',\sigma)$ is not strongly fragile, implying that it is either weakly fragile or not fragile, as desired.

□

**Lemma 2.** *For any e with $\delta > 0$ and any $y \in Y$, there exists $\overline{\sigma} < 1$ such that $W^{I}(y) > W^{NI}(y)$ for all economies $(e, \sigma)$ with $\sigma > \overline{\sigma}$*

*Proof.* The proof of Lemma 2 is divided into four steps as follows.

**Step (i)** : Calculate the level of welfare associated with $y$ under policy regime $NI$. For this case, the value of the objective function 3.2 can be written as a function of the elements of the best-response allocation vector $\mathbf{c}^{NI}$ as follows:[3]

$$W^{NI}(y) = \pi_L u\left(c_1^{NI}\right) + (1-\pi_L)\, {}_s um_{s\in S} q_s \left(\hat{\pi}_s u\left(c_{1s}^{NI}\right) + (1-\hat{\pi}_s)\, u\left(c_{2s}^{NI}\right)\right) + v\left(g^{NI}\right).$$

Using the solutions in C.3 - C.5 and simplifying terms, this expression can be reduced to

$$W^{NI}(y) = \frac{1}{1-\gamma}\left(\pi_L + \delta^{\frac{1}{\gamma}} + \bar{x}^{\frac{1}{\gamma}}\right)^{\gamma}. \tag{C.30}$$

**Step (ii) :** Find a lower bound for the welfare level associated with $y$ under regime $I$. The value of the objective function 3.2 in this case can be written as

$$\begin{aligned} W^{I}(y) \quad = \quad & \sigma\pi_L u\left(\tilde{c}_1^{I}\right) + (1-\sigma)\,\pi_L u\left(c_1^{I}\right) + \\ & {}_s um_{s\in S} q_s \left((1-\pi_L)\left(\hat{\pi}_s u\left(c_{1s}^{I}\right) + (1-\hat{\pi}_s)\, u\left(c_{2s}^{I}\right)\right) + v\left(g_s^{I}\right)\right). \end{aligned}$$

Using the solutions in C.8 - C.11 and simplifying terms, this expression can be reduced to

$$W^{I}(y) = \frac{1}{1-\gamma}\frac{\sigma\pi_L + (1-\sigma)\,\pi_L\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1-\gamma}{\gamma}} + \bar{z}^{\frac{1}{\gamma}}}{\left(\sigma\pi_L + (1-\sigma)\,\pi_L\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1}{\gamma}} + \bar{z}^{\frac{1}{\gamma}}\right)^{1-\gamma}}. \tag{C.31}$$

The definition of $\bar{z}$ in C.13 shows that $\bar{z} > q_L z_L$ holds. Using this fact and $\gamma > 1$, we have

$$\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1-\gamma}{\gamma}} < \left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1}{\gamma}}$$

[3] Note that each element of $\mathbf{c}^{NI}$ depends on the profile of withdrawal strategies $y$, but the dependence is omitted in this expression to save space. The same is true for the terms $x_s$ and $z_s$ in the equations that follow.

and, therefore,

$$W^I(y) > \frac{1}{1-\gamma}\left(\sigma\pi_L + (1-\sigma)\,\pi_L\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1}{\gamma}} + \bar{z}^{\frac{1}{\gamma}}\right)^{\gamma}.$$

**Step (iii) :** Establish a useful intermediate result. Define

$$h(\delta) \equiv \bar{x}^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}} - \bar{z}^{\frac{1}{\gamma}}.$$

Using the definitions of $\bar{x}$ and $\bar{z}$ in C.7 and C.13, we then have

$$h(\delta) = \left({}_s um_{s\in S} q_s x_s\right)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}} - \left({}_s um_{s\in S} q_s \left((x_s)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}\right)^{\gamma}\right)^{\frac{1}{\gamma}}.$$

It is easy to see from this expression that $g(0) = 0$. Differentiating with respect to $\delta$ and simplifying terms yields

$$h'(\delta) = \frac{1}{\gamma}\delta^{\frac{1-\gamma}{\gamma}}\left[1 - \frac{{}_s um_{s\in S} q_s \left((x_s)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}\right)^{\gamma-1}}{\left({}_s um_{s\in S} q_s \left((x_s)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}\right)^{\gamma}\right)^{\frac{\gamma-1}{\gamma}}}\right]. \tag{C.32}$$

Note that for any distinct numbers $\{d_s\} > 0$ and $\varepsilon > 1$, Jensen's inequality implies

$${}_s um_{s\in S} q_s d_s < \left({}_s um_{s\in S} q_s d_s^{\varepsilon}\right)^{\frac{1}{\varepsilon}}.$$

Setting

$$d_s = \left((x_s)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}\right)^{\gamma-1} \qquad \text{and} \qquad \varepsilon = \frac{\gamma}{\gamma-1},$$

we then have

$${}_s um_{s\in S} q_s \left((x_s)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}\right)^{\gamma-1} < \left({}_s um_{s\in S} q_s \left((x_s)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}\right)^{\gamma}\right)^{\frac{\gamma-1}{\gamma}}$$

for all $\delta > 0$, which implies that the term in the square brackets in C.32 is strictly positive. In other words, we have established that function $h$ is strictly positive and strictly increasing for all $\delta > 0$.

**Step (iv) :** Find $\bar{\sigma}$ such that $\sigma > \bar{\sigma}$ implies welfare is necessarily higher with intervention. Using the expressions above, a sufficient condition for $W^I(y)$ to be larger than $W^{NI}(y)$ is

$$\frac{1}{1-\gamma}\left(\sigma\pi_L + (1-\sigma)\,\pi_L\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1}{\gamma}} + \bar{z}^{\frac{1}{\gamma}}\right)^{\gamma} > \frac{1}{1-\gamma}\left(\pi_L + \delta^{\frac{1}{\gamma}} + \bar{x}^{\frac{1}{\gamma}}\right)^{\gamma}$$

or

$$\sigma\pi_L + (1-\sigma)\,\pi_L\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1}{\gamma}} + \bar{z}^{\frac{1}{\gamma}} < \pi_L + \delta^{\frac{1}{\gamma}} + \bar{x}^{\frac{1}{\gamma}}$$

or

$$\left(1-\sigma\right)\pi_L\left(\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1}{\gamma}}-1\right)<\bar{x}^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}-\bar{z}^{\frac{1}{\gamma}}=h\left(\delta\right).$$

In step (iii) we showed that $h\left(\delta\right)>0$ for all $\delta>0$. The definitions of $\bar{x}$, $\bar{z}$, and $z_L$ in C.7, C.12, and C.13 show that each of these terms is independent of $\sigma$. Therefore, if we define

$$\bar{\sigma}\equiv 1-\left[\frac{\bar{x}^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}-\bar{z}^{\frac{1}{\gamma}}}{\pi_L\left(\left(\frac{\bar{z}}{q_L z_L}\right)^{\frac{1}{\gamma}}-1\right)}\right]<1,$$

then $\sigma>\bar{\sigma}$ implies welfare is strictly higher under the regime with intervention, as desired. □

**Lemma 3.** *Assume 3.24 holds. For any $e$ with $\delta>0$, there exists $\overline{\sigma}<1$ such that allowing intervention weakly reduces the fragility type of all economies $(e,\sigma)$ with $\sigma>\overline{\sigma}$*

*Proof.* To establish this result, we need to show that for any $e$ with $\delta>0$, there exists $\bar{\sigma}<1$ such that $\sigma>\bar{\sigma}$ implies

$(a)$ if $(e,\sigma)$ is not fragile under $NI$, it is not fragile under $I$, and

$(b)$ if $(e,\sigma)$ is weakly fragile under $NI$, it is either weakly fragile or not fragile under $I$.

For part $(a)$, if $(e,\sigma)$ is not fragile under regime $NI$, then by Proposition 1 we know $c_1^{NI}\left(y^E\right)<c_{2H_\beta}^{NI}\left(y^E\right)$ holds. Using C.4, we then have

$$\frac{\bar{x}\left(y^E\right)}{x_{H_\beta}\left(y^E\right)}=q_L\frac{x_L\left(y^E\right)}{x_{H_\beta}\left(y^E\right)}+q_{H_\alpha}\frac{x_{H_\alpha}\left(y^E\right)}{x_{H_\beta}\left(y^E\right)}+q_{H_\beta}>\frac{1}{R}. \tag{C.33}$$

Note that all of the terms in this expression are independent of the parameter $\delta$. Next, the ratio of $\bar{z}\left(y^E\right)$ to $z_{H_\beta}\left(y^E\right)$ can be written as

$$\frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}=q_L\left(\frac{x_L\left(y^E\right)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}{x_{H_\beta}\left(y^E\right)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}\right)^{\gamma}+q_{H_\alpha}\left(\frac{x_{H_\alpha}\left(y^E\right)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}{x_{H_\beta}\left(y^E\right)^{\frac{1}{\gamma}}+\delta^{\frac{1}{\gamma}}}\right)^{\gamma}+q_{H_\beta}. \tag{C.34}$$

This ratio is identical to the one in C.33 when $\delta$ is set to zero. Moreover, it is straightforward to show that C.17 implies that C.34 is strictly increasing in $\delta$. It follow that for any economy in which C.33 holds, we also have

$$\frac{\bar{z}\left(y^E\right)}{z_{H_\beta}\left(y^E\right)}>\frac{1}{R}.$$

Using C.10, we then have $\tilde{c}_1^I\left(y^E\right)<c_{2H_\beta}^I\left(y^E\right)$. Continuity and 3.23 then imply that we can find

$\bar{\sigma} < 1$ such that $\sigma > \bar{\sigma}$ implies

$$\mathcal{E}\left(\mathbf{c}^{I}\left(y^{E};\sigma\right)\right) < u\left(c^{I}_{2H_{\beta}}\left(y^{E};\sigma\right)\right),$$

which, by Proposition 2, shows that $(e, \sigma)$ is not fragile under policy regime $I$.

For part $(b)$, if the economy is weakly fragile under regime $NI$, then by Proposition 1 we know that $c_1^{NI}\left(y^E\right) \leq c_{2H_\alpha}^{NI}\left(y^E\right)$ holds. Using C.4, we then have

$$\frac{\bar{x}\left(y^{E}\right)}{x_{H_{\alpha}}\left(y^{E}\right)} = q_{L}\frac{x_{L}\left(y^{E}\right)}{x_{H_{\alpha}}\left(y^{E}\right)} + q_{H_{\alpha}} + q_{H_{\beta}}\frac{x_{H_{\beta}}\left(y^{E}\right)}{x_{H_{\alpha}}\left(y^{E}\right)} \geq \frac{1}{R}. \tag{C.35}$$

Note that, as with C.33, all of the terms in this expression are independent of the parameter $\delta$. Next, the ratio of $\bar{z}\left(y^E\right)$ to $z_{H_\alpha}\left(y^E\right)$ can be written as

$$\frac{\bar{z}\left(y^{E}\right)}{z_{H_{\alpha}}\left(y^{E}\right)} = q_{L}\left(\frac{x_{L}\left(y^{E}\right)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}{x_{H_{\alpha}}\left(y^{E}\right)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}\right)^{\gamma} + q_{H_{\alpha}} + q_{H_{\beta}}\left(\frac{x_{H_{\beta}}\left(y^{E}\right)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}{x_{H_{\alpha}}\left(y^{E}\right)^{\frac{1}{\gamma}} + \delta^{\frac{1}{\gamma}}}\right)^{\gamma}. \tag{C.36}$$

This ratio is identical to the one in C.35 when $\delta$ is set to zero. Moreover, step (ii) in the proof of Proposition 9 establishes that, assuming 3.24 holds, whenever the ratio $\bar{z}\left(y^E\right)/z_{H_\alpha}\left(y^E\right)$ is less than $1/R$, it is strictly increasing in $\delta$.[4] Starting from C.35, which is independent of $\delta$, and using the fact that this ratio is a continuously differentiable function of $\delta$, it follows that

$$\frac{\bar{z}\left(y^{E}\right)}{z_{H_{\alpha}}\left(y^{E}\right)} \geq \frac{1}{R} \tag{C.37}$$

holds for any $\delta > 0$. Therefore, any economy for which C.35 holds will also satisfy C.37 and, using C.10, will necessarily have $\tilde{c}_1^I\left(y^E\right) \leq c_{2H_\alpha}^I\left(y^E\right)$. Following the same logic as in step (i) above, continuity and 3.23 then imply that we can find $\bar{\sigma} < 1$ such that $\sigma > \bar{\sigma}$ implies

$$\mathcal{E}\left(\mathbf{c}^{I}\left(y^{E};\sigma\right)\right) \leq u\left(c^{I}_{2H_{\alpha}}\left(y^{E};\sigma\right)\right).$$

[4]The role of assumption 3.24 in the analysis can be seen by comparing equations C.34 and C.36. The expression in C.34 is always a strictly increasing function of $\delta$, which implies that when $\sigma$ is close to 1 and the incentive distortions associated with bailouts are small, having a larger public sector always reduces the incentive for depositors to run in state $H_\beta$. Working with the expression in C.36 shows that the same is **not** true in state $H_\alpha$. The larger bailout payments associated with a higher value of $\delta$ will lead the policy maker to be less conservative in setting the early payment $\tilde{c}_1$. In some cases, the ratio $\tilde{c}_1/c_{2H_\alpha}$ will actually *rise* when $\delta$ is increased, meaning that larger bailouts can *increase* the incentive for depositors to run in state $H_\alpha$. Step (ii) of the proof shows that this effect cannot arise when the economy is strongly fragile if 3.24 holds. In step (iii), we use this intermediate result to show that allowing intervention cannot move the economy from weakly fragile to strongly fragile when $\sigma$ is close to 1 and 3.24 holds, which allows us to establish the desirability of allowing intervention in such cases in Proposition 10 below.

By Proposition 2, therefore, the economy is not strongly fragile under regime $I$, implying that it is either weakly fragile or not fragile, which completes the proof.

□

### C.2.5 Proof of Proposition 10

**Proposition 10.** *Assume 3.24 holds. For any $e$ with $\delta > 0$, there exists $\overline{\sigma} < 1$ such that allowing intervention strictly increases equilibrium welfare for all economies $(e, \sigma)$ with $\sigma > \overline{\sigma}$.*

*Proof.* The proof of the proposition is divided into two steps.

**Step (i):** Show $W^P\left(y^N\right) > W^P\left(y^E\right) > W^P\left(y^F\right)$ *for* $P = NI, I$.

*Proof:* Using the definitions of these three strategy profiles in 3.5 – 3.7, together with the definition of $\hat{\pi}_s(y)$ as the fraction of the remaining depositors who are impatient after $\pi_L$ withdrawals have been made, we have

$$\begin{aligned} \hat{\pi}_{H_\alpha}\left(y^N\right) &= \hat{\pi}_{H_\alpha}\left(y^E\right) < \hat{\pi}_{H_\alpha}\left(y^F\right), \quad \text{and} \\ \hat{\pi}_{H_\beta}\left(y^N\right) &< \hat{\pi}_{H_\beta}\left(y^E\right) = \hat{\pi}_{H_\beta}\left(y^F\right). \end{aligned}$$

Using C.1 and the definition of $\bar{x}(y)$ in C.7, we then have

$$\bar{x}\left(y^N\right) < \bar{x}\left(y^E\right) < \bar{x}\left(y^F\right). \tag{C.38}$$

Equation C.30 shows that the level of welfare generated by the best-response allocation $\mathbf{c}^{NI}$ is a strictly decreasing function of $\bar{x}(y)$ (recall that $\gamma > 1$). Therefore, we have

$$W^{NI}\left(y^N\right) > W^{NI}\left(y^E\right) > W^{NI}\left(y^F\right).$$

For policy regime $I$, combining C.38 and the definition of $\bar{z}(y)$ in C.13, we have

$$\bar{z}\left(y^N\right) < \bar{z}\left(y^E\right) < \bar{z}\left(y^F\right).$$

Working from equation C.31, it can be shown that the level of welfare generated by the best-response allocation $\mathbf{c}^I$ is a strictly decreasing function of $\bar{z}(y)$.[5] Therefore, we have

[5]To see this result, differentiate the expression for $W^I$ with respect to $\bar{z}$. The resulting expression is lengthy, but can be shown to be strictly positive.

$$W^I\left(y^N\right) > W^I\left(y^E\right) > W^I\left(y^F\right),$$

as desired.

**Step (ii):** Establish the result. Consider any $e$ with $\delta > 0$ and let $y^*_{NI} \in \left\{y^N, y^E, y^F\right\}$ denote equilibrium strategy profile in the policy regime with no intervention if the economy is not/weakly/strongly fragile under that regime. Then equilibrium welfare is $W^{NI}\left(y^*_{NI}\right)$. Lemma 2 establishes that there exists $\bar{\sigma}_1 < 1$ such that $\sigma > \bar{\sigma}_1$ implies

$$W^I\left(y^*_{NI}\right) > W^{NI}\left(y^*_{NI}\right).$$

Lemma 3 establishes that there exists another cutoff point $\bar{\sigma}_2 < 1$ such that $\sigma > \bar{\sigma}_2$ implies the fragility type of the economy under the policy regime with intervention is weakly lower than under the regime with no intervention. Step (i) above establishes that lowering the fragility type of the economy always raises equilibrium welfare. Combining these results shows that whenever $\sigma > \max\left\{\bar{\sigma}_1, \bar{\sigma}_2\right\}$, we have

$$W^I\left(y^*_I\right) > W^{NI}\left(y^*_{NI}\right)$$

and allowing intervention strictly increases equilibrium welfare.

□

**Lemma 4.** *For any economy with $\delta = 0$ and $\sigma < 1$, $W^{NI}(y) > W^I(y)$ holds for all $y \in Y$.*

*Proof.* The proof is divided into two steps.

**Step (i)** : Show that when $\delta = 0$ and $\sigma = 1$, the allocations $\mathbf{c}^{NI}(y)$ and $\mathbf{c}^I(y)$ are equivalent for any $y$. This result follows from the expressions given for the two allocations in Appendix A. When $\delta = 0$, equation C.12 shows that $z_s(y) = x_s(y)$ for all $s$ and $y$. When $\sigma = 1$ also holds, equations C.3 and C.8 show that $c_1^{NI}(y) = \tilde{c}_1^I(y)$ for all $y$; equations C.4 and C.10 then show that $c_{ts}^{NI}(y) = c_{ts}^I(y)$ for $t = 1, 2$, for all $s$, and for all $y$. Using $\delta = 0$ in equations C.5 and C.11 shows that no public good is provided in either allocation. The only difference between the two allocations, therefore, is that the "distorted" payment $c_1^I(y)$ appears in the allocation under regime $I$. When $\sigma = 1$, however, no depositors receive this consumption level. In this sense, the two allocations are equivalent and necessarily generate the same level of welfare,

$$W^{NI}(y) = W^{I}(y).$$

**Step (ii)** : Establish the result. The intuition for this step is clear: the two regimes are equivalent when $\delta = 0$ and $\sigma = 1$, and lowering $\sigma$ below 1 will decrease welfare under regime $I$ while having no effect under regime $NI$. However, $W^{I}(y)$ does not necessarily change monotonically with $\sigma$ for all $\sigma < 1$. To establish the result, therefore, we consider the the auxiliary problem of maximizing

$$\sigma\pi_L u(\tilde{c}_1) + (1-\sigma)\pi_L u(c_1) + (1-\pi_L)\, _sum_{s\in S} q_s\left(\hat{\pi}_s u(c_{1s}) + (1-\hat{\pi}_s) u(c_{2s})\right) \tag{C.39}$$

subject to

$$\sigma\pi_L\tilde{c}_1 + (1-\sigma)\pi_L c_1 + (1-\pi_L)\left(\hat{\pi}_s c_{1s} + (1-\hat{\pi}_s)\frac{c_{2s}}{R}\right) \leq 1. \tag{C.40}$$

The solution to this problem is the best feasible allocation of resources in an economy where investors follow a given strategy profile $y$ and do not value the public good (that is, $\delta = 0$). It is straightforward to show that the first-order conditions characterizing this solution are given by 3.10, 3.11 and 3.14. In other words, the solution to this problem is equivalent to the best-response allocation under the policy regime with no intervention, $\mathbf{c}^{NI}(y)$. Note that the best-response allocation under the regime with intervention, $\mathbf{c}^{I}(y)$, is in the feasible set C.40, but is clearly not equal to the solution because C.9 shows that $\tilde{c}_1^{NI}(y) \neq c_1^{NI}(y)$. Since C.39 is strictly concave, $\mathbf{c}^{I}(y)$ generates strictly lower welfare than $\mathbf{c}^{NI}(y)$ when $\delta = 0$ and $\sigma < 1$, that is,

$$W^{NI}(y) > W^{I}(y)$$

for any $y \in Y$, as desired.

□

**Lemma 5.** *For any economy with $\delta = 0$ and $\sigma < 1$, allowing intervention weakly increases the fragility type of the economy.*

*Proof.* Step (i) of the proof Lemma 4 established that when $\delta = 0$ and $\sigma = 1$, $\mathbf{c}^{NI}(y)$ and $\mathbf{c}^{I}(y)$ are equal and, hence, we have

$$\frac{c_1^{NI}(y)}{c_{2s}^{NI}(y)} = \frac{\tilde{c}_1^{I}(y)}{c_{2s}^{I}(y)}$$

for all $s$ and $y$. It then follows from Propositions 5 and 7 that the fragility type of the economy is the same under both regimes. Proposition 6 established that the fragility type of an economy is independent of $\sigma$ under regime $NI$, while Proposition 8 established that it is weakly decreasing in $\sigma$ under regime $I$. Therefore, for any $\sigma < 1$, the fragility type of the economy must be weakly higher under regime $I$ than under regime $NI$.

□

### C.2.6 Proof of Proposition 11

**Proposition 11.**

*Proof.* Using the result from Lemma 4 with the equilibrium strategy profile under regime $NI$, when $\delta = 0$ and $\sigma < 1$ we have

$$W^{NI}\left(y^*_{NI}\right) > W^{I}\left(y^*_{NI}\right). \tag{C.41}$$

Lemma 5 establishes that the fragility type of the economy under regime $I$ is at least as high as under regime $NI$, while step (i) of the proof of Proposition 10 establishes that increasing the fragility type of an economy strictly lowers welfare under either regime. Combining these two results with C.41 yields

$$W^{NI}\left(y^*_{NI}\right) > W^{I}\left(y^*_{I}\right),$$

which establishes the proposition.

□